C. Sujatha

Vibration, Acoustics and Strain Measurement

Theory and Experiments

Ane Books
Pvt. Ltd.

Springer

Vibration, Acoustics and Strain Measurement

C. Sujatha

Vibration, Acoustics and Strain Measurement

Theory and Experiments

C. Sujatha
Department of Mechanical Engineering
IIT Madras
Chennai, India

To my husband Chandra Mohan and sons Adithya and Aravind

Preface

The current book is a sequel to my earlier book, Vibration and Acoustics: Measurement and Signal Analysis (Tata McGraw Hill Education Private Limited, 2010). The reason for bringing out a book of this nature is that there is no comprehensive book on theory and experiments in vibration, acoustics, and strain measurements.

The basic goal of the current book is to serve as a laboratory handbook to someone who plans to conduct experiments of a pedagogical nature in the areas of vibration, acoustics, and strain measurements (strain gauge based only). It provides just the requisite amount of theory, as well as descriptions and principles of operation of an assortment of transducers available for measurements, and sufficient signal processing theory to enable one to make a proper choice of settings on an analyser while making measurements. Also added is a description of modal analysis theory and related measurement techniques since there are some fundamental experiments in modal analysis which have also been included. Most of the theory and description of transducers and measuring equipment are from my earlier book, included with the permission of my earlier publisher.

I hope that the book will serve as a complete guide for educational institutions which want to set up a laboratory for undergraduate and postgraduate students or for a research and development organization which wants information on setting up a specialized laboratory for measurements in the fields mentioned. I also hope that the book will serve as a complete manual with which the reader will be able to set up standard and involved experiments as well. The book does not expect the reader to have any previous exposure to vibration, acoustics, or stress analysis, nor any rigorous mathematical background. I faced a considerable amount of dilemma in organizing the required material in the text. One option was to describe each experiment completely as would be done in a laboratory manual with all the background theory, objectives, test set-up with transducers, and instrumentation, followed by the test procedures and sample results. The other option was to present theory, description of transducers, experiments, and signal processing techniques separately. However, after much reflection, I decided to present the basic theory behind vibration, acoustics, and solid mechanics in separate chapters. So also, transducers related to these are arranged in separate chapters and the experiments themselves are featured in

different chapters. With this arrangement, the reader may have to navigate back and forth between theory, transduction techniques, and the actual laboratory experiments, instead of having everything provided on a platter for any specific experiment. This is done for the reason that the experimentalist/researcher may have the flexibility of mixing and matching test set-ups and transducers, rather than having to work within a rigid framework of test rigs and pickups. Also, some experiments share a common theory, so I decided to discuss the theory separately. Besides, there are certain experiments for which the theory is a little more involved than the basic theory described in the first few chapters and for these, the detailed theory is elucidated along with these experiments. Therefore, the explanations for certain experiments are concise, while those for others are a little more elaborate.

No amount of theory can compensate for the understanding of a subject that one gets when doing experiments. I sincerely hope that this book will prove useful for prospective experimentalists and will not just dispel the fear of doing experiments, but will in fact make it a joyful process.

Chennai, India   C. Sujatha

Acknowledgement

About the Book

This book is a comprehensive manual for setting up basic and involved experiments in the areas of vibration, acoustics and strain measurement (using strain gauges only). It will serve as a reference to conduct experiments of a pedagogical nature in these areas. It covers the various theoretical aspects of experimental test rigs, as well as the description and choice of transducers/equipment. While the book is targeted to meet the needs of academic and research and development organizations, industry will be benefitted equally. The contents have been formulated so as to help persons with no previous exposure to vibration, acoustics or stress analysis, or any rigorous mathematical background.

The book begins with an explanation of the need for such experimentation and testing, elucidated by the general theory pertaining to each area of vibration, acoustics and stress analysis. This is followed by a comprehensive description of a variety of vibration and acoustic pickups and exciters, as well as strain gauge transducers. The fundamentals of signal processing theory, including the basics of random signals have been included to enable the user to make a proper choice of settings on an analyzer/measuring equipment. Also added is a description of modal analysis theory and related parameter extraction techniques. The book illustrates around fifty experiments in the areas of vibration, acoustics and strain measurements in separate chapters. The book also includes a couple of experiments in modal analysis. The presentation is such that an academician could mix and match these experiments to suit the laboratory courses being taught at undergraduate or postgraduate level.

Suitable anecdotes are included, as appropriate, to make the subject more interesting. Besides, all chapters are provided with conceptual questions which will provoke the reader to think and gain a better understanding of the subjects.

Contents

About the Author

Prof. C. Sujatha is Professor in the Machine Design Section, Department of Mechanical Engineering, Indian Institute of Technology Madras, India. She did her postgraduation in Electrical Engineering in 1987 and Ph.D. in Applied Mechanics in 1991 from IIT Madras. She has 40 years of experience in teaching, research and industrial consultancy in the areas of machine dynamics, vehicular vibration, acoustics, instrumentation and signal analysis, condition monitoring, etc. She has also taught courses at the undergraduate and postgraduate levels in subjects related to these areas at IIT Madras. She has delivered numerous lectures and presented papers in international and national conferences. She has around 215 publications in peer-reviewed international and national journals and proceedings of international and national conferences. She has published a book entitled "Vibration and Acoustics: Measurements and Signal Analysis" with Tata McGraw-Hill Education Private Limited. She also has two monographs to her credit. She has guided 38 Ph.D. and M.S. (by Research) scholars and more than 125 M.Tech. and B.Tech. students. She has coordinated a large number of industrial consultancy projects for more than 250 industries in India. She has also undertaken many sponsored research projects and has been associated in an advisory capacity with prestigious national R&D and defence organizations.

Chapter 1
Introduction

With heavy industrialization and automation of many of the processes in mechanical engineering, the objective has become to operate machines at higher speeds and higher loads. This has led not only to higher power ratings of machines but also to larger vibration levels. These are accompanied by higher structure-borne sound with higher sound pressure levels emanating from the machines and larger dynamic stress reversals, manifesting as larger measured strains. Hence, from the point of view of understanding the dynamic behaviour of the machines and from the point of view of redesign if required, it becomes imperative to measure the vibrations, sounds, and strains produced. Mere measurement will not help without a knowledge of the physics behind the production of vibration, sound, and strain. Also, one cannot afford to treat the transducers and measuring equipment used as black boxes, since the choice of the wrong transducer or wrong setting on an analyser can give rise to erroneous data and interpretation of results. For this reason, it is required to understand the principle of operation of these measuring instruments and also to understand the principles of signal processing, so that the right measurement options may be exercised.

This book describes the theory, transducers, and measurement techniques required for carrying out basic experiments in vibration, acoustics, and strain measurements. It is good to start with the most preliminary pedagogical experiments in the laboratory before graduating to more complicated real-life experiments at site. Hence, experiments in vibrations, acoustics, and strain measurements have been presented with increasing complexity in this book. Of course, there are many readymade test rigs, which can be bought off the shelf. But I feel that the understanding that is gained using rigs fabricated in the workshop for specific experiments gives a better feel for the test conditions, complexities involved in the measurements and analysis of the results got. For very complex experiments, there may be no choice but to buy commercially available test setups. The experiments described in this book suggest simple test rigs with a choice of appropriate measuring transducers and measurement techniques. Sufficient theory about vibration, acoustics, and mechanics of solids has initially been presented. The amount of theory is restricted to the bare minimum as

will be required for an understanding of these experiments and is by no means comprehensive. The principle of operation of various transducers has been described so that the reader can make an intelligent choice of the measuring device. Salient signal processing concepts have also been described in brief to enable choice of appropriate measuring techniques and settings.

Vibration as we all know is caused by periodic back-and-forth motion of the particles of an elastic body or medium, when they are displaced from their equilibrium position. Any body that has inertia and elasticity is capable of vibrating. The simplest form is simple harmonic motion, like that of the pendulum of a clock. Vibration may be classified as forced or free vibration, depending on whether there is sustained external excitation or not. Vibrations are generally undesirable, like those of an automobile, but may also be desirable, like those of a mechanical sieve.

Sound may be thought of as the vibration of mechanical systems and waves in the medium through which it is propagating in the audible frequency range. When a tuning fork is struck, its vibrations periodically compress and displace the air particles around it and we hear this as a continuous tone of a single frequency. When this wave impinges on a person's ear, it encounters the eardrum, causing it to vibrate, in much the same manner that a musical drum vibrates when hit with a drumstick. These vibrations are transformed into electrical signals that the brain interprets as sound. Structural acoustics or vibro-acoustics is the study of the mechanical waves in structures and fluid-structure interaction, leading to the radiation of sound.

Mechanical strain is a geometric measure of the deformation representing the relative displacement between particles in a material body/components. When a material is loaded with a force, it produces a stress, causing it to deform. Strain is thus the response of a system to an applied stress, which may be static or dynamic. Engineering strain is defined as the ratio of the deformation in the direction of the applied force to the initial length of the material and is a dimensionless quantity.

How Does Vibration Create Noise?
The sound produced by a mechanical vibrating system and the strains induced in its components are closely related. Whenever a structure is set into vibration, it produces structure-borne sound and air-borne sound. We can see many examples around us as to how vibration creates noise. For example, when a gong is hit, it is set into vibration and these vibrations produce sound. Many musical instruments, except wind instruments, also produce sound in this fashion. If a panel or a wall is tapped, it vibrates and behaves like a large loudspeaker membrane. We are all familiar with the noise produced by a running road vehicle. This is caused by the vibration of the tyres of the vehicle as they go over a rough pavement; there is also interaction between the tyre treads and the road, resulting in traffic noise. Of course, the vibrations produced by a vehicle's engines create noise with their own typical spectra. We also have squeal from brake components due to the stick-slip action between the brake pads and the

disks causing radiation of sound. A railway vehicle produces a rattle, essentially due to sound radiation from the interaction between wheel and rail irregularities. We also have underwater sound due to vibrating structures and resulting pressure fluctuations in water. Sonar projectors serve as underwater loudspeakers and help in detecting objects like naval vessels and submarines, the surfaces of which vibrate due to the operation of their internal equipment and thus radiate sound. It is not surprising that the world's naval organizations spend enormous amounts of money and effort on noise reduction of these systems.

How Does Sound Produce Vibration?

It is to our advantage that sound produces vibrations giving us the sensation of hearing. Sound is essentially vibration of the air, giving rise to pressure fluctuations causing our eardrums to vibrate and make us hear. We know that windows can vibrate due to the noise from an aircraft flying overhead. We also know that the sound from opera singers can set a wine glass into resonance, causing it to shatter. Sound can cause vibration, which in turn can produce sound. Consider two adjacent rooms in an apartment building. When a noise is created in one room, it makes the common wall vibrate and radiate sound like a loudspeaker into the adjacent room. A substantially heavier wall between the two rooms will vibrate less and radiate less sound into the second room.

How Does Vibration Produce Stress?

A vibrating structure responds to its natural modes, resulting in dynamic stress. Machinery components, pipes and their connections in rotating machinery are often subjected to high-frequency vibrations that cause dynamic stress reversals, leading to failure. This happens because the cyclic forces cause deformation that initiates cracks in the areas of stress concentration. This may happen even when the machinery operates within safe design operating parameters. In most cases, component failures result from high-frequency fatigue. For this reason, it is important to detect and estimate high cycle stresses associated with vibrations as early as possible. Vibration fatigue is a term used in mechanical engineering to describe material fatigue caused by forced vibration. Shaker testing of components is generally done to understand their fatigue failure.

How Does Sound Cause Stress?

Sound-induced stresses have been of concern in aircraft and engine tests, as well as in rocket structures and launch facilities that experience intense pressure fluctuations during a launch. Sounds of high intensity can cause damage, not only to our hearing but also to big structures. In many of the early jet aircraft tested in the laboratory, cracks were caused on structures on the surface of the aircraft. These were attributed to high-frequency bending of the structures, a phenomenon called 'sonic fatigue' that was caused by the sound generated by the jet engines. This sound also had the potential of damaging the buildings in which they were being tested.

Why Measure Vibrations?

Vibrations are generally undesirable in machinery and structures, except in devices like mobile phones where they are deliberately produced. A machine or structure or vehicle may be too difficult to control, be uncomfortable to ride in, wear too fast, not maintain tolerances, fatigue prematurely, make too much noise or break down unexpectedly. All these scenarios necessitate the measurement of vibration, noise, and strain. Vibration measurements are often carried out either in the laboratory or at site. Tests are done during the developmental phase of a product to understand its behaviour or qualification tests are done after design. Typically, vibration measurements are made for the following reasons:

- Determination of the response of a vibrating machine/vehicle/structure.
- Determination of vibration levels in road vehicles, off-road vehicles, and tracked vehicles.
- Characterization of suspensions.
- Finding the response of a human being in a vibrating environment, such as in a vehicle, in a building or while operating a machine.
- Finding out the natural frequencies, damping ratios and mode shapes associated with various modes of vibration (modal analysis).
- Determination of the forces causing a particular vibration.
- Evaluation of vibration isolation.
- Calibration of a transducer.
- Evaluation of the damping of materials.
- Condition monitoring of a machine to find out if it is operating within safe levels; if not the cause of vibration leading to the condition has to be determined, or mitigation measures have to be taken.
- Estimating vibration levels on massive structures like machine foundations, bridges, highways, and buildings.
- Measurement of ground motion during an earthquake and response of structures to earthquake excitation.
- Measurement of ground vibrations due to construction activities, blasting and traffic that may have a detrimental effect on buildings.
- Comparison of vibration levels in similar consumer items of different makes, e.g., air conditioner, refrigerator, etc.
- Dynamic qualification tests for different requirements.
- Estimation of torsional vibration response or system properties.
- Passive and active vibration control.
- Estimation of optimal parameters for fault detection.

Why Do We Measure Sound?

Acoustic measurements are commonly made to characterize a source, for diagnostics, for understanding objective values or for subjective perception. Audio acoustic measurements would typically fall under the following categories.

- Estimation of sound pressure level emanating from a machine and its directivity index.

- Frequency analysis of sound signals (spectrum analysis).
- Estimation of sound power levels of identical machines/products to compare their performances.
- Estimation of sound intensity levels to identify the source of noise in a machine.
- Determination of reverberation time of an enclosure.
- Evaluation of sound absorption coefficients of acoustic materials.
- Studies on transmission of acoustic signals.
- Noise reduction techniques in machines and enclosed spaces.
- Measurement of environmental noise, such as traffic noise, noise near an airport or in a commercial area.
- Measurement of vehicle pass-by noise.
- Calibration of microphones.
- Subjective rating of noisiness.
- Audiometry to ascertain hearing loss.
- Measurement of noise to ascertain machinery condition.
- Measurement of noise in workshops and factories.
- Measurement of construction noise.
- Measurement and analysis of transient signals, such as sonic booms, explosions, gunshots, etc.
- Study of the effects of the above on human beings.

Need for Strain Measurements:
Strain measurements are typically made to quantify the strain a body is subjected to and strain gauges are used for this. In components subjected to mechanical or thermal stresses, strain gauges are used to measure average strain at a particular location and in a given direction. From the measured strain, it is possible to calculate the stresses knowing the modulus of elasticity. Measurements may be of a static or dynamic nature and are carried out on equipment found in power plants, ships, refineries, automobiles and industry at large. Strain gauges are extensively used in railways, aerospace industry and in many areas of mechanical and civil engineering.

The range of applications of strain gauges include:

- Measurement of force, pressure, tension, weight, torque, etc.
- Measurements in geotechnical applications on structures such as tunnels, underground cavities, buildings, bridges, concrete, masonry dams, embedment in soil/concrete, etc.
- Measurement of the torque applied by a motor, turbine, or engine to fans, generators, wheels, or propellers.
- Measurement of power by measuring the strain on a shaft and its rotational speed.
- Stress measurement on railway lines.
- Stress measurement on deflecting aircraft wings.
- Strain measurement during aircraft component testing.
- Determination of rotational strain on turbines, wheels, fans, propellers and motors.
- Determination of strains while testing ship hulls.

- Determination of strains while testing structural components of bridges and buildings.
- Determination of stresses in automobiles and their components.
- Detecting and localizing damage in structural health monitoring.
- Measurement of vibration modes using strain gauges in modal analysis.

Organization of the Book

The book has been arranged into 12 chapters, the first being an introductory one. Chapter 2 deals with basic vibration theory, while Chap. 3 describes transducers for measurement as well as for generation of vibration. Chapter 4 examines basic acoustic theory, while Chap. 5 reviews transducers and equipment used for sound measurement/generation. Chapter 6 gives an introduction to basic solid mechanics. Chapter 7 discusses strain gauge-based transducers and measurement techniques. Chapter 8 gives insight into fundamental signal processing concepts, including those of random signals. Chapter 9 throws light on the basics of experimental modal analysis, excitation and measurement techniques as well as modal parameter extraction methods. Chapter 10 describes 17 laboratory experiments in vibration. Chapter 11 gives details of 13 experiments in acoustics and Chap. 12 details 19 experiments in stress analysis using strain gauges. As mentioned earlier, this arrangement of the material was adopted after much debate, so as to allow the wide range of topics covered to be presented coherently. All chapters are provided with some interesting information, which hopefully will kindle more interest in the subjects discussed. Besides, all chapters are also provided with conceptual questions, which will make the reader think and gain a better understanding of the subjects. SI units have been used throughout the book and lists of symbols used have been provided wherever required. In this computational age, it is hoped that this book will motivate readers with a mechanical engineering background to conduct experiments and thus get an intuitive feel for vibrations, sound, and stresses.

Chapter 2
Theory of Vibration

INTERESTING FACTS: Studies in vibration probably started with the first musical instruments, namely, the whistles and the drums. As early as 4000 B.C., music had evolved well and was enjoyed by Chinese, Hindus, Japanese, and the Egyptians. Many scientists have made contributions to the field of vibration. Vibration theory was first proposed by the Pythagoreans in the fifth century BC, along with the theory of music and acoustics. They studied the natural frequency of vibrating systems and proved that it was a system property and not a function of excitation. Since China experienced many earthquakes, Chinese scientist Zhang Heng found the necessity and invented in 132 A.D. the world's first seismograph. Galileo was intrigued by the movements of a lamp in his church and studied the behaviour of a simple pendulum. He found that the time period of the oscillations was independent of the amplitude of vibration. He also observed the dependence of the frequency of vibration on the length of the simple pendulum and the phenomenon of resonance. Sir Isaac Newton's second law of motion is extensively used in modern books on vibrations to obtain the equations of motion of a vibratory system. The analytical solution to the problem of the vibrating string was found in 1713 by the English mathematician Brook Taylor, who was also responsible for the famous Taylor's theorem on infinite series. He obtained the natural frequency of vibration from the equation of motion and found it to agree with the experimental values observed by Galileo and Mersenne. Many later scientists such as Daniel Bernoulli, Jean D'Alembert, Leonard Euler, Lagrange, Charles Coulomb, E. F. F. Chladni, G. R. Kirchhoff, and Lord Baron Rayleigh contributed immensely to this field.

2.1   Introduction

This book deals primarily with the measurement and analysis of vibration, acoustic and strain signals, and common laboratory experiments; hence for a better understanding, some fundamental theories about vibration are discussed in this chapter. The study of vibration is concerned with the oscillatory motions of bodies. Many engineering structures and machines undergo vibrations which may be inherent and unwanted and occur because of the dynamic effects of manufacturing tolerances, clearances, rolling and rubbing contact between machine parts, unbalanced forces in rotating and reciprocating parts or due to resonance of structural elements as in the case of vibrations of a vehicle induced by ground undulations. On the other hand, certain vibrations perform a useful job and are purposely introduced, as in the case of a mechanical sieve. We also generate vibration intentionally in concrete compactors, ultrasonic cleaning baths, rock drills, etc. Vibration shakers/exciters are designed to impart a controlled level of vibration energy to machines/components/structures to examine their response and to ascertain if they can withstand the operational vibration environments. Irrespective of whether the vibrations are inherently or deliberately produced, an accurate description of the vibration is required through measurement and analysis.

The parameters of concern in the study of vibration are amplitudes, which may be expressed as displacement, velocity or acceleration, and excitation forces. In the metric system, the practical units for these quantities would be mm or μm (one thousandth of a mm) for displacement, mm/s for velocity, m/s^2 or g (one g being equal to 9810 mm/s^2) for acceleration and Newton for force.

The frequencies of interest in any vibration study are the natural frequencies of the system and the excitation frequencies, which can be related to their running speeds and their harmonics. The natural frequencies depend on the system masses and stiffness. Table 2.1 shows some typical values for the natural frequencies of some practical systems. The values given are indicative only; the exact values would depend on the system geometry and the material properties. The exciting or forcing frequencies of mechanical systems are usually in the range 1–160,000 in revolutions per minute (RPM), i.e., 1/60–2667 Hz as shown in Table 2.2.

Any system, which has mass and elasticity can vibrate. Most practical structures around us have multiple degrees of freedom (DOF) in the sense that a large number of independent coordinates are required to completely describe the motion of these systems. However, even very complicated multi-degree-of-freedom (MDOF) systems can be reduced to the simplest model, namely that of a single-degree-of-freedom (SDOF) system. The study of vibration consists of free and forced vibration. Free vibration occurs due to forces inherent in the system, while forced vibration is induced by externally impressed forces. If the excitation frequency of a structure coincides with any of its natural frequencies, a condition of resonance occurs. Failure of major structures like bridges, aeroplane wings, and buildings has been found to be due to resonance.

Table 2.1   Typical natural frequencies

System	Natural frequencies (Hz)
Offshore oil rig	1 or 2
Bridge	1–10
Car-bounce	1
Car-pitch	2.5
Military tank-bounce	0.5–1.5
Motorcycle weave	0.2–4
Motorcycle wobble	4–9
Ship	0.5–10
Multi-storeyed building	1–10
Railway car-bounce	1
Railway bogie-bounce	5
Railway wheelset-bounce	200
Human trunk	2.5–5
Human vertebrae	100s
Power train of sports utility vehicle (SUV)	50–300
SUV compartment cavity	100s
Hard disk drive	100s–1000s
Turbine blade low pressure stage	10s–100s
Turbine blade high pressure stage	100s–1000s
Inner/outer races of rolling element bearings	Few 1000s

Table 2.2   Typical running speeds

System	Normal running speed (RPM)
Kiln	1
Ball mill	15–20
Low speed marine diesel engine	60–480
Compressors	500–3000
Induced draft (ID)/Forced draft (FD) fan	600–1200
Induction motor (4% slip)	1440 for 50 Hz supply frequency
Direct current (DC) motor	100–6000
Steam turbine	3000 for 50 Hz supply frequency
Car (including race car) engine	600–15,000
Hard disc drives	5400–15,000
Kitchen blender motor	0–20,000
High speed spindles of lathes	60,000 and above
Turbochargers	150,000
Dental drill	Up to 600,000

2.2 Single-Degree-of-Freedom System

DID YOU KNOW that most of the work in the area of vibration started with treatises on sound and the vibration of strings? One of the earliest publications on vibrating bodies was in 1638 by Galileo. The SDOF system is the simplest model of a vibratory system. The motion of such a system can be described by a single coordinate, which could correspond to translation or rotation. Reduction of an MDOF system to one with a single DOF is done by lumping all masses to an equivalent one, or effectively finding the total mass and by finding the equivalent stiffness (essentially determined by the softest spring) and by finding damping through experiments. It is recommended that while doing a structural analysis of complex systems like automobiles, ships, oil rigs, aeroplanes, windmills, water towers, aircraft wings, bridges, multi-storeyed buildings and machinery with multiple DOF, even if the complete geometry and material properties are known, the system should first be reduced to one with a single DOF to understand it. As far as torsional vibration studies are concerned, Charles Coulomb (1784) conducted analytical and experimental studies on the torsional vibration of a metal cylinder suspended by a wire and obtained its equation of motion.

2.2.1 Translatory System

The SDOF vibratory model consists of a non-elastic mass, which is suspended from a linear massless spring (Fig. 2.1a). Figure 2.1b shows the free body diagram (FBD) of the mass of a simple SDOF undamped system, which is assumed to move in the vertical direction only and which is described by a single coordinate x.

When the system is set into motion, it vibrates at a frequency f_n called the natural frequency, which is a function of its mass and stiffness parameters. If the deformation of the spring corresponding to the static equilibrium position is assumed to be Δ,

Fig. 2.1 Undamped SDOF system: **a** translatory SDOF system, **b** free body diagram of mass m
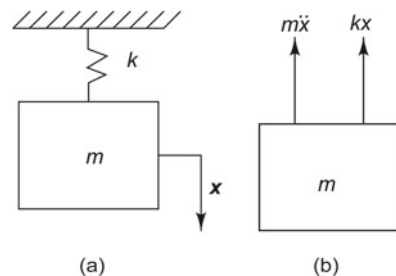
the spring force $k\Delta$ is equal to the gravitational force $W = mg$ acting on the mass. Therefore, by measuring the dynamic displacement x from the static equilibrium position, the spring force and the gravitational force nullify each other and need not be considered in the equation of motion as long as the spring is linear. If the system is disturbed by moving the mass downwards by a distance x and released, the spring k deforms in proportion to the force acting on it according to Hooke's Law. The mass m experiences an acceleration proportional to the force acting on it. While drawing the free body diagram, a fictitious inertia force equal to $m\ddot{x}$ which is in a direction opposite to that of the acceleration, is introduced on the mass, to reduce the problem to one of statics. Since there are only two forces keeping the mass in equilibrium, their sum must be equal to zero. Hence applying Newton's second law of motion to the mass we obtain

$$m\ddot{x} + kx = 0 \tag{2.1}$$

This equation is a homogeneous, second-order, linear differential equation. Its solution can be derived mathematically and is found to be harmonic. It can be expressed in the form

$$x(t) = X \exp(i\omega t) \tag{2.2}$$

The system has only one mode of vibration, which is defined by the motion along the x coordinate and which is obtained by substituting Eq. (2.2) in (2.1) to give

$$(-m\omega^2 + k)X = 0 \tag{2.3}$$

or

$$\omega = \pm \left(\frac{k}{m}\right)^{1/2} = \pm\omega_n \tag{2.4a}$$

Here ω_n is the natural frequency in rad/s. The natural frequency in Hz is

$$f = \pm\frac{1}{2\pi} \left(\frac{k}{m}\right)^{1/2} = \pm f_n \tag{2.4b}$$

The period of oscillation in seconds is

$$T = \frac{1}{f_n} = \frac{2\pi}{\omega_n} \tag{2.4c}$$

It is to be noted that the natural frequency and the period depend only on the mass and stiffness of the system, which are system properties. The general solution to Eq. (2.1) is of the form

$$x(t) = A \exp(i\omega_n t) + B \exp(-i\omega_n t) \tag{2.5}$$

The constants A and B can be obtained by substituting the known values of the initial conditions of the problem, i.e., the known values of the initial displacement and initial velocity.

2.2.2 Torsional System

Besides translational motion, we often come across angular oscillations, e.g., a section of a shaft moving clockwise or anticlockwise with respect to the other cross sections. Such a case is said to be one of torsional vibration. The whole shaft rotates with uniform speed with superposition of angular accelerations or decelerations of some sections, causing twisting or untwisting of the shaft. This is not perceptible to the naked eye, unlike in the case of longitudinal or lateral oscillations. However, this can be discerned as a frequency modulated tone by an experienced ear. Such problems are prevalent in machines with rotating parts such as punch presses, shearing machines, reciprocating compressors, pumps, turbines, and reciprocating engines in which the input or output torque is fluctuating. The results described above for translational oscillations are applicable to all SDOF systems with rotation, for which case the mass is to be replaced by a mass moment of inertia and the linear spring by a torsional spring.

Let J be the mass moment of inertia of the rotor about its axis and q the torsional stiffness (torque needed per unit angular displacement) of the shaft in Fig. 2.2. Let θ denote the angular rotation of the rotor with respect to a section of uniform velocity. Applying Newton's law, the sum of the torques acting on the rotor must be equal to

Fig. 2.2 Single-degree-of-freedom torsional system
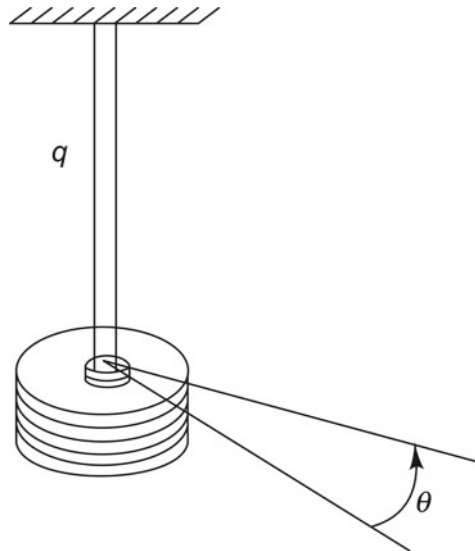
zero. Hence

$$J\ddot{\theta} + q\theta = 0 \qquad (2.6)$$

Assuming a harmonic solution as in the case of translation, we get

$$-J\omega_n^2\Theta + q\Theta = 0 \qquad (2.7)$$

The natural frequency of torsional oscillation is

$$\omega_n = \sqrt{\frac{q}{J}} \; \text{rad/s} \qquad (2.8)$$

It is to be noted here that the rotor and the shaft being in series, the component with the smaller torsional stiffness, namely, the shaft, is taken as the elastic member and the component with the larger mass moment of inertia, namely, the rotor, is assumed to contribute towards the inertia, though in reality both these members have both inertia and elasticity to varying degrees. Typically, the angular displacements that we come across in practice are very small, around 2° or 0.05 radians at the most; the shear strain and the resulting dynamic shear stress nevertheless can still be large. The lowest torsional natural frequency of practical systems may be of the order of 70–2000 rad/s (11–318 Hz approximately). As in the case of lateral vibration, there are cases when the torsional natural frequency of a system is in the vicinity of the forcing frequency, causing resonance. A typical example is a steam turbine for power generation where natural frequencies of some of its components have been observed to be close to the running speed (3000 rpm for 50 Hz power supply).

2.2.3 Damping

In any practical vibratory system, the physical parts of the system contributing to the inertia and elasticity can easily be identified. In addition to mass and stiffness, there is a third very important parameter, which is responsible for the dissipation of energy in a vibrating system. This parameter is called damping and is not so explicitly seen, since most mechanical structures are inherently lightly damped. Damping is obvious only in cases like the shock absorber in a vehicle or the dampers in an automatic door closer or a gun recoil system. However, it is clearly seen that the vibration amplitude keeps falling in the case of free vibration after a system has been set into motion and that amplitude at resonance during a forced vibration does not shoot to infinity (because the energy supplied by the excitation is balanced by loss of energy), indicating that there is some mechanism present in all vibratory systems for the dissipation of energy. Ignoring damping can give rise to grossly wrong results in a forced vibration study, especially at frequencies close to the natural frequencies, though not at frequencies far away from resonance.

Damping is brought about by a large number of phenomena like intermolecular friction or air resistance. It is difficult to get a closed form expression for the combined effect of these phenomena and one has to resort to simplified models or experimentally obtained values instead. Three damping models, which are commonly used are as follows:

 (i) Viscous damping model.
 (ii) Structural damping model.
(iii) Coulomb damping model.

2.2.4   Free Damped Vibration

In the model shown in Fig. 2.1, there is no damping. To simulate practical situations though, one needs to include damping as described in the earlier section. The viscous damping model is the most popular one, the reason being that the introduction of this linear damper is very simple and it also adequately explains the response observed experimentally. Besides, it gives rise to a linear differential equation with constant coefficients, the solution to which is well known and simple. Referring to the SDOF system with damping coefficient c and the free body diagram shown in Fig. 2.3, the equation of motion is a homogeneous differential equation of the form

$$m\ddot{x} + c\dot{x} + kx = 0 \tag{2.9}$$

Traditionally a solution of the form shown below is assumed

$$x(t) = X \exp(i\omega t) \tag{2.10}$$

On plugging in this solution into Eq. (2.9), the roots of the quadratic equation obtained are given by

Fig. 2.3 Single-degree-of-freedom system with damping: **a** schematic drawing, **b** free body diagram
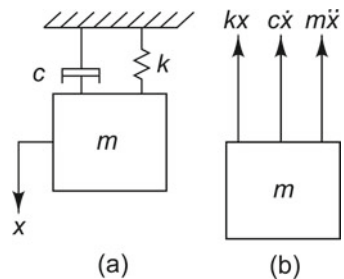
$$\omega_{1,2} = \frac{-\frac{c}{m} \pm \sqrt{\left(\frac{c}{m}\right)^2 - \frac{4k}{m}}}{2} = \sqrt{\frac{k}{m}} \left[\frac{-c}{2\sqrt{km}} \pm \sqrt{\left(\frac{c}{2\sqrt{km}}\right)^2 - 1} \right] \quad (2.11a)$$

Since $\omega_n = \sqrt{k/m}$, the natural frequency without damping, the natural frequency with damping is given by the equation below for $(c/m)^2 < (4k/m)$.

$$\omega_{1,2} = \omega_n \left[\frac{-c}{2\sqrt{km}} \pm i\sqrt{1 - \left(\frac{c}{2\sqrt{km}}\right)^2} \right] \quad (2.11b)$$

$c/2\sqrt{km}$ is dimensionless and can be expressed as a quantity $\zeta = c/c_c$ known as damping ratio, where c_c is known as critical damping coefficient and is equal to $2\sqrt{km}$. Then

$$\omega_{1,2} = -\zeta\omega_n \pm i\omega_n\sqrt{1 - \zeta^2} \quad (2.12)$$

Hence the solution to Eq. (2.9) is given by

$$x = e^{-\zeta\omega_n t} \left(c_1 \cos\left(\omega_n\sqrt{1 - \zeta^2}t\right) + c_2 \sin\left(\omega_n\sqrt{1 - \zeta^2}t\right) \right) \quad (2.13a)$$

The constants c_1 and c_2 may be obtained from the initial displacement and initial velocity. With initial conditions $x(0)$ and $\dot{x}(0)$, the above equation can be simplified as

$$x = e^{-\zeta\omega_n t} \left[x(0) \cos\left(\omega_n\sqrt{1 - \zeta^2}t\right) + \frac{\dot{x}(0) + \zeta\omega_n x(0)}{\omega_n\sqrt{1 - \zeta^2}} \sin\left(\omega_n\sqrt{1 - \zeta^2}t\right) \right] \quad (2.13b)$$

For most engineering problems considering materials like mild steel or aluminium, $\zeta \ll 0.1$ and hence $\sqrt{1 - \zeta^2}$ can be taken as 1. Vibrations with such low ζ are said to be underdamped and the corresponding motion is oscillatory. The response can then be simplified to

$$x = e^{-\zeta\omega_n t} (c_1 \cos \omega_n t + c_2 \sin \omega_n t) \quad (2.14)$$

ζ is found to be frequency dependent and of the order of 10^{-1}–10^{-3} for materials like mild steel and aluminium. It may also be noted that when $\zeta = 1$, $c = 2\sqrt{km}$. This damping is known as critical damping. When $\zeta = 1$, the two roots of Eq. (2.11) are equal. The system comes to rest in the minimum possible time with no oscillation as given by the solution

$$x = e^{-\omega_n t} (c_1 + c_2 t) \quad (2.15)$$

Critical damping is incorporated in systems where oscillations are to die down within the shortest period once the system is excited, as in the case of the recoil system in the firing of a bullet or closing of a door using an automatic door closer. In instruments like accelerometers and vibrometers, eddy current damping is deliberately introduced so as to have a damping ratio of 0.65–0.7 to increase their useful frequency range. The case corresponding to $\zeta > 1$ is said to be overdamped or non-oscillatory motion; the free vibration record shows an exponentially decreasing function of time and is aperiodic for this case.

2.2.4.1   Determination of Damping Ratio ζ by Logarithmic Decrement Method

One of the easiest methods to determine the damping ratio of a system is by the logarithmic decrement method which is applied to the decaying time response of the system. This decaying response curve for an MDOF system contains a mixture of modes. The unfiltered signal may be used as such for the determination of damping ratio of the first mode. However to find the damping associated with the higher modes using this method, the time decay data will have to be band pass filtered before applying the logarithmic decrement method and is therefore a little cumbersome. It might be more convenient to use the frequency domain-based half-power method instead.

Equation (2.13) has been plotted in Fig. 2.4. Here the amplitudes of successive waves are separated in time by $T_d = 2\pi/\omega_d$ and the envelope of the decaying oscillation is an exponential curve, which lends itself to a convenient definition of damping. The ratio of any two successive amplitudes is given by
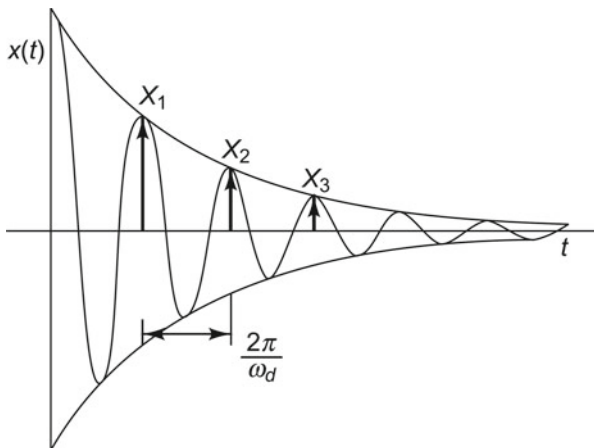
Fig. 2.4 Damped system response

$$\frac{X_1}{X_2} = e^{\zeta \omega_n \left(t - t + \frac{2\pi}{\omega_d}\right)} = e^{2\pi \zeta / \sqrt{1-\zeta^2}} \tag{2.16}$$

where X_1 and X_2 are the amplitudes of the first and second cycles, respectively. The natural logarithm of the ratio of any two successive displacements known as logarithmic decrement is given by

$$\delta = \ln\left(\frac{X_1}{X_2}\right) = \frac{2\pi \zeta}{\sqrt{1-\zeta^2}} \tag{2.17a}$$

The logarithmic decrement can also be defined as

$$\delta = \frac{1}{n} \ln\left[\frac{X_1}{X_{n+1}}\right] \tag{2.17b}$$

where X_{n+1} is the amplitude of the $(n+1)^{\text{st}}$ cycle. Equation (2.17b) is especially convenient for computation of δ and hence ζ from measurements, since measuring two successive amplitudes (as required in Eq. 2.17a) without error may be difficult for the reason that the two values are very close to each other. The relationship between the logarithmic decrement and the damping ratio of the system, ζ is

$$\zeta = \frac{\delta}{\sqrt{\delta^2 + (2\pi)^2}} \approx \frac{\delta}{2\pi} \tag{2.17c}$$

2.2.4.2 Viscous Damping Model

The most common damping model used is that of viscous damping described by a force proportional to the velocity of vibration. This model is based on the viscous resistive force of the oil when a piston moves in an oil-filled cylinder. The damping force is expressed by the equation

$$F_d = c\dot{x} \tag{2.18}$$

where c is a constant of proportionality called damping coefficient and has the dimension of force per unit velocity (Ns/m in SI units). This damping model may not always result in the best idealization, since it assumes that the damping is frequency independent. The energy dissipated per cycle of oscillation due to a damping force F_d can be calculated as

$$W_d = \oint F_d dx \tag{2.19a}$$

The force–displacement curve will depend on the type of damping present, but will always enclose an area called the hysteresis loop, which is proportional to the energy lost per cycle W_d. Such a loop is shown in Fig. 2.5a for the case of viscous damping.
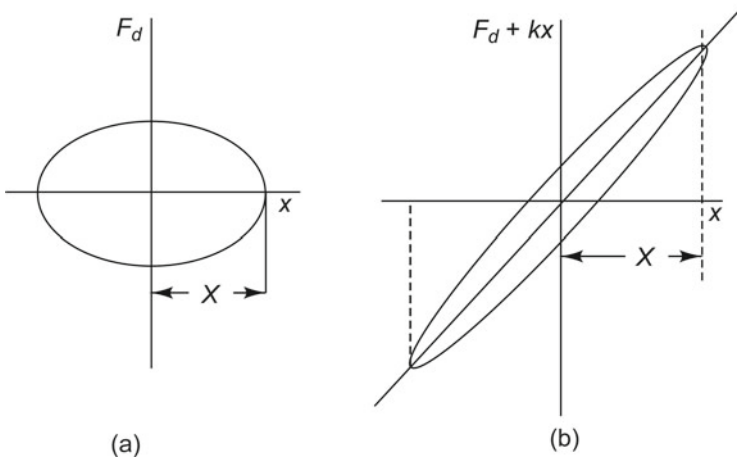
Fig. 2.5 Hysteresis loops for energy dissipated due to viscous damping: **a** viscous damping model, **b** Voigt model

Figure 2.5b shows the loop for a Voigt model where the spring is parallel to the damper (Fig. 2.3a). In the case of harmonic vibration of the form described in Eq. (2.2) with viscous damping

$$W_d = \oint c\dot{x}\, dx = \oint c\dot{x}^2\, dt = \pi c\omega X^2 \qquad (2.19b)$$

For any damping model other than viscous, the equivalent damping C_{eq} can be obtained by setting equal the energy dissipated by the viscous damping force to that of the non-viscous damping force. Harmonic motion is implied in this equation.

$$\pi C_{eq}\omega X^2 = W_d \qquad (2.19c)$$

Here W_d corresponds to the work done per cycle for the particular type of damping force considered.

2.2.5 Structural/Hysteretic Damping

Many real-life structures exhibit a frequency dependent damping behaviour. Stiffness of materials like rubber can be expressed in the form $k + ih = k(1 + i\xi)$, h being a complex stiffness and ξ the hysteretic or structural damping ratio. The equation of motion when such materials are used as a spring is given by

$$m\ddot{x} + (k + ih)x = m\ddot{x} + kx(1 + i\xi) = 0 \qquad (2.20)$$
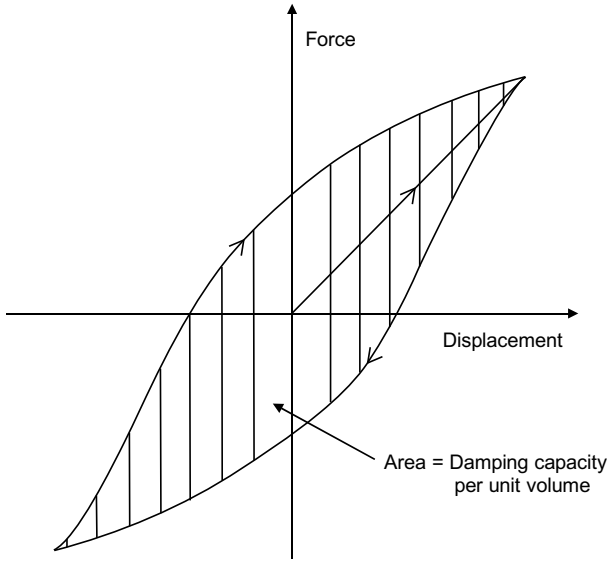
Fig. 2.6 Hysteresis loop for structural damping

Table 2.3 Structural damping of common materials

Material	ξ
Aluminium	$0.0002-0.001$
Steel	$0.01-0.06$
Cast iron	$0.03-0.06$
Concrete	0.15
Masonry structures	$0.3-0.6$
Rubber	$0.1-0.5$
Cork	$0.1-0.2$

When this is compared with Eq. (2.9), it is seen that the term $c\omega$ is replaced by $k\xi$. This damping model poses difficulties for a rigorous free vibration analysis and is generally used for forced vibration analysis. Figure 2.6 shows the hysteresis loop for the case of structural damping. Table 2.3 shows the structural damping loss factor ξ for typical materials that we come across in engineering practice. These values are rough indicative values and may change depending on the actual constitution of the material.

Fig. 2.7 Forced vibration of
undamped system

F cos ωt (External force)
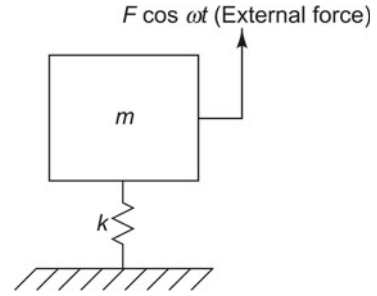
2.2.6   Forced Vibration

As in the case of free vibration, here also we shall study the undamped SDOF system first, followed by the damped system. The simplest case of forced vibration is that with harmonic excitation, for which the steady-state response is also harmonic at the forcing frequency.

2.2.6.1   Without Damping

Let us now consider a simple undamped SDOF system subjected to excitation forces which may come from say, a drive motor in a machine. The governing differential equation for such a system (Fig. 2.7) would then be

$$m\ddot{x} + kx = F \cos \omega t \tag{2.21}$$

The solution to this equation has two parts: we have the homogeneous differential equation, which is the complementary function corresponding to the free damped vibration of the system when $f(t)$ is equal to zero, and which would disappear in a matter of seconds (even for very small values of ζ). The second part of the solution is the particular integral, corresponding to $f(t)$ not equal to zero and is the response to the excitation force and is called the steady-state response. The particular solution is obtained by substituting $x = X \cos \omega t$ in Eq. (2.21) to give

$$- m\omega^2 X + kX = F \tag{2.22}$$

Hence

$$X = \frac{F}{k - m\omega^2} = \frac{(F/k)}{(1 - \eta^2)} \tag{2.23a}$$

where

$$\eta = \frac{\omega}{\omega_n} = \frac{\omega}{\sqrt{k/m}} \tag{2.23b}$$
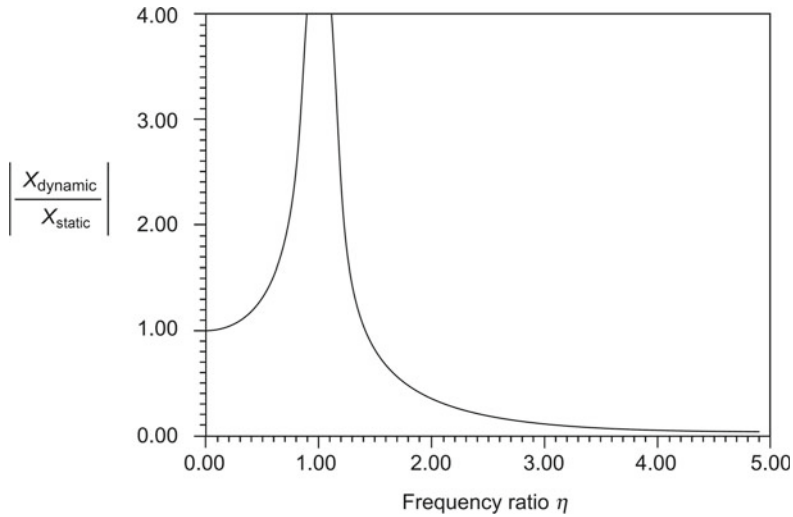
Fig. 2.8 Dynamic magnifier for various frequency ratios

η is known as the frequency ratio, the ratio of the external forcing frequency to the natural frequency of the system. Figure 2.8 shows the response as a dynamic magnifier, i.e., ratio of the dynamic displacement to static displacement as a function of the frequency ratio. The static deflection $X_{st} = (F/k)$ corresponds to $\eta = 0$. When $\eta = 1$, the forcing frequency is equal to the natural frequency; this corresponds to a condition of resonance and the dynamic magnifier becomes ∞. As η becomes very large, i.e., as the forcing frequency becomes much higher than the natural frequency, the dynamic deflection tends to zero.

2.2.6.2 With Damping

Figure 2.9a shows an SDOF system with viscous damping. The governing differential equation can be written as

$$m\ddot{x} + c\dot{x} + kx = F\cos\omega t \tag{2.24}$$

Assuming a harmonic solution of the form

$$x = X\cos(\omega t - \phi) \tag{2.25}$$

Equation (2.24) can be rewritten as

$$(-m\omega^2 + \mathrm{i}c\omega + k)X = F \tag{2.26}$$
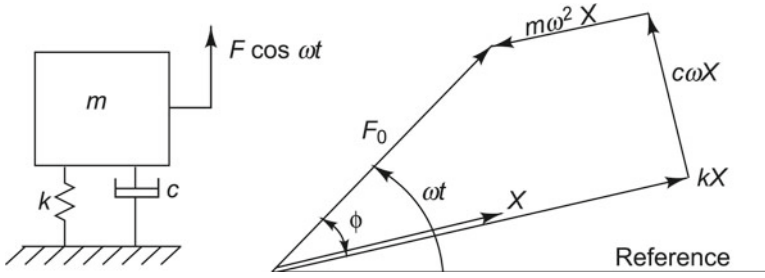
Fig. 2.9 Forced vibration of damped SDOF system: **a** schematic of SDOF system, **b** vector diagram showing various forces

Figure 2.9b shows the vectors representing the inertia, damping and elastic forces, as well as the external force in the above equation.

Hence, the particular solution can be expressed as

$$X = \frac{F \cos \omega t}{(k - m\omega^2 + i c\omega)} \tag{2.27}$$

Dividing the numerator and denominator by k and using the non-dimensional values $\zeta = c/c_c$, $\eta = \omega/\omega_n = \omega/\sqrt{k/m}$, we get the expression for the dynamic magnifier as

$$\left| \frac{X_{\text{dynamic}}}{X_{\text{static}}} \right| = \frac{X}{(F/k)} = \frac{1}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \tag{2.28a}$$

Here X_{static} is the static displacement. The angle of lag ϕ is given by

$$\tan \phi = \frac{2\zeta\eta}{(1 - \eta^2)} \tag{2.28b}$$

Figure 2.10 indicates the variation of the dynamic magnifier and phase angle ϕ as a function of η. At $\eta = 1$ (resonance), the angle of lag is always 90° for any damping ratio. For values of $\eta \ll 1$, the phase angle is close to 0°, whereas it is close to 180° for $\eta \gg 1$. At resonance the dynamic magnifier is equal to $(1/2\zeta)$.

For hysteretic damping, the dynamic magnifier $(1/2\zeta)$ at resonance gets replaced by $(\sqrt{km}/2c) = (1/\xi)$; or in other words, it is nothing but the reciprocal of the hysteretic damping coefficient.

2.2.6.3 Evaluation of Damping by Half-Power Method

One common way of determining damping ratio is by measuring the frequency bandwidth, between frequencies on the response curve, for which the response is some fraction of the resonant response of the system. The usual procedure is to

Fig. 2.10 Forced vibration with damping: **a** dynamic magnification factor, **b** phase angle

(a)

(b)
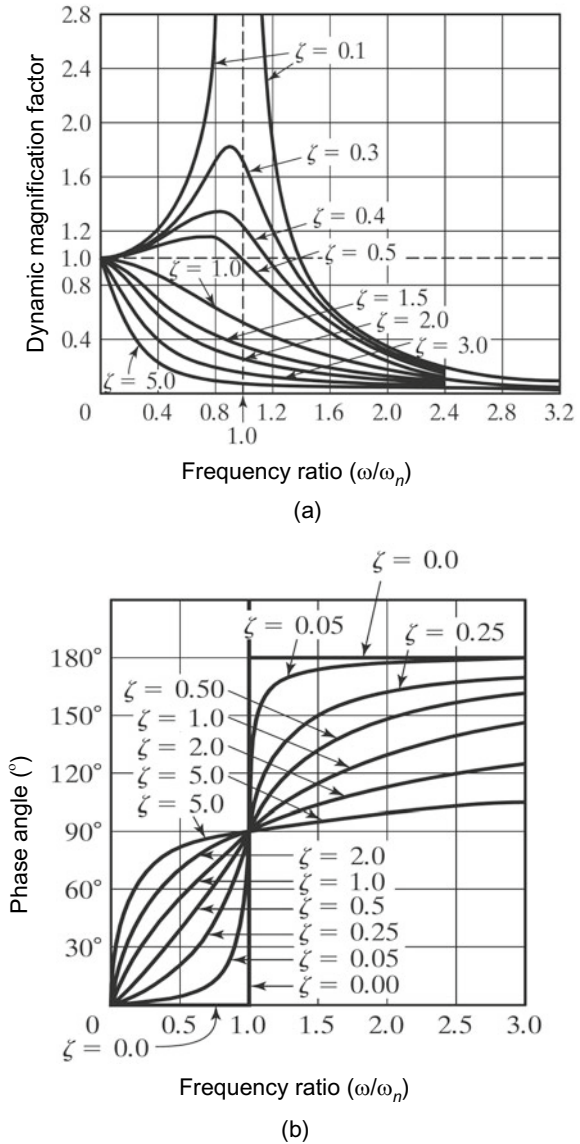
consider two frequencies on either side of the resonant frequency where the amplitude of the response is $1/\sqrt{2}$ times the maximum response. The bandwidth corresponding to these frequencies is called the "half-power bandwidth", a term used in the analysis of electrical systems in which the amplitude squared of the current is a measure of the electric power. This amplitude ratio of $1/\sqrt{2}$ corresponds to a change in amplitude of $20\log_{10}(1/\sqrt{2}) = -3.01$ dB or in other words, a reduction of 3 dB. Hence, a
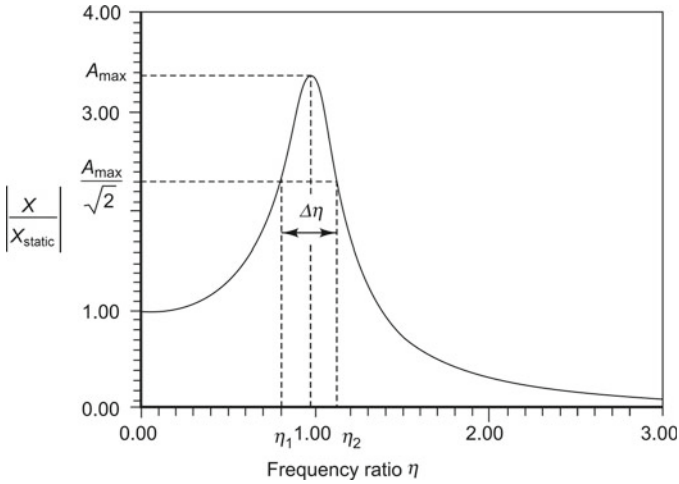
Fig. 2.11 Determination of damping from half-power method

measurement associated with an amplitude ratio of $1/\sqrt{2}$ is frequently described as "3 dB bandwidth" or half-power bandwidth and is given by $\Delta\omega = (\omega_2 - \omega_1)$ (Fig. 2.11). The damping ratio can be got from the following formula.

$$\zeta = \frac{\Delta w}{2\omega_n} \tag{2.29a}$$

and

$$\xi = 2\zeta \tag{2.29b}$$

Generally, the damping of a system can be determined from the bandwidth between any two points A and B associated with any amplitude ratio $1/n$, where n is greater than 1. Although there are practical limits of measurements, values of n greater than $\sqrt{2}$ or less than $\sqrt{2}$ can occasionally be used.

2.2.7   Response with Unbalanced Forces

One of the most important applications of forced vibration is the case of vibration due to rotating unbalances. All manufactured rotors have some amount of eccentricity. There is no rotor which has zero eccentricity. This may be due to the presence of keys, casting defects, and due to non-uniformity in welding. Besides, when three or more parts are assembled to form a combined rotor, the individual eccentricities due to the permitted tolerances in an assembly may add up to worsen the quality of manufacture of the combined rotor. Let us consider the rotor of mass m shown in

Fig. 2.12 Unbalanced rotor
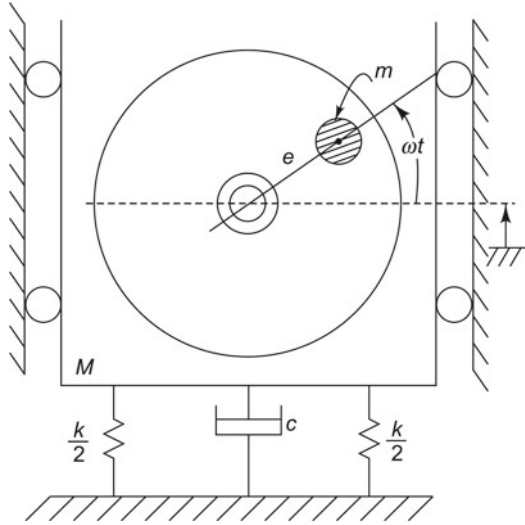
Fig. 2.12. Let it have an eccentricity of e (a quality of balance ωe for a rated speed of ω).

Let the displacement of the non-rotating mass $M - m$ be x. Then the displacement of the rotor is obtained as $x + e \sin \omega t$. The equation of motion is given by

$$(M - m)\ddot{x} + m \frac{d^2}{dt^2}(x + e \sin \omega t) = -c\dot{x} - kx \qquad (2.30a)$$

This can be rewritten as

$$M\ddot{x} + c\dot{x} + kx = me\omega^2 \sin \omega t \qquad (2.30b)$$

This is the expression for forced vibration with excitation force amplitude $me\omega^2$. The steady-state displacement response can hence be written as

$$X = \frac{me\omega^2}{[(k - M\omega^2)^2 + (\omega c)^2]^{1/2}} \qquad (2.31)$$

This can be reduced to

$$\frac{MX}{me} = \frac{\eta^2}{[(1 - \eta^2)^2 + (2\zeta\eta)^2]^{1/2}} \qquad (2.32)$$

The variation of MX/me with η is shown in Fig. 2.13 for zero damping. When $\eta = 0$, (when the rotor is not rotating), $X = 0$. When $\eta = 1$, the response is infinity.
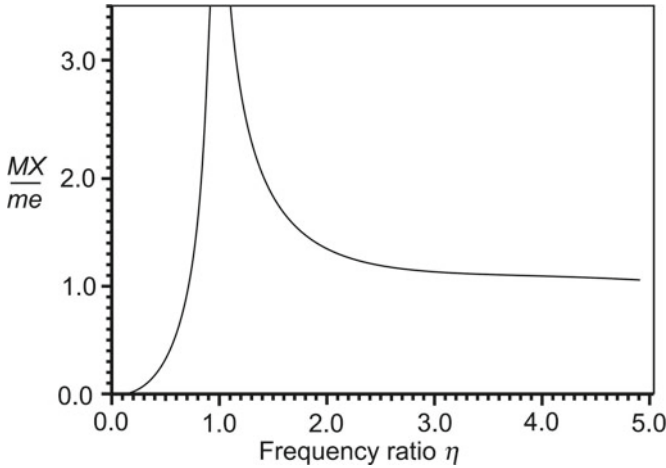
Fig. 2.13 Response of an unbalanced rotor

2.2.7.1 Synchronous Whirl of Rotating Shafts

Consider a shaft carrying a rotor (disk) and mounted between two bearings as shown
in Fig. 2.14. Let the mass of the shaft be negligible in comparison to that of the rotor.
Let G be the centre of gravity (CG) of the disk and O be the centre of the shaft in
the plane of the disk. Distance OG is the eccentricity e of the disk caused by internal
material defects, manufacturing errors, etc. The line CSD corresponds to the axis of
rotation. If e is expressed in mm, then the rotor is said to have a quality of balance
or balance grade ωe for a rated speed of ω rad/s. The deflected shape of the shaft is
as shown in the figure. Let the dynamic deflection of the rotor be x in the plane of
the disk. Then the centrifugal force experienced at the CG is $m\omega^2(x + e)$ radially
outwards. The restoring force experienced by the shaft at O is kx where k is the
restoring force per unit deformation of the shaft at its centre, or in other words, the
stiffness of the shaft when treated as a beam with length L, area moment of inertia
I and Young's modulus E is

$$k = \frac{48EI}{L^3} \tag{2.33}$$

Let us assume that the damping force due to air-friction is negligible. The shaft
rotates about its own axis; also the plane containing the bent shaft and the line of
bearings rotates about an axis which coincides with the line of bearings. When these
two rotational speeds are identical, we have what is called a condition of synchronous
whirl.
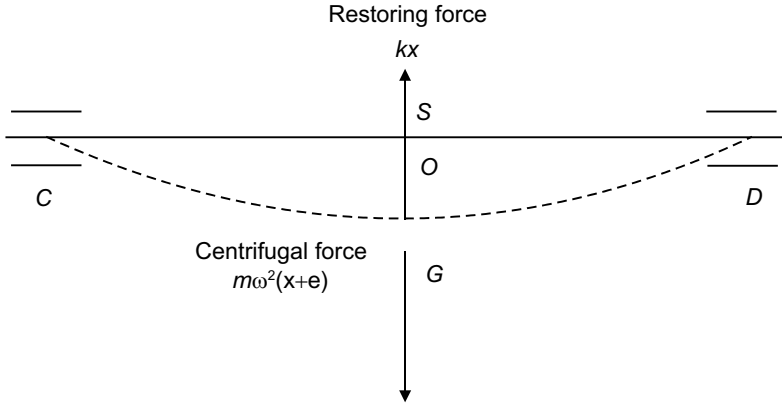
Fig. 2.14 Unbalanced rotor

Equating the forces in the vertical direction, we have

$$m\omega^2(x + e) = kx \qquad (2.34)$$

Therefore,

$$x = \frac{m\omega^2 e}{k - m\omega^2} \qquad (2.35)$$

Or

$$\frac{x}{e} = \frac{\omega^2 e}{(k/m) - \omega^2} = \frac{\eta^2}{1 - \eta^2} \qquad (2.36)$$

where

$$\eta = \frac{\omega}{\sqrt{(k/m)}} = \frac{\omega}{\omega_n} \qquad (2.37)$$

ω_n is called the critical speed of the shaft. The response of the shaft is as shown in Fig. 2.13. When the running speed of the system and the natural frequency of lateral vibrations coincide, the shaft tends to bow out with a large amplitude. This speed is known as the whirling speed or critical speed. It is necessary to avoid such a resonance condition in actual practice since the rotor cannot run at this speed continuously. For values of η less than 0.1, x/e will be less than 0.01. Since the eccentricity itself is quite small, the dynamic deflection will also be negligibly small and the rotor can be treated as rigid.

Fig. 2.15 Force
transmissibility
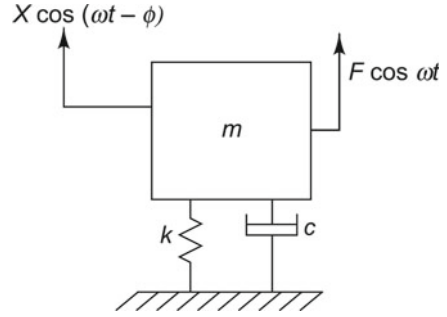
2.2.8 Vibration Isolation

Vibration isolation is a very important problem to be dealt with in dynamic sys-
tems and is of concern in steady-state disturbances. There are two types of isolation
problems present:

 (i) Force isolation, in which it is to be ensured that a machine with inherent vibratory
    forces is mounted in such a way that it does not transmit the disturbances to the
    surroundings, and
(ii) Displacement isolation, in which it is to be made sure that a sophisticated instru-
    ment which is to be placed in a vibratory environment is not affected by the base
    excitation. Ensuring good ride comfort to passengers inside a vehicle subjected
    to undulations from rough roads is an example of this category.

2.2.8.1 Force Transmissibility

Consider a compressor or an engine mounted on a bed experiencing a vibratory force,
which in the simplest form, may be assumed to be harmonic of the form $F \cos \omega t$
(Fig. 2.15). It is to be ensured that these vibrations are not transmitted through the bed
to other nearby machines. Let the mass of the bed be m and let the spring stiffness
and damping coefficient of the elastic mount on which it is supported be k and c
respectively. Let the displacement response of the bed be $X \cos(\omega t - \phi)$. Then we
know that the response X is of the form given in Eq. (2.27).

Dividing numerator and denominator of the response by k to non-dimensionalize
the equation, we get

$$X = \frac{(F/k)}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \tag{2.38}$$

Force transmitted to the ground by the mount is $kx + c\dot{x}$. The magnitude of this force is given by

$$F_T = \frac{X\sqrt{k^2 + (c\omega)^2}}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \qquad (2.39)$$

On non-dimensionalizing this equation and expressing it as the ratio of force transmitted to the force produced, we get

$$\left|\frac{F_T}{F}\right| = \frac{\sqrt{1 + (2\zeta\eta)^2}}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \qquad (2.40)$$

This quantity is known as force transmissibility and its variation with respect to η is shown in Fig. 2.16. Transmissibility is found to be maximum when $\eta = 1$, irrespective of the value of damping ratio, and shoots to infinity when there is no damping. Good vibration isolation, or in other words, reduced transmissibility, i.e., transmissibility less than 0.1, is obtained only for large values of η, typically $\eta > 3$. Considering that one often does not have control over the excitation frequencies, reduction in transmissibility can be achieved by making the natural frequency of the system a fraction of the forcing frequency, at least one-third. Since the forcing frequencies of most machines are known, one can calculate the required low natural frequencies of the combined machine-bed-isolator systems. ω_n can be made small by either increasing the mass of the machine bed or by choosing an isolator of low stiffness. Another very important observation to be made is that in the frequency range of effective vibration isolation, an undamped spring is better than a damped spring as seen from Fig. 2.16, though it is required to have high damping in the vicinity of resonance for reduced response. This may be achieved through the use of vibration stops.

2.2.8.2 Displacement Transmissibility

The case of displacement isolation will now be considered. This is shown in Fig. 2.17. Let a system be subjected to a displacement excitation represented by $y = Y \cos \omega t$. Let m be the mass of the system and let the stiffness and damping of the isolators be k and c, respectively. Let the response of the system at any instant of time be x. Let the relative displacement be $z = (x - y)$. Referring to Fig. 2.17, the equation of motion of the system is given by

$$m\ddot{x} + c(\dot{x} - \dot{y}) + k(x - y) = 0 \qquad (2.41)$$

Since

$$z = x - y, \; \dot{z} = \dot{x} - \dot{y}, \; \ddot{z} = \ddot{x} - \ddot{y} \qquad (2.42)$$

Substituting Eq. (2.42) in (2.41) we have

$$m\ddot{z} + c\dot{z} + kz = m\omega^2 Y \cos \omega t \qquad (2.43)$$

(a)

(b)
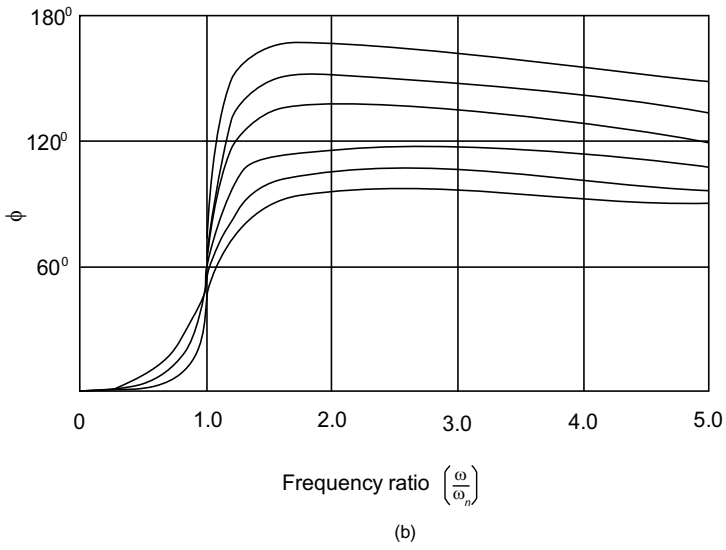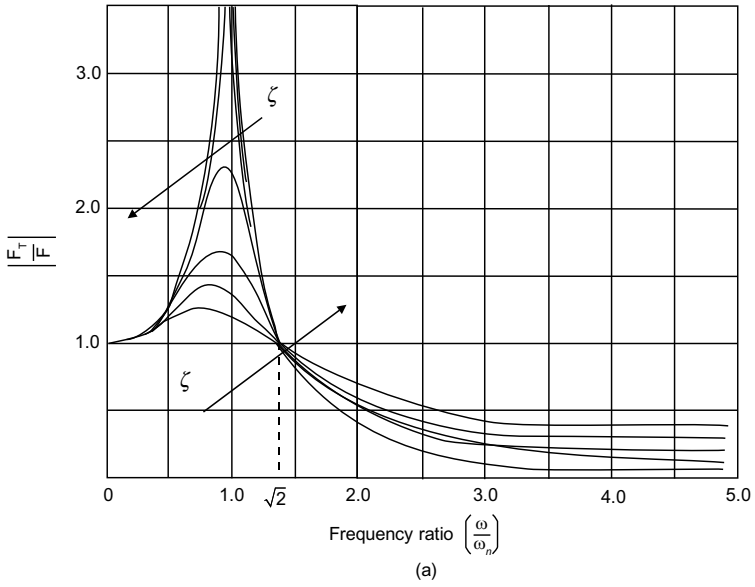
Fig. 2.16 Transmissibility as a function of frequency ratio: **a** magnitude, **b** phase angle

Fig. 2.17 Displacement
transmissibility
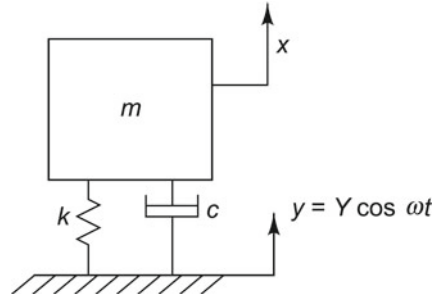
This is the equation for a forced vibration problem. Hence, the solution for z is given
by $z = Z \cos(\omega t - \phi)$ and

$$\left| \frac{Z}{(m\omega^2 Y/k)} \right| = \frac{1}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \qquad (2.44)$$

Rewriting,

$$\left| \frac{Z}{Y} \right| = \frac{\eta^2}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \qquad (2.45)$$

The plot for $|F_T/F|$ in Fig. 2.16 as a function of η, for various values of ζ holds
good for $|Z/Y|$ also. All the salient points put forth for force transmissibility are valid
here also. A typical example of displacement isolation is the case of ride comfort in
automobiles which have road excitation frequencies anywhere up to 50 Hz depending
on the vehicle speed. Their lowest natural frequency is designed to be around 1 Hz
to ensure good ride comfort.

Wherever the exciting force is purely in the vertical direction, one can conveniently
use non-metallic isolators like cork in compression, besides helical springs. If the
disturbance is emanating from a rotating unbalance with axis horizontal, the exciting
force has components of equal intensity in the vertical direction and in a horizontal
direction, both at right angles to the axis of rotation. For purely vertical exciting
forces, helical springs with axis vertical and cork pads in compression are the ideal
choice. For horizontal exciting forces, rubber in shear in the horizontal plane or
inclined rubber isolators can be used. Rubber bushings are well suited for torsional
and axial excitation.

2.3 Two-Degree-of-Freedom System

INTERESTING FACTS: A 2-degree-of-freedom sys-
tem requires 2 independent coordinates to describe its
motion and is the simplest of N degree-of-freedom
systems. These could be translational or rotational.

A typical example of a 2DOF system is the combined system consisting of a
vibratory SDOF system and an attached vibration absorber. The tuned mass
damper (TMD) concept was first proposed by Frahm in 1909 to reduce the
rolling motion of ships as well as ship hull vibrations. A theory for the TMD
was later presented by Ormondroyd and Den Hartog in their paper in 1928.
They showed that by adding damping to Frahm's absorber, its performance
could be significantly improved. Subsequently Den Hartog in his book on
mechanical vibrations in 1940 gave a detailed discussion of optimal tuning
and damping. R. Sarazin (1937) and R. Chilton (1938) are credited with the
invention of the centrifugal pendulum absorber. Typical examples of multi-
degree-of-freedom systems modelled as 2DOF systems are: (i) a two-wheeler
or 4-wheeler vehicle with translational DOF corresponding to the sprung and
unsprung masses, or a translational motion corresponding to bounce and a
rotational motion corresponding to pitching of the vehicle, (ii) a torsional
system consisting of two disks on a shaft, (iii) an aeroplane wing with bounce
and pitch DOF, (iv) a double pendulum with lateral translational motions of
the two masses, and (v) a shaft-coupled motor-generator system.

2.3.1 Translatory System

Consider a 2DOF undamped vibratory system consisting of two springs and two
masses as shown in Fig. 2.18. The masses could very well be the sprung and unsprung
masses of a vehicle and the stiffness those of the suspensions and tyres. Let x_1 and
x_2 denote the vertical displacements measured from the static equilibrium positions
of the two masses m_1 and m_2, respectively.

   If the system is given some arbitrary initial conditions, say, an initial displacement
to mass m_2, the system will perform free vibrations which are periodic in nature.
A harmonic analysis of the motion of masses m_1 and m_2 shows that it is made up
of vibrations of two frequency components. Using Newton's laws, the differential
equations of motion may be written as

$$m_1\ddot{x}_1 + (k_1 + k_2)x_1 - k_2x_2 = 0 \qquad (2.46a)$$
$$m_2\ddot{x}_2 + k_2x_2 - k_2x_1 = 0 \qquad (2.46b)$$

Fig. 2.18 2DOF model
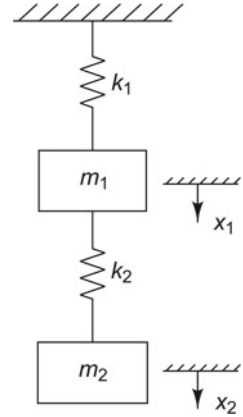
These are coupled equations in two unknowns x_1 and x_2. Equation (2.46) can be put in matrix form as

$$[M]\{\ddot{x}\} + [K]\{x\} = \{0\} \tag{2.47}$$

where

$$[M] = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \qquad [K] = \begin{bmatrix} (k_1 + k_2) & -k_2 \\ -k_2 & k_2 \end{bmatrix} \tag{2.48}$$

The solution may be assumed to be harmonic of the form

$$\begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix} \exp(i\omega t) \tag{2.49}$$

Substituting this solution in Eq. (2.47) we get

$$\begin{bmatrix} (k_1 + k_2 - m_1\omega^2) & -k_2 \\ -k_2 & (k_2 - m_2\omega^2) \end{bmatrix} \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \tag{2.50a}$$

i.e.,

$$\begin{vmatrix} (k_1 + k_2 - m_1\omega^2) & -k_2 \\ -k_2 & (k_2 - m_2\omega^2) \end{vmatrix} = 0 \tag{2.50b}$$

The solution to Eq. (2.50) assumes harmonic motion for both the masses m_1 and m_2 at frequency ω, but with different amplitudes, X_1 and X_2. For a normal mode of vibration, each mass undergoes harmonic motion at that natural frequency, and passes through the equilibrium position simultaneously. For a non-trivial solution for Eq. (2.50),

$$m_1 m_2 \omega^4 - \omega^2[(k_1 + k_2)m_2 + k_2 m_1] + k_1 k_2 = 0 \tag{2.51}$$

This results in a quadratic equation in ω^2, the roots of which are

$$\omega_{1,2}{}^2 = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \tag{2.52}$$

where

$$
\begin{aligned}
A &= m_1 m_2 \\
B &= -[(k_1 + k_2)m_2 + k_2 m_1] \\
C &= k_1 k_2
\end{aligned}
\tag{2.53}
$$

Strictly there are four roots for ω, i.e., $\pm\omega_1$ and $\pm\omega_2$ and the system will perform harmonic motion when the frequency ω in Eq. (2.51) has one of the values given by Eq. (2.52). Hence the complete solution has four vector constants and is as shown in Eq. (2.54).

$$\begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \{\phi_1\}e^{+i\omega_1 t} + \{\phi_2\}e^{-i\omega_1 t} + \{\phi_3\}e^{+i\omega_2 t} + \{\phi_4\}e^{-i\omega_2 t} \tag{2.54}$$

When $\omega = \omega_1$, the first of Eq. (2.54) becomes

$$(k_1 + k_2 - m_1\omega_1{}^2)X_1 - k_2 X_2 = 0 \tag{2.55}$$

Therefore, when $\omega = \omega_1$, X_1 and X_2 have the following ratio

$$\left.\frac{X_2}{X_1}\right|_{\omega=\omega_1} = \frac{k_1 + k_2 - m_1\omega_1^2}{k_2} \tag{2.56}$$

If X_1 is assigned an arbitrary value of unity,

$$X_2 = \frac{(k_1 + k_2 - m_1\omega_1{}^2)}{k_2} \tag{2.57}$$

The vector (X_1, X_2) for the first mode can be written as

$$\begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(1)} = \begin{Bmatrix} 1 \\ (k_1 + k_2 - m_1\omega_1^2)/k_2 \end{Bmatrix} \tag{2.58}$$

Similarly when $\omega = \omega_2$, we can get a similar equation relating X_1 and X_2 from the first of Eq. (2.54) as

$$\left.\frac{X_2}{X_1}\right|_{\omega=\omega_2} = \left\{\frac{(k_1 + k_2 - m_1\omega_2^2)}{k_2}\right\} \tag{2.59}$$

Arbitrarily assigning a value of unity to X_1 as before, we get the vector of displacements (X_1, X_2) for the second mode as

$$\begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(2)} = \begin{Bmatrix} 1 \\ (k_1 + k_2 - m_1\omega_2^2)/k_2 \end{Bmatrix} \tag{2.60}$$

The two vectors described in Eqs. (2.58) and (2.60) give the displacement configurations of the system when it is performing purely harmonic motion and are called the normal modes of the system. The solution in Eq. (2.49) may now be expressed as

$$\begin{Bmatrix} x_1(t) \\ x_2(t) \end{Bmatrix} = \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(1)} (C_1 e^{i\omega_1 t} + C_2 e^{-i\omega_1 t}) + \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(2)} (C_3 e^{i\omega_2 t} + C_4 e^{-i\omega_2 t}) \tag{2.61}$$

where the constants C_i, $i = 1, 2, 3, 4$ are obtained by substituting the initial conditions of the motion. Equation (2.61) can be expressed in the alternate form

$$\begin{aligned} \begin{Bmatrix} x_1(t) \\ x_2(t) \end{Bmatrix} &= \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(1)} (D_1 \cos \omega_1 t + D_2 \sin \omega_1 t) \\ &+ \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(2)} (D_3 \cos \omega_2 t + D_4 \sin \omega_2 t) \end{aligned} \tag{2.62}$$

where

$$\begin{aligned} D_1 &= (C_1 + C_2), \quad D_2 = i(C_1 - C_2) \\ D_3 &= (C_3 + C_4), \quad D_4 = i(C_3 - C_4) \end{aligned} \tag{2.63}$$

Assuming the initial displacements to be $x_1(0)$ and $x_2(0)$ for masses m_1 and m_2 and the initial velocities $\dot{x}_1(0)$ and $\dot{x}_2(0)$ to be zero, we get

$$\begin{Bmatrix} x_1(0) \\ x_2(0) \end{Bmatrix} = \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(1)} D_1 + \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(2)} D_3 \tag{2.64a}$$

$$\begin{Bmatrix} 0 \\ 0 \end{Bmatrix} = \omega_1 \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(1)} D_2 + \omega_2 \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(2)} D_4 \tag{2.64b}$$

From Eq. (2.64) we get

$$D_1 = \frac{x_1(0) X_2^{(2)} - x_2(0) X_1^{(2)}}{X_1^{(1)} X_2^{(2)} - X_2^{(1)} X_1^{(2)}} \tag{2.65a}$$

$$D_3 = \frac{X_1^{(1)} x_2(0) - X_2^{(1)} x_1(0)}{X_1^{(1)} X_2^{(2)} - X_2^{(1)} X_1^{(2)}} \tag{2.65b}$$

and $D_2 = D_4 = 0$. Hence the solution in Eq. (2.62) can be written as

$$\begin{Bmatrix} x_1(t) \\ x_2(t) \end{Bmatrix} = \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(1)} D_1 \cos\omega_1 t + \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(2)} D_2 \cos\omega_2 t \qquad (2.66)$$

It can be seen from Eq. (2.65) that if

$$\begin{Bmatrix} x_1(0) \\ x_2(0) \end{Bmatrix} = \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(1)} \qquad (2.67)$$

then $D_1 = 1$, $D_3 = 0$ and the complete motion is in the first mode and is harmonic with frequency ω_1. If on the other hand

$$\begin{Bmatrix} x_1(0) \\ x_2(0) \end{Bmatrix} = \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(2)} \qquad (2.68)$$

then $D_1 = 0$, $D_3 = 1$, and the motion is in the second mode and is harmonic with frequency ω_2. For arbitrary initial conditions, it is evident from Eq. (2.66) that the motion will be the superposition of two harmonic motions at frequencies ω_1 and ω_2 with both the modes participating.

2.3.2 Coupled Translation and Rotation

The differential equations of motion for the 2DOF system in general are coupled, in that both coordinates appear in each equation. In this section, equations corresponding to a system with coupled translation and rotation (as in pitch and bounce modes) are considered. The equations can be expressed in matrix form and the type of coupling present is evident from them. Mass or dynamical coupling exists if the mass matrix is not diagonal, whereas stiffness or elastic coupling exists if the stiffness matrix is non-diagonal.

Consider the undamped 2DOF system as shown in Fig. 2.19a with coordinates x and θ, where x is the linear displacement of the centre of mass G, J is its mass moment of inertia and θ is the rotation of the bar.

With the translational and rotational coordinates described through the centre of gravity (Fig. 2.19a and b), the equations have coordinate coupling and are of the form

$$\begin{bmatrix} m & 0 \\ 0 & J \end{bmatrix} \begin{Bmatrix} \ddot{x} \\ \ddot{\theta} \end{Bmatrix} + \begin{bmatrix} (k_1 + k_2) & (k_2 l_2 - k_1 l_1) \\ (k_2 l_2 - k_1 l_1) & (k_1 l_1^2 + k_2 l_2^2) \end{bmatrix} \begin{Bmatrix} x \\ \theta \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \qquad (2.69)$$

Stiffness or static coupling exists in this case and the stiffness matrix is non-diagonal. Assuming harmonic motion, the equations of motion of the system can be written as
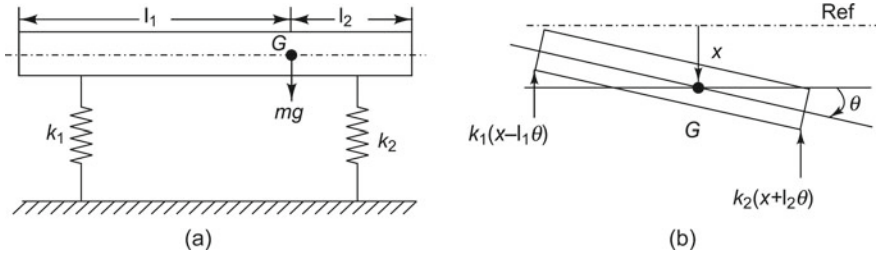
Fig. 2.19 2DOF system with coordinate coupling: **a** 2DOF system, **b** static coupling

$$\begin{bmatrix} (k_1 + k_2 - \omega^2 m) & -(k_1 l_1 - k_2 l_2) \\ -(k_1 l_1 - k_2 l_2) & (k_1 l_1^2 + k_2 l_2^2 - \omega^2 J_c) \end{bmatrix} \begin{Bmatrix} X \\ \Theta \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \tag{2.70}$$

2.3.3 Undamped Vibration Absorber

In applications where it is not possible to reduce the amplitude of vibration of an SDOF system of mass m_1 and stiffness k_1 by adding damping or by redesigning the system so as to change its natural frequency, a vibration absorber is used. This is nothing but another spring-mass system with stiffness k_2 and mass m_2 chosen judiciously and added to the original system to suppress its vibration. The combined system now takes on the form of a 2DOF system as shown in Fig. 2.18. If the absorber is tuned such that its natural frequency coincides with the exciting frequency, i.e., $\omega^2 = k_2/m_2$, the amplitude of vibration of the main mass m_1 is reduced to zero. Making the substitution,

$$\omega_{11}^2 = \frac{k_1}{m_1}; \qquad \omega_{22}^2 = \frac{k_2}{m_2} \tag{2.71}$$

Assuming a harmonic excitation force as we have done before, the amplitude X_1 of the main mass m_1 takes the form

$$X_1 = \frac{(k_2 - \omega^2 m_2)F}{(k_1 + k_2 - \omega^2 m_1)(k_2 - \omega^2 m_2) - k_2^2} \tag{2.72a}$$

The amplitude of the absorber is

$$X_2 = \frac{k_2 F}{(k_1 + k_2 - \omega^2 m_1)(k_2 - \omega^2 m_2) - k_2^2} \tag{2.72b}$$

Or in other words,

$$\frac{X_1 k_1}{F} = \frac{\left[1 - \left(\frac{\omega}{\omega_{22}}\right)^2\right]}{\left(1 + \frac{k_2}{k_1} - \left(\frac{\omega}{\omega_{11}}\right)^2\right)\left(1 - \left(\frac{\omega}{\omega_{22}}\right)^2\right) - k_2^2} \qquad (2.72c)$$

Hence, the design philosophy behind a vibration absorber is to keep $\omega_{22} = \omega_{11}$ or $k_2/m_2 = k_1/m_1$. When this condition is satisfied, the amplitude of the absorber mass will be $X_2 = -F/k_2$.

2.4 Multi-degree-of-Freedom System

DID YOU KNOW that the first devices to solve differential equations were built only in the 1800s? In 1836, the French physicist Gaspard Gustave Coriolis designed a mechanical device for integration of differential equations of the first order. James Thomson in 1876 published the first description of a device called "integrating machine" which could integrate differential equations of any order. His description of the device, along with two further descriptions by his younger brother, Lord Kelvin, led to the invention of the differential analyzer. Thomson's integrating machine was later developed by Arthur Pollen into an electrically driven, mechanical analogue computer, which was completed around 1912. In 1914 Italian mathematician Ernesto Pascal developed so-called integraphs for mechanical integration. The first practical general-purpose differential analyzer was made by Harold Locke Hazen and Vannevar Bush in the United States during the period 1928–1931. It consisted of six wheel-and-disk mechanical integrators and so could solve two, third-order differential equations simultaneously; this work also brought about the development of precise servomechanisms. During the 1930s and early 1940s similar machines were developed in a number of countries involved in World War 2 for military applications. This war brought about the development of the operational amplifier and other electronic devices used in computers. The Electronic Numerical Integrator and Computer (ENIAC) was the first electronic general-purpose computer which was brought out in 1945; it was able to solve "a large class of numerical problems". Today we have sophisticated digital computers and algorithms which do the job.

A system requiring more than one coordinate to describe its motion is called a multi-degree-of-freedom system. Most practical engineering structures are continuous, have infinite degrees of freedom, and give rise to inhomogeneous differential equations. However, we often approximate them to systems with only a finite number of degrees of freedom. The size of the model chosen depends on the frequency range of interest in vibration analysis (dictated by the frequency range of excitation)
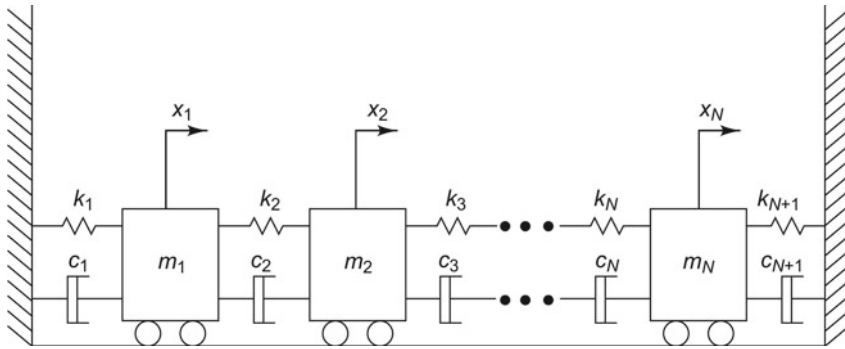
Fig. 2.20 NDOF vibratory model

and the vibration modes that are of significance in this frequency range. The effectiveness of the analysis depends on the assumptions made in the model developed and on the number and type of degrees-of-freedom chosen. An NDOF system has N natural frequencies and N normal modes or in other words N eigenvalues and N eigenvectors which are obtained from the solution of N simultaneous equations of motion of the system. Figure 2.20 represents an NDOF viscously damped system consisting of mass, stiffness, and damping elements. Such a model is typically obtained by lumping the properties of a continuous structure or from a finite element (FE) discretization of the structure. All concepts applicable to SDOF systems are applicable to MDOF systems also. For systems with a large number of degrees of freedom, matrix methods are essential.

2.4.1 Free Vibration of Undamped Multi-degree-of-Freedom System

Let us first consider the undamped case. Normal mode vibrations depend only on the masses and stiffness of the system and their distribution. When a system vibrates at any of its normal modes, all points in the system undergo simple harmonic motion and pass through their equilibrium positions simultaneously. A system must be given specific initial conditions corresponding to its normal mode to excite it. With general initial conditions such as an impulse, the vibrating system may vibrate at all normal modes simultaneously. The N equations of motion for the undamped system may be represented as

$$m_1\ddot{x}_1 + (k_1 + k_2)x_1 - k_2x_2 = 0 \tag{2.73}$$
$$m_2\ddot{x}_2 - k_2x_1 + (k_2 + k_3)x_2 - k_3x_3 = 0$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$\begin{array}{ccccc} . & . & . & . & \\ . & . & . & . & \end{array}$$

$$m_N \ddot{x}_N - k_N x_{N-1} + (k_N + k_{N+1})x_N = 0$$

They can be rewritten in matrix form as

$$[M]\{\ddot{x}\} + [K]\{x\} = 0 \tag{2.74}$$

Here $[M]$ is the $N \times N$ mass matrix and $[K]$ is the $N \times N$ stiffness matrix. $\{x\}$ is the $N \times 1$ coordinate displacement vector.

The solution is assumed to be harmonic as

$$\{x\} = \{X\}\exp(i\omega t) \tag{2.75}$$

The solution satisfies the differential equation if

$$[K - \omega^2 M]\{x\} = \{0\} \tag{2.76}$$

Equation (2.76) is a system of N homogeneous simultaneous equations and for the existence of a non-trivial solution we must have

$$|K - \omega^2 M| = 0 \tag{2.77}$$

This is a polynomial in ω^2 of order N and there will be exactly N roots of ω^2, for which Eq. (2.77) is satisfied. If the matrices $[K]$ and $[M]$ are positive definite (or semi-definite) which is satisfied if the matrices are symmetric, then real solutions exist for ω^2 given by $\omega_1^2, \omega_2^2, \ldots, \omega_N^2$, which are the squares of the natural frequencies of the system. These solutions are called the eigenvalues. Substituting any one of these natural frequencies into Eq. (2.76), we can obtain the system configuration at that natural frequency which is a vector solution given by $\{\phi_k, k = 1, 2, \ldots, N\}$ and which represents the mode shapes or eigenvectors of the system. There are N such natural mode shapes for the structure corresponding to the N natural frequencies. It is clear from the type of solution assumed in Eq. (2.75) that the structure can vibrate harmonically at each of these N natural frequencies, with the structural configuration at those frequencies defined by the corresponding mode shapes.

2.4.2 Free Vibration of Multi-degree-of-Freedom System with Damping

Equation (2.78) represents a set of N coupled second-order differential equations for an MDOF system with viscous damping.

$$m_1\ddot{x}_1 + (c_1 + c_2)\dot{x}_1 - c_2\dot{x}_2 + (k_1 + k_2)x_1 - k_2x_2 = 0$$
$$m_2\ddot{x}_2 - c_2\dot{x}_1 + (c_2 + c_3)\dot{x}_2 - c_3\dot{x}_3 - k_2x_1 + (k_2 + k_3)x_2 - k_3x_3 = 0$$

$$m_N\ddot{x}_N - c_N\dot{x}_{N-1} + (c_N + c_{N+1})\dot{x}_N - k_Nx_{N-1} + (k_N + k_{N+1})x_N = 0 \quad (2.78)$$

It can be rewritten in matrix form as follows:

$$[M]\{\ddot{x}\} + [C]\{\dot{x}\} + [K]\{x\} = \{0\} \tag{2.79}$$

Here the matrices $[M]$, $[K]$, and $[C]$ are the mass, stiffness, and damping matrices, respectively, describing the spatial properties of the discretized structure. Hence this model is called the spatial model. $\{\ddot{x}\}$, $\{\dot{x}\}$ and $\{x\}$ correspond to the acceleration, velocity and displacement vectors. Equation (2.79) can be solved if two initial conditions of motion are specified for each of the N coordinates.

2.4.3 Orthogonality of Natural Modes

The mode shape vectors satisfying the eigenvalue problem possess "orthogonality" properties which are due to the symmetry of the $[K]$ and $[M]$ matrices. The ith natural mode corresponding to the ith natural frequency is a column vector with N elements, each element representing the displacement of the corresponding DOF in that mode. Each of the mode shapes satisfies the equations of motion individually. Hence, for the ith and jth modes, we have

$$-\omega_i^2[M]\{\phi_i\} + [K]\{\phi_i\} = \{0\} \tag{2.80a}$$
$$-\omega_j^2[M]\{\phi_j\} + [K]\{\phi_j\} = \{0\} \tag{2.80b}$$

Pre-multiplying Eq. (2.80a) by ϕ_T^j and Eq. (2.80b) by ϕ_i^T and writing the resulting expression without matrix brackets, we have

$$\omega_i^2\phi_j^T M\phi_i = \phi_i^T K\phi_i \tag{2.81a}$$
$$\omega_j^2\phi_i^T M\phi_j = \phi_i^T K\phi_j \tag{2.81b}$$

Taking transpose of both sides of Eq. (2.81b), we obtain

$$\omega_j^2\phi_j^T M^T \phi_i = \phi_j^T K^T \phi_i \tag{2.82}$$

For most practical structures, the mass and stiffness matrices are symmetric and therefore $M^T = M$, $K^T = K$. Using this in Eq. (2.82) we obtain

$$\omega_j^2 \phi_j^T M \phi_i = \phi_j^T K \phi_i \tag{2.83}$$

Subtracting Eq. (2.83) from (2.81a), we get

$$(\omega_i^2 - \omega_j^2) \phi_j^T M \phi_i = 0 \tag{2.84}$$

However, $\omega_i \neq \omega_j$; therefore

$$\phi_j^T M \phi_i = 0 \quad \text{for} \quad i \neq j \tag{2.85a}$$

Substituting Eq. (2.85a) in Eq. (2.81a), it follows that

$$\phi_j^T K \phi_i = 0 \quad \text{for} \quad i \neq j \tag{2.85b}$$

However, if $i = j$, then

$$\begin{aligned} \phi_i^T M \phi_i &= \mu_i \\ \phi_i^T K \phi_i &= \kappa_i \end{aligned} \tag{2.86a}$$

$$\frac{\phi_i^T K \phi_i}{\phi_i^T M \phi_i} = \frac{\kappa_i}{\mu_i} = \omega_i^2 \tag{2.86b}$$

Equations (2.85) and (2.86) define the orthogonality conditions for normal modes. μ_i and κ_i are called the generalized or modal mass and stiffness corresponding to the ith mode.

2.4.4 Free Vibration of Multi-degree-of-Freedom System in Terms of Modal Coordinates

The free vibration response of an MDOF system can be expressed in terms of the normal modes as mentioned earlier. It is to be noted that the absolute magnitudes of the eigenvectors of the ith mode ϕ_i, $i = 1, 2, \ldots, N$ are not known, but their directions are. Since the absolute magnitudes are not known and depend on the initial conditions, one can normalize the mode shape vectors with respect to the modal masses and these "normal mode shapes" are then called the mass-normalized or weighted eigenvectors, ψ_i, $i = 1, 2, \ldots, N$.

$$\psi_i = \frac{\phi_i}{\sqrt{\mu_i}}, \quad i = 1, 2, \ldots, N \tag{2.87}$$

A modal coordinate is associated with every mode. The physical coordinate $\{x\}$ is related to the modal coordinate $\{q\}$ by the equation

$$\{x(t)\} = [\phi]\{q(t)\} \tag{2.88}$$

Here $[\phi]$ is the $N \times N$ matrix called the raw modal matrix and is obtained by arranging all the eigenvectors side by side. The vector corresponding to the ith column of this matrix is the modal vector corresponding to the ith mode. Substituting for $\{x\}$ from Eq. (2.88) into Eq. (2.74) and omitting the brackets, we get

$$M\phi\ddot{q} + K\phi q = 0 \tag{2.89}$$

Pre-multiplying Eq. (2.89) by ϕ^T we get

$$\phi^T M\phi\ddot{q} + \phi^T K\phi q = 0 \tag{2.90}$$

Using the orthogonality conditions for normal modes, Eq. (2.90) becomes

$$[\mu]\{\ddot{q}\} + [\kappa]\{q\} = \{0\} \tag{2.91}$$

Here μ and κ are diagonal matrices containing generalized masses and stiffness. Equation (2.91) constitutes a set of N uncoupled equations of the form

$$\mu_i \ddot{q}_i + \kappa_i q_i = 0, \qquad i = 1, 2, \ldots, N \tag{2.92}$$

Equation (2.91) simply corresponds to N SDOF systems as μ and κ are diagonal matrices. Therefore, one can solve for $\{q(t)\}$ using SDOF theory and then use the transformation relation in Eq. (2.88) to calculate the physical motion vector $\{x(t)\}$. The response coordinates $\{q(t)\}$ are called the principal coordinates or normal coordinates.

Instead of the transformation used in Eq. (2.88), if we use one involving the mass-normalized eigenvectors of the form

$$\{x(t)\} = [\psi]\{q(t)\} \tag{2.93}$$

then Eq. (2.92) takes on the form

$$\ddot{q}_i + \omega_i^2 q_i = 0, \qquad i = 1, 2, \ldots, N \tag{2.94}$$

Here again the equation corresponds to N SDOF systems. If we include the effect of the damping matrix however, as in Eq. (2.79), we will find that the transformations of the form given in Eqs. (2.88) and (2.93) do not diagonalize the equations of motion. However, a form of viscous damping, known as proportional or Rayleigh damping results in diagonalization of the damping matrix also. Rayleigh damping assumes that the damping matrix, $[C]$ is given by $\alpha[M] + \beta[K]$, where α and β

are proportionality constants. Then the normal mode matrix $[\psi]$, which diagonal-ized the equations of motions along with the earlier transformation represented by Eq. (2.93), can now diagonalize the damping matrix $[C]$ also, to yield N uncoupled SDOF damped equations as shown below

$$\ddot{q}_i + 2\zeta_i\omega_i\dot{q}_i + \omega_i^2 q_i, \qquad i = 1, 2, \ldots, N \tag{2.95}$$

Here $2\zeta_i\omega_i$ is called the modal damping and is defined by the equation

$$2\zeta_i\omega_i = \alpha + \beta\omega_i^2 \tag{2.96}$$

It is worthwhile knowing that the values of α and β may be obtained, if experimental values of ζ_i and ω_i are got for any two modes of the system through experiments. The solution corresponding to each of Eqs. (2.91) can be obtained by plugging in the initial conditions $\{x(0)\}$ and $\{\dot{x}(0)\}$ in the coordinates $\{q\}$ which are obtained from the transformation Eq. (2.88).

$$\begin{aligned} \{q(0)\} &= [\phi]^{-1}\{x(0)\} \\ \{\dot{q}(0)\} &= [\phi]^{-1}\{\dot{x}(0)\} \end{aligned} \tag{2.97}$$

The solution to Eq. (2.91) is then

$$q_i(t) = q_i(0)\cos\omega_i t + \frac{\dot{q}_i(0)}{\omega_i}\sin\omega_i t \tag{2.98}$$

Here ω_i is the ith natural frequency which is related to the modal mass and stiff-ness matrices as described in Eq. (2.86). After solving for q_i, $i = 1, 2, 3, \ldots, N$, the response in coordinates $\{x(t)\}$ is obtained using Eq. (2.88).

2.4.5 Forced Response to Harmonic Excitation: Frequency Response Function (FRF)

Once the free vibration response is known, the next step is determination of the steady-state response of the structure to sinusoidal excitation at different frequencies. This is called its frequency response function. Knowledge of the mass, stiffness and damping properties of the structure can be used to obtain the FRF or transfer function. The response of the MDOF system is derived for the proportional viscous damping case alone. The governing equations are obtained based on the assumption that each mass is acted upon by a force $f_i(t), i = 1, \ldots, N$. The equations of motion with forced excitation can be written as

$$[M]\{\ddot{x}\} + [C]\{\dot{x}\} + [K]\{x\} = \{f(t)\} \tag{2.99}$$

The eigenvalues and normalized eigenvectors obtained by solving the homogeneous eigenvalue Eq. (2.74), ignoring the damping term are ω_i and ψ_i, $i = 1, 2, 3, \ldots, N$. Using the transformation in Eq. (2.93) and substituting in Eq. (2.99) we get

$$[M][\psi]\{\ddot{q}\} + [C][\psi]\{\dot{q}\} + [K][\psi]\{q\} = \{\tilde{f}(t)\} \tag{2.100}$$

Pre-multiplying both sides of Eq. (2.100) by ψ^T and using the principle of orthogonality of the eigenvectors, we get the equation corresponding to the jth coordinate, q_j as

$$\{\ddot{q}_j\} + [2\zeta_j\omega_j]\{\dot{q}_j\} + [\omega_j^2]\{q_j\} = \{\sigma_j(t)\} \tag{2.101}$$

where $\{\sigma\} = [\psi]^T\{\tilde{f}(t)\}$ is the generalized force vector.

Equation (2.101) constitutes N number of uncoupled equations of motion corresponding to N SDOF systems. The complete solution in the jth coordinate, q_j is

$$q_j = e^{-\zeta_j\omega_j t}(A_j \cos \omega_{d_j} t + B_j \sin \omega_{d_j} t) + q_{js}(t) \tag{2.102}$$

where

$$\omega_{d_j} = \omega_j(1 - \zeta_j^2)^{1/2} \tag{2.103}$$

and $q_{js}(t)$ is the steady-state solution in the jth coordinate and can be readily obtained when the form of forcing function $\sigma_j(t)$ is known. If $\sigma_j(t)$ is harmonic and is expressible as

$$\sigma_j(t) = \sigma_{jo} \cos \omega t \tag{2.104}$$

then it can easily be shown that $q_{js}(t)$ can be expressed as

$$q_{js}(t) = \frac{\sigma_{jo} \cos \omega t}{\omega_j^2 - \omega^2 + i2\zeta_j\omega_j\omega}, \quad j = 1, 2, \ldots, N \tag{2.105}$$

The amplitude of response in the jth coordinate to harmonic excitation of unit amplitude in the jth coordinate or in other words the term $H_{jj}(\omega)$ in the frequency response function is given by

$$H_{jj}(\omega) = [\omega_j^2 - \omega^2 + i2\zeta_j\omega_j\omega]^{-1} \tag{2.106}$$

The total steady-state response amplitude in the jth coordinate due to the contribution from all the N modes is given by

$$Q_{js} = \sum_{r=1}^{N} \frac{(_r\psi_j)\sigma_{ro}}{\omega_j^2 - \omega^2 + i2\zeta_j\omega_j\omega}, \quad j = 1, 2, \ldots, N \tag{2.107}$$

Here $(_r\psi_j)$ is the response of the jth coordinate in the rth mode. The total steady-state response in the jth coordinate to a unit harmonic force in the kth coordinate is given by

$$H_{jk}(\omega) = \sum_{r=1}^{N} \frac{(_r\psi_j)(_r\psi_k)}{[\omega_j^2 - \omega^2 + i2\zeta_j\omega_j\omega]} \tag{2.108}$$

$H(\omega)$ is the called the receptance matrix or frequency response function matrix and has all the dynamic characteristics of the system. Each element in the receptance matrix $H_{jk}(\omega)$ describes the relation between the response at location j due to an excitation at location k. The form given in Eq. (2.107) is an efficient way to obtain the response. More details regarding obtaining and interpreting the FRF matrix are discussed in Chap. 9.

2.5 Vibrations of Continuous Systems

INTERESTING FACTS: A beam is a slender horizontal structural member that resists lateral loads applied to it, by bending. Beams are widely used in engineering structures, such as supporting members in high-rise buildings, roof beams for supporting roof trusses, beams in railway tracks, long-span bridges, flexible satellites, aeroplane wings, etc. Beams are also used to construct frames for trolley ways, elevators, trailer and truck beds, platforms, and common building constructions. Studies on thin beams with different boundary conditions were conducted by Leonhard Euler in 1744 and Daniel Bernoulli in 1751. The latter derived the differential equation governing the motion of a vibrating beam, while the former studied the shape of elastic beams under different boundary conditions. Their study has come to be known as the Euler–Bernoulli or thin beam theory. Their beam model tends to overestimate the natural frequencies of a vibrating beam. In 1877, Lord Rayleigh included the effect of rotational inertia of the cross-sectional area, in order to improve the model; the natural frequencies were however still overestimated. The Timoshenko beam model was developed in the early 20th century by Stephen Timoshenko. This model added the shear distortion effects to the Rayleigh model, making it more accurate.

2.5.1 Introduction to Continuous Systems

We have in the earlier sections discussed discrete mass systems modelled as single, two, or multi DOF systems. In these systems, lumped masses, stiffness, and damping elements were assumed to be present only at discrete points. There are systems, on

the other hand, which have continuously distributed mass, elasticity, and damping and are called continuous systems; these systems have infinite degrees of freedom with an infinite number of natural frequencies and corresponding mode shapes. It may be mentioned that the response of the continuous system depends on time and space coordinates, whereas, in the case of discrete systems, the response is a function of time only. Hence, while the equations of motion of discrete systems are written in terms of ordinary differential equations, those of continuous systems are written in terms of partial differential equations. It is important to note that the concept of orthogonality is applicable to both discrete and continuous systems. The response of the system will depend on the boundary conditions. There are two different types of boundary conditions, viz., geometric and natural boundary conditions which have to be satisfied. Geometric or essential boundary conditions are due to conditions of purely geometric compatibility (for example in the case of a clamped–clamped beam, the deflections and slopes are zero at both the ends). Natural boundary conditions, also known as additional or dynamic boundary conditions, result from the balance of moments or forces at the boundary. For example, in the case of a free–free beam which may be a model of a flying aeroplane or a spacecraft, shear force and bending moments are zero at both the ends; hence they constitute the natural boundary conditions. In many systems one will find a mix of both the geometric and natural boundary conditions. For example, in the case of a simply supported beam, both deflection and bending moment at the end points are zero. Hence, the boundary condition in this case may be termed as mixed boundary conditions.

2.5.2 Longitudinal Vibration of a Rod

Consider an elastic rod of length l and uniform cross-sectional area A subjected to an axial force P at time $t = 0$ and then released. It will undergo longitudinal vibration. Figure 2.21 shows a small element of this rod of length dx.

Let $u(x, t)$ be the axial displacement of the element dx of the rod at x. The displacement at $x + dx$ will be $u + (\partial u/\partial x)dx$. From Hooke's law, the ratio of unit stress to unit strain is the modulus of elasticity E. Let p be the axial force acting at x. Hence

$$\frac{p}{A} = E\frac{\partial u}{\partial x} \tag{2.109}$$

Let us now apply Newton's law of motion to the element.

$$dm\frac{\partial^2 u}{\partial t^2} = \left(p + \frac{\partial p}{\partial x}dx\right) - p \tag{2.110a}$$

i.e.,

$$\rho A dx\frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x}\left(AE\frac{\partial u}{\partial x}\right)dx \tag{2.110b}$$
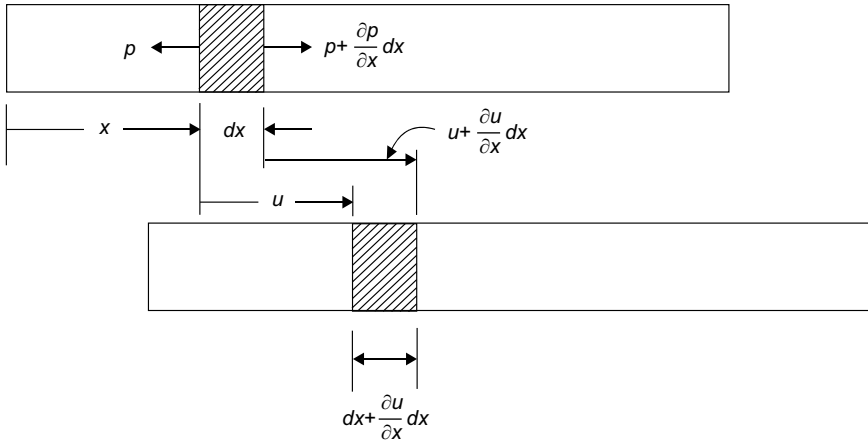
Fig. 2.21 Free body diagram of rod element dx

where $dm = \rho A dx$ is the mass of element dx and ρ is the mass per unit volume.

In Eq. (2.110) if AE is constant, by eliminating $\partial P / \partial x$ between Eqs. (2.109) and (2.110), we get the partial differential equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{AE}{\rho A} \frac{\partial^2 u}{\partial x^2} \tag{2.111}$$

or

$$\frac{\partial^2 u}{\partial t^2} = \frac{E}{\rho} \frac{\partial^2 u}{\partial x^2} \tag{2.112}$$

or

$$\frac{\partial^2 u}{\partial t^2} = c_0^2 \frac{\partial^2 u}{\partial x^2} \quad \text{where } c_0 = \sqrt{\frac{E}{\rho}} \tag{2.113}$$

It may be observed that this corresponds to the wave equation where c_0 represents the velocity of propagation of the displacement or stress wave in the rod. To find the response of the rod, one may use the method of separation of variables as can be shown below:

$$u(x, t) = \phi(x) q(t) \tag{2.114}$$

$\phi(x)$ is known as the mode shape of the system and $q(t)$ represents time modulation. Now Eq. (2.113) reduces to

$$\phi(x) \frac{\partial^2 q}{\partial t^2} = c_0^2 q(t) \frac{\partial^2 \phi}{\partial x^2} \tag{2.115a}$$

or

$$c_0^2 \frac{1}{\phi} \left(\frac{\partial^2 \phi}{\partial x^2} \right) = \frac{1}{q} \left(\frac{\partial^2 q}{\partial t^2} \right) \tag{2.115b}$$

Since the left side of Eq. (2.115b) is independent of time t and the right side is independent of x, the equality is valid for all values of t and x. Therefore, each side must be a constant. Since the right side term is equal to a constant, it implies that the acceleration term $\partial^2 q / \partial t^2$ is proportional to displacement term $q(t)$; the proportionality constant may be assumed equal to $-\omega^2$ to have a simple harmonic motion in the system; a negative constant is taken to ensure that the response will decay exponentially and make the system stable. Hence, one may write Eq. (2.115b) as

$$c_0^2 \frac{1}{\phi} \left(\frac{\partial^2 \phi}{\partial x^2} \right) = \frac{1}{q} \left(\frac{\partial^2 q}{\partial t^2} \right) = -\omega^2 \tag{2.116}$$

Hence,

$$\frac{d^2 q}{dt^2} + \omega^2 q = 0 \tag{2.117}$$

Also

$$\frac{d^2 \phi}{dx^2} + \frac{\omega^2}{c_0^2} \phi = 0 \tag{2.118}$$

The solution of Eqs. (2.117) and (2.118) can be written as

$$q(t) = C_1 \sin \omega t + C_2 \cos \omega t \tag{2.119}$$

$$\phi(x) = A \sin \left(\frac{\omega}{c_0} \right) x + B \cos \left(\frac{\omega}{c_0} \right) x \tag{2.120}$$

Hence,
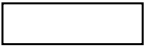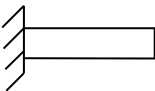
$$u(x, t) = \left(A \sin \left(\frac{\omega}{c_0} \right) x + B \cos \left(\frac{\omega}{c_0} \right) x \right) (C_1 \sin \omega t + C_2 \cos \omega t) \tag{2.121}$$

The coefficients A and B are to be determined from the boundary conditions and the coefficients C_1 and C_2 from the initial conditions. The various boundary conditions for some of the typical cases of longitudinal vibration of a rod of length l are given in Table 2.4.

2.5.3 Transverse Vibration of Euler–Bernoulli Beam

In this section, we will consider the vibration of a simple continuous system, a beam with constant cross-sectional properties. Transverse vibrations cause bending

Table 2.4 Boundary conditions for the longitudinal vibration of rod

Case	Boundary condition: left $x = 0$	Boundary condition: right $x = l$
Free-free	$\dfrac{\partial u(0, t)}{\partial x} = 0$	$\dfrac{\partial u(l, t)}{\partial x} = 0$
Fixed-free	$u(0, t) = 0$	$\dfrac{\partial u(l, t)}{\partial x} = 0$
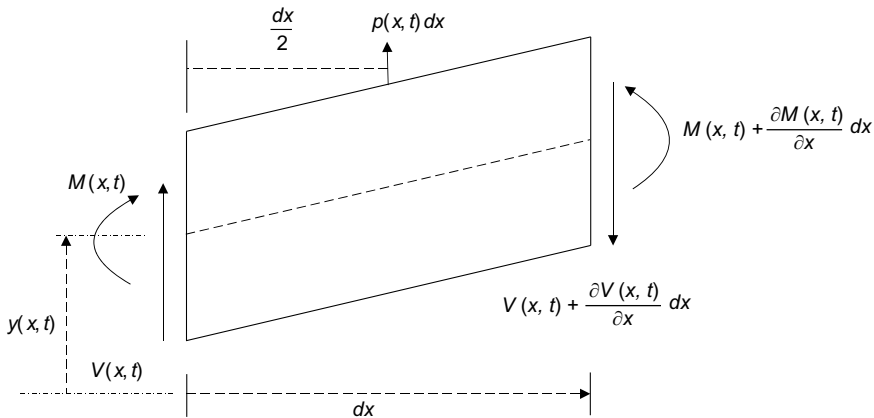
Fig. 2.22 Free body diagram of Euler–Bernoulli beam element

deformation of the beam. When the transverse deflection of a beam is assumed to be due to bending, then the beam is called an Euler–Bernoulli beam. For the present analysis, Euler–Bernoulli beam theory is considered, which assumes that a plane cross section remains plane after bending and remains perpendicular to the neutral axis of the beam before and after bending. In order to obtain the differential equation for the transverse vibration of a beam, we can consider the forces and moments acting on an element of the beam as shown in Fig. 2.22.

Let $y(x, t)$ be the transverse displacement at distance x from the left end at time t. Let the force per unit length in the transverse direction be $p(x, t)$. $V(x, t)$ and $M(x, t)$ are the shear force and bending moment respectively acting on the beam. Summing the forces in the y-direction, we have

$$-\left(V + \frac{\partial V}{\partial x}dx\right) + V + p(x, t)dx = \mu dx \frac{\partial^2 y}{\partial t^2} \tag{2.122}$$

where μ is the mass per unit length of the beam.

From strength of materials we know that

$$V(x, t) = \frac{\partial M(x, t)}{\partial x} \tag{2.123a}$$

Hence

$$\frac{\partial V(x, t)}{\partial x} = \frac{\partial^2 M(x, t)}{\partial x^2} \tag{2.123b}$$

We also know that

$$M(x, t) = EI \frac{\partial^2 y(x, t)}{\partial x^2} \tag{2.124}$$

where E and I are Young's modulus of elasticity and area moment of inertia of the beam cross section.

Substituting Eq. (2.123b) into Eq. (2.122) we have

$$-\frac{\partial^2 M}{\partial x^2} + p(x, t) = \mu \frac{\partial^2 y}{\partial t^2} \tag{2.125}$$

Substituting (2.124) into Eq. (2.125) we get

$$EI \frac{\partial^4 y}{\partial x^4} + \mu \frac{\partial^2 y}{\partial t^2} = p(x, t) \tag{2.126}$$

The above equation is the governing partial differential equation of motion. Setting the external force $p(x, t)$ to zero, the above equation can be written as

$$c^2 \frac{\partial^4 y(x, t)}{\partial x^4} + \frac{\partial^2 y(x, t)}{\partial t^2} = 0 \quad \text{where } c^2 = EI/\mu \tag{2.127}$$

Since the equation of motion involves a fourth-order derivative with respect to x and a second-order derivative with respect to time, four boundary conditions and two initial conditions are required for finding a unique solution $y(x, t)$. The boundary conditions for the transverse vibrations for various configurations of the beam are given in Table 2.5.
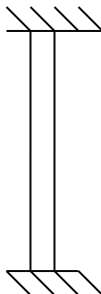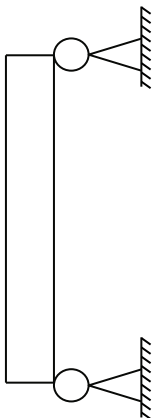
In all the above cases, the equation of motion of the system reduces to that of the wave equation, which can be given by

$$\frac{\partial^2 y}{\partial t^2} + \frac{EI}{\mu} \frac{\partial^4 y}{\partial x^4} = 0 \tag{2.128}$$

To find the response of the system we can use separation of variables as before by using the following equation.

$$y(x, t) = \phi(x) q(t) \tag{2.129}$$

Table 2.5 Boundary conditions for transverse vibration of beam of length l

Case	BCs at left end; $x = 0$	BCs at right end; $x = l$
Clamped–clamped (deflection, slope = 0)	$y(0, t) = 0$; $\dfrac{\partial y(0, t)}{\partial x} = 0$	$y(l, t) = 0$; $\dfrac{\partial y(l, t)}{\partial x} = 0$
Pinned–pinned (deflection, moment = 0)	$y(0, t) = 0$; $\dfrac{\partial^2 y(0, t)}{\partial x^2} = 0$	$y(l, t) = 0$; $\dfrac{\partial^2 y(l, t)}{\partial x^2} = 0$
Sliding–sliding (slope, shear = 0)	$\dfrac{\partial y(0, t)}{\partial x} = 0$; $\dfrac{\partial^3 y(0, t)}{\partial x^3} = 0$	$\dfrac{\partial y(l, t)}{\partial x} = 0$; $\dfrac{\partial^3 y(l, t)}{\partial x^3} = 0$
Free-free (moment, shear = 0)	$\dfrac{\partial^2 y(0, t)}{\partial x^2} = 0$; $\dfrac{\partial^3 y(0, t)}{\partial x^3} = 0$	$\dfrac{\partial^2 y(l, t)}{\partial x^2} = 0$; $\dfrac{\partial^3 y(l, t)}{\partial x^3} = 0$

Now Eq. (2.128) reduces to

$$\phi(x)\frac{\partial^2 q(t)}{\partial t^2} + \frac{EI}{\mu}\frac{\partial^4 \phi}{\partial x^4}q(t) = 0 \tag{2.130a}$$

Or

$$-\frac{EI}{\mu}\left(\frac{1}{\phi(x)}\right)\left(\frac{\partial^4 \phi(x)}{\partial x^4}\right) = \frac{1}{q(t)}\left(\frac{\partial^2 q(t)}{\partial t^2}\right) \tag{2.130b}$$

Since the left side of Eq. (2.130b) is independent of time t and the right side is independent of x, the equality is valid for all values of t and x. Taking the proportionality constant equal to $-\omega^2$ as before, we may write Eq. (2.130) as

$$-\frac{EI}{\mu}\left(\frac{1}{\phi(x)}\right)\left(\frac{\partial^4 \phi(x)}{\partial x^4}\right) = \frac{1}{q(t)}\left(\frac{\partial^2 q(t)}{\partial t^2}\right) = -\omega^2 \tag{2.131}$$

Hence,

$$\left(\frac{d^2 q(t)}{dt^2}\right) + \omega^2 q(t) = 0 \tag{2.132a}$$

Also

$$\frac{\partial^4 \phi(x)}{\partial x^4} - \frac{\mu\omega^2}{EI}\phi(x) = 0 \tag{2.132b}$$

Let

$$\beta^4 = \frac{\mu\omega^2}{EI} \tag{2.133}$$

Equation (2.132b) can be written as

$$\frac{\partial^4 \phi(x)}{\partial x^4} - \beta^4 \phi(x) = 0 \tag{2.134}$$

The solution of Eqs. (2.132a) and (2.132b) can be written as

$$q(t) = C_1 \sin \omega t + C_2 \cos \omega t \tag{2.135a}$$
$$\phi(x) = A \sinh(\beta x) + B \cosh(\beta x) + C \sin(\beta x) + D \cos(\beta x) \tag{2.135b}$$

Table 2.6 $(\beta_n l)^2$ for different boundary conditions of beam

Beam configuration	Fundamental mode $(\beta_1 l)^2$	Second mode $(\beta_2 l)^2$	Third mode $(\beta_3 l)^2$	Fourth mode $(\beta_4 l)^2$
Simply supported	9.87	39.5	88.9	157.91
Cantilever	3.52	22.0	61.7	120.90
Free-free*	22.4	61.7	120.9	199.86
Clamped–clamped	22.4	61.7	120.9	199.86
Clamped-hinged	15.42	49.96	104.25	178.27
Hinged-free*	15.42	49.96	104.25	178.27

$*(\beta l)^2 = 0$ for rigid body mode

Hence,

$$y(x, t) = (A \ \sinh \beta x + B \cosh \beta x + C \sin \beta x + D \cos \beta x)$$
$$\times (C_1 \sin \omega t + C_2 \cos \omega t) \tag{2.136}$$

Constants C_1 and C_2 can be got from the initial conditions and constants A, B, C and D from the boundary conditions. Let us now find the mode shapes of a simply supported beam, fixed-fixed beam, cantilever beam and free–free beams. From Eq. (2.133) one may write the expression for the natural frequency as

$$\omega_n = \beta_n^2 \sqrt{\frac{EI}{\mu}} = (\beta_n l)^2 \sqrt{\frac{EI}{\mu l^4}} \tag{2.137}$$

Here the number β_n depends on the boundary conditions of the problem. Table 2.6 shows the numerical values of $(\beta_n l)^2$ for typical boundary conditions.

Figures 2.23, 2.24 and 2.25 show the first four modes of a simply supported beam, a cantilever beam and a beam fixed at both ends.

Chapter 3 gives a description of commonly used vibration transducers and exciters. Chapter 10 explains some common laboratory experiments in vibration.
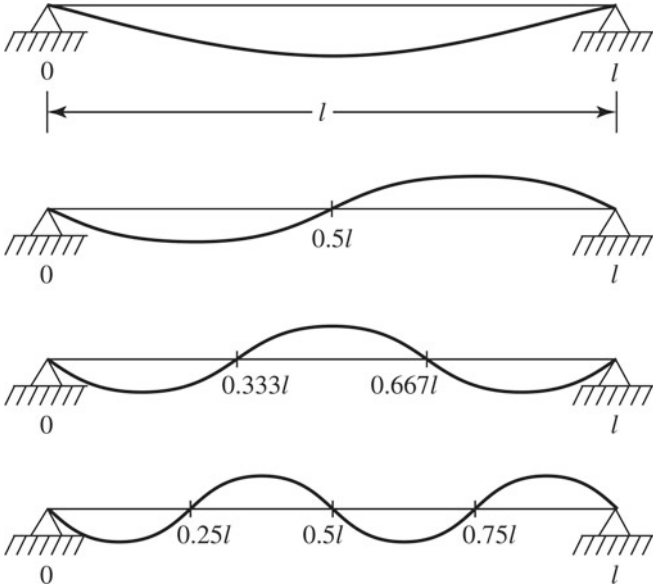
Fig. 2.23   First four modes of simply supported beam
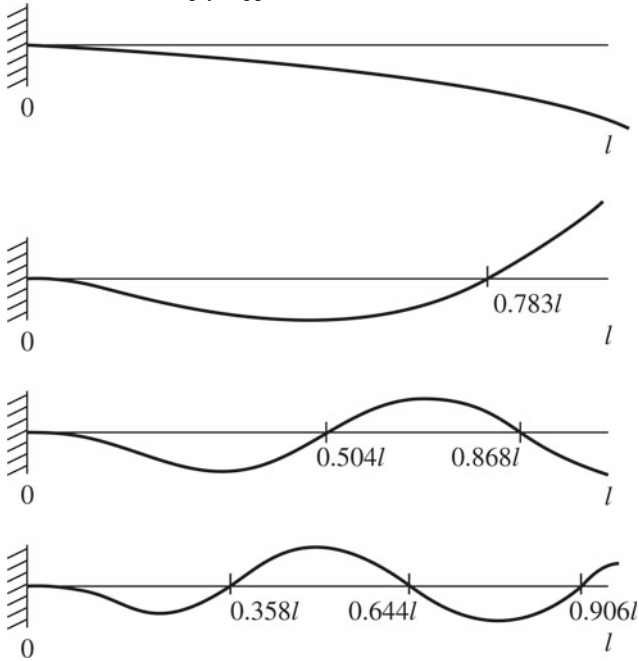
Fig. 2.24   First four modes of cantilever beam
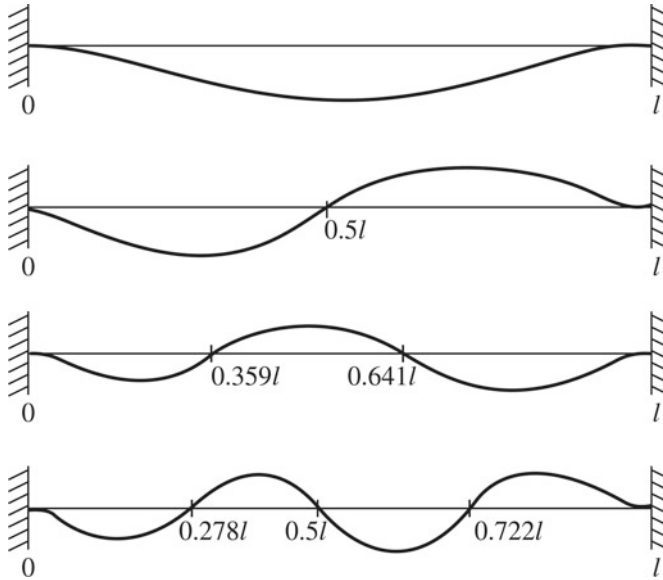
Fig. 2.25 First four modes of beam fixed at both ends

List of Symbols

Symbol	Meaning	SI unit
A	uniform cross-sectional area of rod	m^2
A, A_j	constants	
B, B_j	constants	
c_1, c_2	constants	
c, c_1, c_2, ..., c_{N+1}	viscous damping coefficients	N.s/m
c_c	critical damping coefficient	N.s/m
c_0	velocity of propagation of displacement wave	m/s
C, C_1, C_2, C_3, C_4	constants	
C_{eq}	equivalent damping coefficient	N.s/m
$[C]$	$N \times N$ damping matrix	N.s/m
dm	mass of element dx	kg
dx	length of small element of rod	m
D, D_1, D_2, D_3, D_4	constants	
e	eccentricity	m
E	Young's modulus	Pa
f	frequency	Hz
f	force	N

Symbol	Meaning	SI unit
f_n	natural frequency	Hz
F	amplitude of force $f(t)$	N
F_d	damping force	N
F_T	transmitted force	N
$\{f(t)\}$	$N \times 1$ excitation force vector	N
$\{\tilde{f}(t)\}$	$N \times 1$ transformed excitation force vector	N.kg$^{-1/2}$
g	acceleration due to gravity	m/s^2
h	complex stiffness	N/m
$H(\omega)$	receptance FRF matrix	m/N
$H_{jj}(\omega), H_{jk}(\omega)$	elements of receptance FRF matrix	m/N
i	$\sqrt{-1}$	
I	area moment of inertia	m^4
J	mass moment of inertia	kg.m^2
$k, k_1, \ldots, k_N, k_{N+1}$	stiffness	N/m
$[K]$	$N \times N$ stiffness matrix	N/m
l	length of rod	m
L	length of shaft	m
l_1, l_2	distances of k_1 and k_2 from centre of gravity	m
m	mass of dynamic system	kg
$m_1, m_2, \ldots, m_N$	masses	kg
M	total mass of system	kg
$M(x, t)$	bending moment acting on beam	N.m
$[M]$	$N \times N$ mass matrix	kg
n	cycle number	
N	number of degrees of freedom	
$p(x, t)$	force per unit length in transverse direction	N/m
p	axial force	N
q	torsional stiffness	N.m/rad
q_i	ith principal coordinate	m
$q(t)$	displacement	m
$q_{js}(t)$	steady-state solution in jth coordinate	m
$\{q\}, \{q(t)\}$	$N \times 1$ vector of principal or normal coordinates	m
$\{q(0)\}$	$N \times 1$ vector of principal coordinates at time $t = 0$	m
$\{\dot{q}(0)\}$	$N \times 1$ vector of derivatives of principal coordinates at time $t = 0$	m/s
Q_{js}	total steady-state response amplitude in jth coordinate due to contribution from all N modes	m

Symbol	Meaning	SI unit
t	time	s
T	natural period of oscillation	s
T_d	natural period of oscillation with damping	s
$u(x, t)$	axial displacement of element dx at x	m
$V(x, t)$	shear force acting on beam	N
W	weight of body	N
W_d	energy lost per cycle	J
x	dynamic displacement	m
$x_1, x_2, \ldots, x_N$	dynamic displacements of masses $m_1, m_2, \ldots, m_N$	m
$\dot{x}$	velocity of vibration	m/s
$\ddot{x}$	vibration acceleration	m/s^2
$x_1(0), x_2(0)$	displacements of m_1, m_2 at $t = 0$	m
$\dot{x}_1(0), \dot{x}_2(0)$	velocities of m_1, m_2 at $t = 0$	m/s
X	amplitude of $x(t)$	m
X_{dynamic}	amplitude of dynamic displacement	m
X_{static}	static displacement	m
X_1, X_2	amplitudes of $x_1(t)$ and $x_2(t)$ or amplitudes of 1$^{\text{st}}$ and 2$^{\text{nd}}$ cycles of $x(t)$	m
$\{x\}, \{x(t)\}$	$N \times 1$ displacement vector	m
$\{X\}$	$N \times 1$ displacement vector of peak amplitudes	m
$\{\dot{x}\}$	$N \times 1$ velocity vector	m/s
$\{\ddot{x}\}$	$N \times 1$ acceleration vector	m/s^2
$\{x(0)\}$	$N \times 1$ displacement vector at time $t = 0$	m
$\{\dot{x}(0)\}$	$N \times 1$ velocity vector at time $t = 0$	m/s
y	excitation displacement	m
$y(x, t)$	transverse displacement of beam at distance x at time t	m
Y	amplitude of excitation $y(t)$	m
z	relative displacement of mass	m
Z	amplitude of relative displacement $z(t)$	m
α, β, β_n	constants	
δ	logarithmic decrement	
Δ	static displacement	m
$\Delta\omega$	half-power bandwidth	rad/s
ζ	viscous damping ratio	
ζ_i	viscous damping ratio of ith mode	
η	frequency ratio $= \omega/\omega_n$	
θ	dynamic angular displacement	rad
Θ	amplitude of $\theta(t)$	rad

Symbol	Meaning	SI unit
κ_i	generalized or modal stiffness of ith mode	N/m
$[\kappa]$	$N \times N$ generalized diagonal stiffness matrix	N/m
μ	mass per unit length of beam	kg/m
μ_i	generalized or modal mass of ith mode	kg
$[\mu]$	$N \times N$ generalized diagonal mass matrix	kg
ξ	structural/hysteretic damping ratio	
ρ	mass per unit volume	kg/m^3
$\{\sigma\}$	generalized force vector $= [\psi]^T \{\tilde{f}(t)\}$	N/kg
σ_{j0}	amplitude of generalized force of jth DOF	N/kg
$\sigma_j(t)$	generalized force of jth DOF	N/kg
ϕ	phase angle	rad
$[\phi]$	$N \times N$ raw modal matrix	
$\{\phi_1\}, \{\phi_2\}, \ldots,$ $\{\phi_N\}$	raw modal vectors of 1st, 2nd, ..., Nth modes	
$\psi_i, i = 1, 2, \ldots, N$	mass-normalized or weighted eigenvectors	kg$^{-1/2}$
$(_r\psi_j), (_r\psi_k)$	mass-normalized displacements of jth and kth coordinates in rth mode	kg$^{-1/2}$
$[\psi]$	$N \times N$ mass-normalized modal matrix	kg$^{-1/2}$
ω	excitation frequency	rad/s
$\omega_1, \omega_2, \ldots, \omega_i, \omega_j$	1st, 2nd, ..., ith, jth natural frequency	rad/s
ω_{11}	natural frequency of main system	rad/s
ω_{22}	natural frequency of absorber system	rad/s
ω_d	natural frequency with damping	rad/s
$\omega_{d,j}$	natural frequency with damping of jth DOF	rad/s
ω_n	natural frequency (undamped), critical speed of rotor	rad/s

Abbreviations

2DOF	2-degrees-of-freedom
CG	centre of gravity
DC	direct current
DOF	degree-of-freedom
ENIAC	Electronic Numerical Integrator and Computer
FBD	free body diagram
FD	forced draft
FRF	frequency response function
ID	induced draft
MDOF	multi-degrees-of-freedom
NDOF	N-degrees-of-freedom

RPM   revolutions per minute
SDOF   single-degree-of-freedom
SUV   sports utility vehicle
TMD   tuned mass damper

Questions

1. Give two examples of forced vibration: one desirable and one undesirable.
2. What are the three components of a vibratory system?
3. Does damping reduce the amplitude ratio for all values of the excitation frequency for an SDOF system?
4. What is the difference between a discrete system and a continuous system?
5. Define degrees of freedom in vibration analysis.
6. What is the effective stiffness of two springs in series?
7. What is the effective stiffness of two springs in parallel?
8. What are the three most common damping models used?
9. Why is damping always a model of convenience?
10. What is meant by the principle of superposition and for what class of systems is it valid?
11. If energy is lost from a vibrating system, is the system considered to be damped?
12. How can you test a vibratory system to find out if it is linear or non-linear?
13. Under what circumstances do systems exhibit dangerously large vibration amplitudes?
14. What is the characteristic behaviour of an undamped vibratory system?
15. What is meant by critical damping? Give examples of systems tuned to critical damping.
16. How are frequency in Hertz, frequency in radians per second and time period related?
17. When is a system said to perform forced vibration?
18. Why, as engineers, do we like to impart harmonic excitation in studies of system behaviour?
19. Comment on the amplitude, frequency, and phase of steady-state vibration in relation to the applied harmonic force for an undamped system.
20. When is a vibratory system said to be under a condition of resonance?
21. What is the number of degrees of freedom of a simple pendulum?
22. What is meant by the Q-factor of a vibratory system?
23. If the main parts of a vibrating system slide on a dry surface, what is the type of damping the system is said to have?
24. Why are the half-power frequencies also called the 3 dB frequencies?
25. What is meant by simple harmonic motion?
26. What is meant by frequency spectrum? How is it related to the time domain signal?

27. Can a non-linear vibration problem be identified from its governing differential equation? How?
28. When a system vibrates in a fluid medium, which damping model can be used?
29. What is the spring constant of a cantilever beam with end mass m?
30. Why is the vibration due to unbalance in a rotating machine higher at higher frequencies?
31. How can you make an SDOF model of a complicated MDOF system?
32. How can we obtain the frequency, phase, and amplitude of a harmonic motion from (i) the corresponding rotating vector? (ii) the time domain plot? and (iii) the spectrum?
33. Why is the frequency response function of a vibratory system with damping complex?
34. What is meant by static and dynamic coupling in the context of an MDOF system?
35. How can you reduce unbalanced forces in a rotating machine?
36. How can you eliminate coupling of the equations of motion of a vibrating system?
37. What is the electrical analogy of damping in a vibratory system?
38. A vibration displacement amplitude of 4 mm (peak–peak) at 240 RPM is more/less severe than one of 100 g (peak–peak) at 2000 Hz. Strike off what is not applicable and give reasons.
39. In a system vibrating freely, the vibration amplitude reduces to 1/20th of the original value in 15 cycles. What is the damping ratio?
40. What is meant by mass-controlled, stiffness-controlled, and damping-controlled regions of a vibration frequency response function?
41. What is the difference between force transmissibility and displacement transmissibility?
42. If a vehicle undergoes severe vibration while moving on a uniformly bumpy road, will it reduce if the speed is reduced?
43. What is meant by Rayleigh damping?
44. Does the phase angle of the response of an SDOF system depend on the system parameters m, c, k? Explain the physics behind it.
45. What is meant by normal mode summation in forced vibration analysis? When is it used?
46. Why is the dynamic magnification factor close to 1 for frequencies well below and well above resonance? Give a physical explanation.
47. How can you decouple the equations of motion of an MDOF system?
48. How many arbitrary constants are required to solve a second-order, linear differential equation? How are these constants determined?
49. What is meant by a quasi-static problem?
50. What is an eigenvalue problem?
51. How does the response of a system with Coulomb damping differ from that of systems with other types of damping?
52. How can you obtain the Rayleigh damping constants from experiments?
53. Does the vibration of a system depend on the coordinate system chosen?
54. On what basis will you choose the damping of the shock absorbers in a vehicle?

55. Why is it required to know the natural frequency of a vibrating system?
56. Under what circumstances can you ignore damping in vibration analysis?
57. What is meant by equivalent viscous damping? Is it a constant?
58. How do you define the structural damping loss factor?
59. What do you mean by inertia coupling, velocity coupling, and elastic coupling?
60. Can the damped natural frequency in some cases be larger than the undamped natural frequency of a system?
61. When do you use vibration isolation and when do you use a vibration absorber?
62. During free vibration, do different degrees-of-freedom oscillate (i) with different amplitudes? (ii) with different frequencies?
63. What are the considerations in the design of engine mounts?
64. Give a practical application of a torsional pendulum.
65. What are the main issues in suspension design?
66. Will the amplitude of vibration of an undamped system change with time?
67. What vehicle parameters are required to compute its bounce and pitch natural frequencies?
68. When do you use vibration isolation and when do you use damping?
69. Is the natural frequency of a vibratory system with damping smaller or greater than that without damping?
70. How can you obtain effective vibration isolation from road undulations in a vehicle?
71. Why is a viscous damping model used in most cases rather than other types of damping?
72. Does the nature of the coupling between DOF depend on the coordinates used?
73. Outline a procedure for obtaining damping ratio of any vibratory mode from experiment.
74. If the characteristic roots have a zero imaginary part, what can you comment on the response of a vibratory system?
75. What is a degenerate system? Give an example of a physical system that is degenerate.
76. What happens to the energy dissipated by damping?
77. Differentiate between viscous and structural damping loss factors.
78. How can a system be made to vibrate in one of its natural modes?
79. What is the advantage of a proportional damping model in the analysis of an MDOF system?
80. What is meant by complex stiffness?
81. What is the effect of a tensile force on the natural frequencies of a beam?
82. Does the motion reduce to zero in both underdamped and overdamped vibratory cases?
83. How can you find the natural frequency of a system by measuring its static deflection?
84. What is meant by reciprocity theorem?
85. What is the difference between transient and steady-state responses?
86. How does a continuous system differ from a discrete system in the nature of its equation of motion?

87. When a mass vibrates in a vertical direction, can its weight be neglected while deriving the equation of motion? Why?
88. What parameters correspond to m, c, k, and x for a torsional system?
89. How can you change the natural frequency of a system?
90. What do you mean by principal coordinates? When are they used?
91. Can a system vibrating in air be considered a damped system?

Fill in the Blanks

92. The number of degrees of freedom of a continuous system is _____.
93. If a system vibrates due to initial disturbances alone, it is said to perform _____ vibration.
94. A turbine with a speed of 3000 RPM takes _____ seconds to complete one revolution.
95. When the stress–strain curve of the material of a vibrating system exhibits a hysteresis loop, the damping is said to be _____.
96. If a machine is running at 2000 revolutions per minute, its frequency in radians per second is _____.
97. To achieve low force transmissibility, damping in the system should be _____.
98. For good ride comfort, suspension stiffness should be _____.
99. The main use of the orthogonality conditions for an MDOF system is _____.

Indicate Whether Each of the Following Statements Is True or False. Give Reasons

100. The relative amplitudes of different DOF in an MDOF system depend on the natural frequency.
101. The modal vectors of a system denote the normal modes of vibration.

Bibliography

1. Anderson, R. A. (1967). *Fundamentals of vibrations*. New York: Macmillan.
2. Beards, C. (1996). *Structural vibration analysis and damping*. Butterworth-Heinemann.
3. Bernhard, R. K. (2010). *Mechanical vibrations theory and application. An introduction to practical dynamic engineering problems in the structural field*. Read Books Design.
4. Bishop, R. E. D., Gladwell, G. M. L., & Michaelson, S. (1965). *The matrix analysis of vibration*. London: Cambridge University Press.
5. Brandt, A. (2011). *Noise and vibration analysis*. Wiley.
6. Bottega, W. J. (2006). *Engineering vibrations*. CRC Press Online.
7. Clough, R. W., & Penzien, J. (1993). *Dynamics of structures*. New York: McGraw Hill.
8. Den Hartog. J. P. (1985). *Mechanical vibrations*. New York: Dover.

9. Dimarogonas, A. D., & Haddad, S. (1992). *Vibration for engineers*. New Jersey: Prentice Hall.
10. Dukkipati, R. V. (2004). *Vibration analysis*. Harrow: Alpha Science International Ltd.
11. Ginsberg, J. H. (2001). *Mechanical and structural vibrations: Theory and applications*. New York: Wiley.
12. Kelly, S. G. (1996). *Schaum's outline of theory and problems of mechanical vibrations*. New York: McGraw-Hill.
13. Hansen, C. H., & Snyder, S. D. (1997). *Active control of noise and vibration*. London: Taylor & Francis.
14. Hansen, C. (2018). *Foundations of vibroacoustics*. Routledge.
15. Harker, R. J. (1983). *Generalised methods of vibration analysis*. New York: Wiley.
16. Huston, D. V. (1981). *Applied mechanical vibrations*. New York: McGraw-Hill.
17. Inman, D. J. (2001). *Engineering vibration*. Upper Saddle River: Prentice Hall.
18. Inman, D. J. (2006). *Vibration with control*. Chichester: Wiley.
19. Iyengar, R. N. (2011). *Elements of mechanical vibration*. I K International Publishing House Pvt. Ltd.
20. Jacobsen, L. S., & Ayre, R. S. (1958). *Engineering vibrations*. New York: McGraw-Hill.
21. Ji, X. J. Q. (2016). *Handbook of mechanical vibrations and noise engineering*. Auris Reference Limited.
22. Kelly, S. G. (2011). *Mechanical vibrations: Theory and applications*. Cengage Learning.
23. Kovacic, I., & Radomirovic, D. (2017). *Mechanical vibration: Fundamentals with solved examples*. Wiley.
24. Lalanne, C. (2002). *Mechanical vibration & shock*. North America: Taylor & Francis.
25. Lalanne, M., Berthier, P., & Hagopian, D. J. (1983). *Mechanical vibration for engineers*. New York: Wiley.
26. Leissa, A., & Qatu, M. (2011). *Vibration of continuous systems*. McGraw-Hill Education.
27. Meirovitch, L. (1967). *Analytical methods in vibrations*. New York: Macmillan.
28. Meirovitch, L. (1970). *Methods of analytical dynamics*. New York: McGraw Hill.
29. Meirovitch, L. (1986). *Elements of vibration analysis*. New York: McGraw-Hill.
30. Myklestad, N. O. (1956). *Fundamentals of vibration analysis*. New York: McGraw-Hill.
31. Nataraj, C. (2011). *Vibration of mechanical systems*. Cengage Learning India.
32. Nilsson, A., & Bilong, L. (2015). *Vibro-acoustics* (Vol. 1). Berlin: Springer.
33. Pain, H. J., & Rankin, P. (2015). *Introduction to vibrations and waves*. Wiley.
34. Ramamurti, V. (2000). *Mechanical vibration practice with basic theory*. New Delhi: Narosa Publishing House.
35. Rao, S. S. (1986). *Mechanical vibrations*. Reading: Addison-Wesley.
36. Rao, S. S. (2007). *Vibration of continuous systems*. Wiley.
37. Reynolds, D. D. (1981). *Engineering principles of acoustics: Noise and vibration*. Boston: Allyn & Bacon.
38. Rossing, T. D., & Fletcher, N. H. (2004). *Principles of vibration and sound*. New York: Springer.
39. Schmitz, T. L., & Smith, K. S. (2012). *Mechanical vibrations*. Springer.
40. Shabana, A. A. (1996). *Theory of vibration: An introduction*. New York: Springer.
41. Shabana, A. A. (1997). *Vibration of discrete and continuous systems*. Springer.
42. Sinha, A. (2014). *Vibration of mechanical systems*. Cambridge University Press.
43. Srinivasan, P. (1990). *Mechanical vibration analysis*. New Delhi: Tata McGraw-Hill Publishing Company Limited.
44. Sujatha, C. (2010). *Vibration and acoustics: Measurement and signal analysis*. Tata McGraw-Hill Education.
45. Svetlitsky, V. A., Chechin, V. A., & Merzon, G. I. (2004). *Engineering vibration analysis: Worked problems* (Vol. 1). New York: Springer.
46. Svetlitsky, V. A., Lidvansky, A. S., & Mukhamedshin, R. A. (2004). *Engineering vibration analysis: Worked problems* (Vol. 2). New York: Springer.
47. Thomson, W. T., & Dahleh, M. D. (2003). *Theory of vibration with applications*. India: Pearson Education India.
48. Thomson, W. T. (2004). *Theory of vibration with applications*. Bristol: Taylor & Francis.

49. Tse, F. S., Morse, I. E., & Hinkle, R. T. (1978). *Mechanical vibrations theory and applications*. Boston: Allyn and Bacon.
50. Vasques, C. M. A., & Rodrigues, J. D. (2014) *Vibration and structural acoustics analysis: Current research and related technologies*. Springer.
51. Weaver, W., Timoshenko, S., & Young, D. H. (1990). *Vibration problems in engineering*. New York: Wiley-IEEE.

Chapter 3
Equipment for Measurement and Generation of Vibration

DID YOU KNOW that the world's first seismograph was invented in China in 132 AD? But it was in no way similar to the present-day instruments. China was affected by many earthquakes in ancient times. Zhang Heng, who was a historian and astronomer, saw the need to develop an instrument to measure earthquakes. He had a device cast in fine bronze, shaped like a wine jar with a diameter of eight chi (a chi being 0.237 m). Inside the jar was a mechanism consisting of pendulums surrounded by a group of eight levers pointing in eight directions. Outside the seismograph were eight dragon figures, with bronze balls in their mouths. Each dragon looked upon a toad placed beneath it with its mouth open upward. The idea was that in the event of a strong earthquake in any direction, the corresponding pendulum would tilt in that direction and trigger the lever in the dragon head. This would open the mouth of the dragon, allowing the bronze ball to fall into the mouth of the toad with a clanging sound. This would warn the monitoring personnel to know the time and direction of occurrence of the earthquake.

3.1 Introduction to Vibration Transducers

We are often faced with the task of measuring the vibration level in a machine to decide if it is within allowable limits. For this we need to use a vibration transducer. Considering that the market is flooded with a variety of transducers, choosing the appropriate transducer for the application at hand becomes difficult and requires answers to the following questions:

  (i) Do I need to measure displacement, velocity, or acceleration?
 (ii) Can I use contact type of transducers without introducing mass loading?
(iii) Do I need any additional signal conditioning devices to go with the transducers?
 (iv) What are the frequency range, dynamic range, and sensitivity required?
  (v) Is the choice of the transducer dictated by any environmental constraints like dust, electromagnetic fields, humidity, shock levels, etc.?
 (vi) What is my budget?

One should arrive at the appropriate transducer only after one has the answers to all these questions. Otherwise, the very purpose of measurement may be defeated. If one were to make vibration measurements on a ball mill rotating at 100 revolutions per minute (RPM) using a piezoelectric accelerometer, one would find that all the low-frequency vibrations have not been picked up. If one were to use a heavy accelerometer to measure the vibrations of a light panel, one would find that the mass loading effects have altered the system dynamics completely. The transducer being the front end in any measurement system, care has to be taken in selecting an appropriate one. All further signal conditioning and analysis will be a waste of effort if the signals have not been correctly captured in the first place. This necessitates that the user has a thorough understanding of the various types of transducers available in the market, their principles of operation, their advantages and limitations, typical specifications, etc. This chapter aims at providing all this information.

Vibration transducers can be classified into the following categories based on the following:

  (i) Parameters being measured:

  - Displacement.
  - Velocity.
  - Acceleration.

 (ii) Electrical output:

  - Passive.
  - Active.

(iii) Type of output:

  - Relative.
  - Absolute.

(iv) Proximity of the transducer to the vibrating object:

- Contact.
- Non-contact.

3.1.1 Choice of Vibration Parameter—Displacement, Velocity, and Acceleration

In its simplest form, a vibration displacement signal may be represented by the harmonic function

$$x(t) = X \sin \omega t \tag{3.1}$$

By differentiation, we have velocity and acceleration given by

$$\dot{x}(t) = X\omega \cos \omega t \tag{3.2}$$

$$\ddot{x}(t) = -X\omega^2 \sin \omega t \tag{3.3}$$

Equations (3.1)–(3.3) show how displacement, velocity, and acceleration components of a complex signal are dependent on the angular frequency ω. Thus, the amplitude of velocity is proportional to frequency and that of acceleration to frequency squared. Any vibratory system may be represented by an equation of the form given below as discussed in Chap. 2.

$$m\ddot{x} + c\dot{x} + kx = F(t) \tag{3.4}$$

Here the terms on the left-hand side are those of inertia, damping, and elastic forces in that order. By substituting Eqs. (3.1)–(3.3) into (3.4), we find that the relative magnitudes of the various forces in Eq. (3.4) are frequency dependent. An important observation to be made is that at high frequencies, the force component due to acceleration may dominate, resulting in large forces, even with very low displacement amplitudes, on account of the ω^2 effect. It is indeed these high-frequency components, which remain undetected by displacement criteria and cause sudden catastrophic fatigue failures. One can therefore conclude that displacement alone, as often applied, can be a poor measure of vibration severity, since this should always be considered in conjunction with frequency. To overcome this difficulty, maximum velocity criteria are used so as to include the frequency parameter and the total dynamic force concept in the measurement of machinery vibration.

The frequencies of interest in machinery vibration analysis form a wide continuous spectrum from the infrasound range of a few Hz through the ultrasonic range approaching 100 kHz. Within this range are present subharmonic 'whirl' frequencies of the order of 10s of Hz, once per revolution frequencies from say 20 to 700 Hz, vane or blade or gear tooth passing frequencies from 2 to 30 kHz and finally antifriction bearing frequencies from 20 to 100 kHz. With such a vast range of frequencies, it is
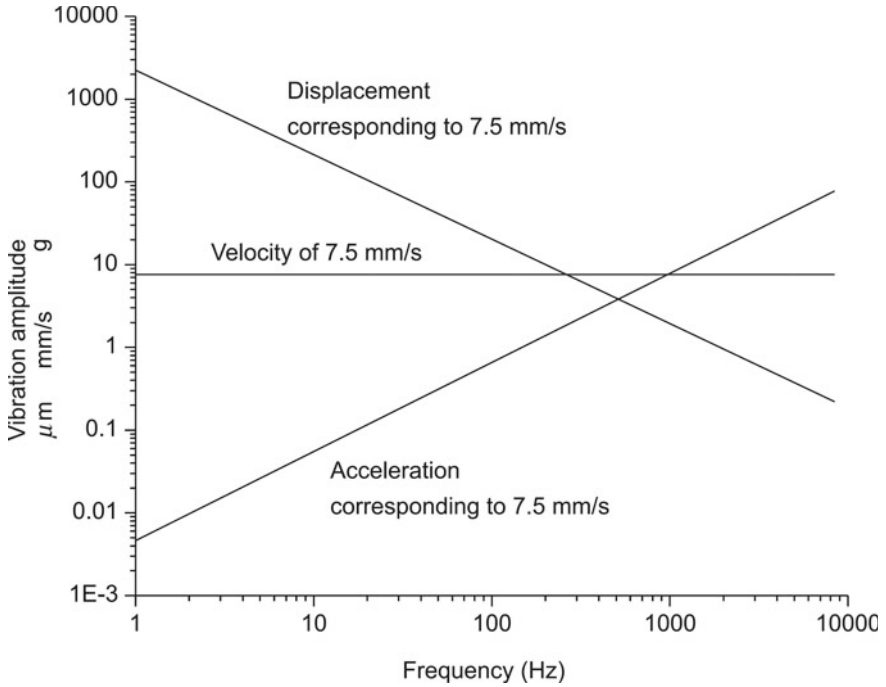
Fig. 3.1 Displacement, velocity, and acceleration nomograph

imperative that the vibration engineer acquires and presents data in an effective and meaningful manner. Figure 3.1 gives vibration displacement and acceleration as a function of frequency for a constant velocity level of 7.5 mm/s. This velocity corresponds to a vibration displacement of 1.194 mm (1194 μm) at 1 Hz which amounts to a very small acceleration of 47.1 mm/s^2 (0.0048 g). On the other hand, at 1000 Hz, the same vibration velocity of 7.5 mm/s corresponds to a high acceleration of 47,100 mm/s^2 (4.8 g) with a displacement of only 1.194×10^{-3} mm (1.194 μm).

It is clear from the figure that transducers are like windows through which portions of the frequency spectrum may be observed. In general, displacements are large at low frequencies and displacement sensors, or vibrometers as they are called, are preferred for these frequencies. Acceleration values are large at high frequencies and for such applications accelerometers are preferred. Velocity measurements are useful at intermediate frequencies where the displacements are likely to be small, or where the frequency range of measurements is not known beforehand. The displacement pickup has a maximum frequency range from 0 to about 1000 Hz, the velocity pickup from 10 to about 2500 Hz, and the accelerometer from 20 Hz to well above 50 kHz.

Table 3.1 Energy conversion used in active vibration transducers

Original form	Transduced form	Device or effect	Reversible
Mechanical	Electrical	Magnetic induction Electric induction	Yes Yes
Pressure	Electrical	Piezoelectric	Yes
Light radiation	Electrical	Photoelectric	No
Acoustic	Electrical	Microphone	Yes
Electrical	Acoustic	Loudspeaker	Yes

3.1.2 Active and Passive Transducers

Two different principles are involved in the process of converting non-electrical variables into electrical signals. One of these is energy conversion; transducers based on this principle are called active transducers. In theory, active transducers can utilize any known physical principle for converting non-electrical energy, typically mechanical energy, into electrical energy. In active transducers, the same transduction principle used to convert from a non-electrical form of energy can also be used in the reverse direction to convert electrical energy into non-electrical forms. Examples are the electrodynamic transducer and microphone, which give a voltage output proportional to vibration or pressure fluctuations. Table 3.1 shows some methods of energy conversion used in active transducers.

Passive transducers use a different principle to generate electrical signals other than voltage or current. The principle involved here is control of a direct current (DC) excitation voltage or modulation of an alternating current (AC) carrier signal. The mechanical signal is converted into a parameter such as a change in a circuit resistance or inductance or capacitance. Examples of passive transducers are strain gauges, linear variable differential transformers (LVDT), and capacitance pickups, which give output resistance, inductance, and capacitance, respectively, proportional to the input signal.

3.1.3 Absolute and Relative Measuring Instruments

Fixed reference instruments or absolute measuring devices have one terminal of the instrument attached to a point which is fixed in space, whereas the other terminal is attached to the vibrating object. The measurements made using such transducers are independent of the user. Many of them incorporate a so-called seismic mass, which offers a fixed frame of reference with respect to which measurements are made. Relative measuring instruments, on the other hand, display different values with different users since measurements are not made with respect to a fixed frame.
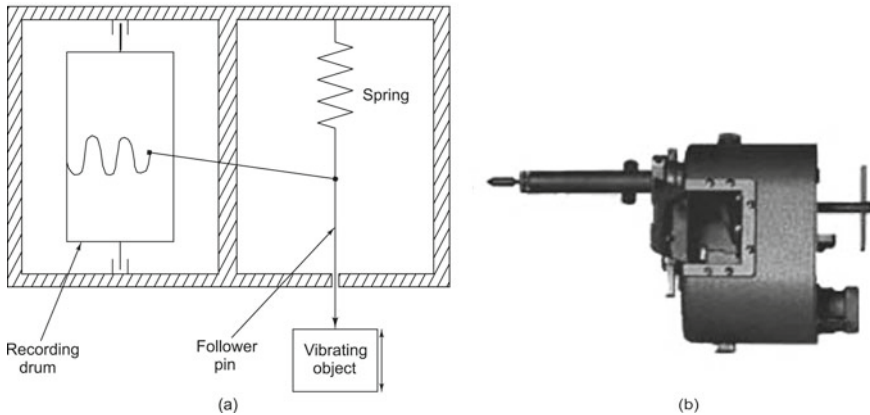
Fig. 3.2 Hand vibrograph: **a** schematic drawing, **b** photograph (Courtesy of IIT Madras)

Quite often the readings depend on the spring force with which the user keeps the transducer in contact with the vibrating object.

3.1.4 Contact and Non-contact Transducers

Contact type of transducers may be used for conducting vibration tests on heavy specimens such as machine foundations, large machines, road vehicles, ships, bridges, and buildings, but care must be exercised in using these on lightweight structures since they are likely to cause mass loading effects. Hence, non-contact transducers have to be used for such applications. Further, contact pickups, whether heavy or light, cannot be used on rotating machines and non-intrusive measurement techniques have necessarily to be used.

3.1.5 Earliest Vibration Transducer: The Hand Vibrograph

A hand vibrograph (Fig. 3.2) is an easy-to-handle robust device used for the measurement of vibrations. This instrument is one of the earliest vibration transducers and is essentially of historic importance since it has become obsolete. It is a relative measuring instrument, which measures displacement and has no fixed reference point (like a seismic mass) as in the case of absolute measuring instruments (e.g., LVDT). Owing to its compactness and small weight, it is versatile in its application and can be used for measurements without any special preparations.

It consists of a housing having two chambers. In one chamber, a rotating recording drum is present. Paper is wound around this rotating drum, so that the pen or stylus

will scribe the vibration records on it. The pen is coupled to the follower pin or feeler through a system of levers. The pin is attached to a spring present in another chamber. It protrudes out of the housing and during operation is pressed against the vibrating object in the direction of motion so that it can freely move up and down between its stops. The vibrations are then magnified and recorded on the rotating drum. When the vibrating object moves with a particular acceleration in the downward direction, the spring force has to overcome the inertia force of the follower pin and system of levers, to keep the pen pressed against the object, so that it will not chatter. A greater spring force admits measurement of higher accelerations.

The vibrograph can be held to the vibrating object in two ways and the frequency response of the recording depends on this, since it is a function of the relative movement between the pin and the housing. When holding the vibrograph with both hands, it acts as a vibrometer. This differs from a standard vibrometer as it forms an inert mass itself and the holding mass together with the spring causes the flexibility effect. It is recommended that the instrument be used only above the natural frequency, i.e., typically above 3–5 Hz in which case the recorded amplitude will always be the product of vibration displacement and amplification. If clamped in a vice, the instrument becomes a simple displacement instrument; this requires a little more preparatory work and in this mode it can also be used for low frequencies down to 0 Hz. The limit of acceleration, as in the case of vibrograph application, depends upon the force of tension or rigidity of the clamping. The upper frequency limit is dependent upon the rigidity of the clamping. In the case of ideal rigid clamping, the record will be proportional to the vibration amplitude throughout the usable frequency range.

Typical Specifications

Frequency range: 5–100 Hz.
Amplitude range: 25 μm–2 mm.
Magnification: 1, 5, 20, 50.
Maximum acceleration: 20.
Weight: 15 N.

Advantages

 (i) It is a compact device.
 (ii) It can be made to work in any desired direction and does not require any preparatory measures for attaching it to the vibrating object.
(iii) It is simple to use with built-in battery.

Disadvantage Readings taken by different operators can be different due to differing spring forces.

Applications

 (i) It can be used for measurement of critical speeds, unbalances, etc. on steam turbines.

(ii) Rough preliminary displacement measurements can be made on machines, machine tools like lathe, grinding machine, power plants, automobiles, and buildings.

(iii) It provides early warning of breakage or excessive wear.

3.2 Absolute Measuring Transducers or Seismic Transducers

DID YOU KNOW that a seismic transducer is an instrument consisting of a mass attached to a fixed base? This transducer evolved from the seismometer meant for measuring ground motion during an earthquake. When the earth shakes, the base of the instrument moves and the mass does not. The motion of the base relative to the mass is commonly transformed into an electrical voltage. Due to the stationary mass which serves as a reference, it is called a seismic transducer and is an absolute measuring instrument. Later seismic transducers that measure velocity and acceleration were developed.

3.2.1 Seismic Transducer Theory

The absolute measuring or seismic transducer can be modelled as shown in Fig. 3.3. It has a seismic mass, i.e., one that remains stationary in space over the usable frequency range of the transducer. Depending on the frequency range used, displacement, velocity, or acceleration is indicated by the relative motion of the suspended mass with respect to the casing. The equation of motion of the system shown in Fig. 3.3 may be written as

$$m\ddot{x} = -c(\dot{x} - \dot{y}) - k(x - y) \tag{3.5}$$

Fig. 3.3 Vibratory model of seismic transducer
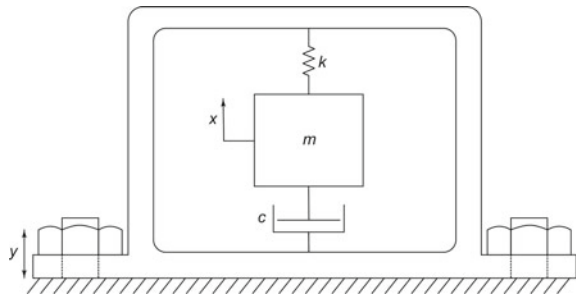
where x and y are the displacements of the seismic mass and vibrating body, respectively, measured with respect to an inertial frame of reference. Letting the relative displacement of the mass m with respect to the case attached to the vibrating body be

$$z = (x - y) \qquad (3.6)$$

and assuming sinusoidal motion $y = Y \sin \omega t$ of the vibrating body, we obtain the equation

$$m\ddot{z} + c\dot{z} + kz = m\omega^2 Y \sin \omega t \qquad (3.7)$$

The steady-state solution is

$$z = Z \sin(\omega t - \phi) \qquad (3.8)$$

The frequency response function (FRF) Z/Y is a complex quantity with magnitude and phase as shown below:

$$\begin{aligned} Z &= \frac{m\omega^2 Y}{\sqrt{(k - m\omega)^2 + (c\omega)^2}} \\ &= \frac{Y(\omega/\omega_n)^2}{\sqrt{\{[1 - (\omega/\omega_n)^2]\}^2 + [2\zeta(\omega/\omega_n)]^2}} \end{aligned} \qquad (3.9a)$$

or

$$\left|\frac{Z}{Y}\right| = \frac{\eta^2}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \qquad (3.9b)$$

where $\eta = \omega/\omega_n = $ the frequency ratio and ζ is the damping ratio. The phase lag is given by

$$\tan \phi = \frac{\omega c}{k - m\omega^2} = \frac{2\zeta(\omega/\omega_n)}{1 - (\omega/\omega_n)^2} \qquad (3.10)$$

Figure 3.4 shows a plot of Eqs. (3.9b) and (3.10) as a function of η. It is evident from the figure that when $\eta = \omega/\omega_n \gg 1$, i.e., when the natural frequency ω_n of the instrument is low in comparison to the vibration frequency ω to be measured, the ratio $|Z/Y|$ approaches unity and the relative displacement Z approaches Y irrespective of the value of damping ratio ζ. This means that the mass remains undisturbed in space. Therefore, the relative motion between the casing and the mass is the true displacement of the casing. Likewise, the relative velocity between the casing and the mass is the true velocity of the casing. Since these transducers work satisfactorily for frequency ratios greater than 3, they must necessarily have a low natural frequency, requiring incorporation of a large mass and soft spring in the transducer.

The choice of damping in such instruments is very critical as seen from the plot of $|Z/Y|$ in Fig. 3.4. When ζ is of the order of 0.6–0.7, the percentage error in Z as compared to Y is very small. Introduction of artificial damping improves the usable lower frequency of the vibrometer, making it usable right from η equal to around 2.
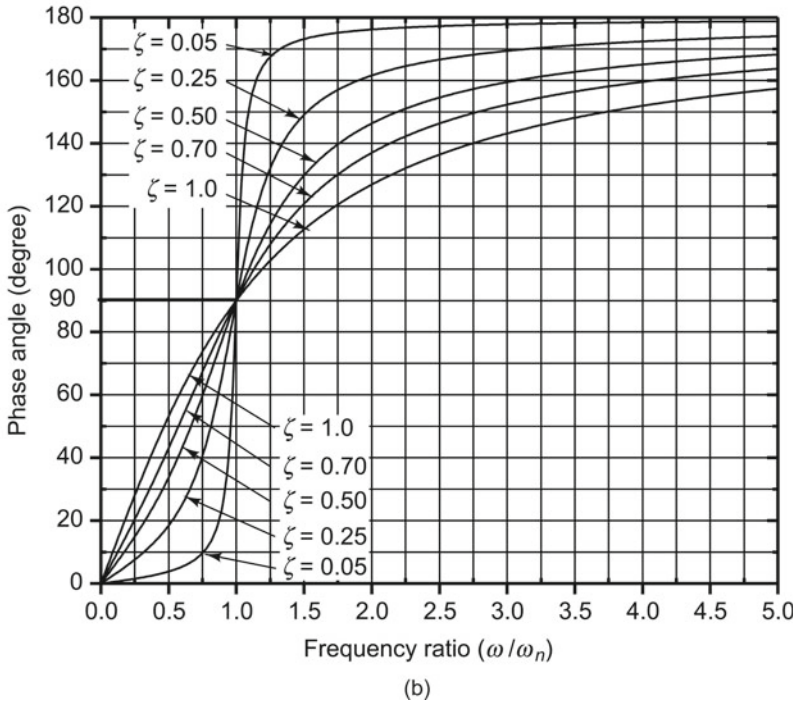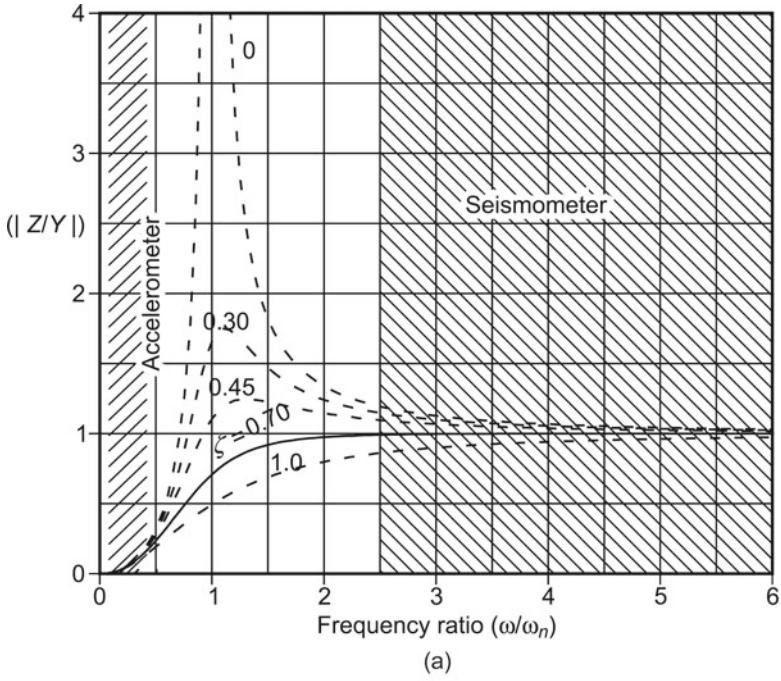
Fig. 3.4 Response of seismic transducer: **a** magnitude, **b** phase

Besides, this value of ζ also offers a reasonably small settling time and low phase distortion.

On the contrary, for values of η far less than 1, $Z/Y = \eta^2$. The acceleration of the casing will then be proportional to the relative displacement of the seismic mass with respect to the casing. These instruments are known as accelerometers. They have very high natural frequencies which are obtained using small masses and very stiff springs and are used up to around 0.5 times their natural frequencies.

For such transducers, the frequency response function is

$$\left| \frac{Z}{Y} \right| = \frac{\omega^2}{\omega_n^2 \sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \tag{3.11}$$

It is clear that in the above equation, the denominator of the right-hand side shown below approaches unity for $\omega/\omega_n \to 0$, i.e.,

$$\frac{1}{\sqrt{\{[1 - (\omega/\omega_n)^2]\}^2 + [2\zeta(\omega/\omega_n)]^2}} \longrightarrow 1 \text{ as } \frac{\omega}{\omega_n} \to 0 \tag{3.12a}$$

Hence

$$Z = \frac{\omega^2 Y}{\omega_n^2} = \frac{\text{acceleration}}{\omega_n^2} \tag{3.12b}$$

Thus, Z becomes proportional to the acceleration of the motion to be measured with a factor of $1/\omega_n^2$. The useful range of the accelerometer can be seen from Fig. 3.5, which is a magnified plot of the right-hand side term in Eq. (3.11) for various values of damping ratio ζ as a function of η. The corresponding phase plot is the same as shown in Fig. 3.4.

Figure 3.5 shows that the useful frequency range of the undamped accelerometer is rather limited. When the damping ratio is between 0.6 and 0.7, the correction factor is within 4% of unity for η up to 0.5. As in the case of the displacement transducer, the usable frequency range of the accelerometer is extended by choosing a damping ratio between 0.6 and 0.7; in this case, it is extended up to 0.5 times its natural frequency. Besides, $\zeta = 0.6 - 0.7$ ensures a good transient response. Another consideration is that this value of ζ also minimizes phase distortion for complex waves by ensuring linear phase, i.e., $\phi = \pi\omega/2\omega_n$ is satisfied with good accuracy.

Let the vibration sensed by an accelerometer be

$$y(t) = Y_1 \sin \omega_1 t + Y_2 \sin \omega_2 t + Y_3 \sin \omega_3 t + \cdots \tag{3.13a}$$

Then the output of a seismic accelerometer with $\zeta = 0.7$ would be

The y-axis is labeled:

$$\dfrac{1}{\sqrt{\left[1-\left(\dfrac{\omega}{\omega_n}\right)^2\right]^2+\left[2\zeta\left(\dfrac{\omega}{\omega_n}\right)\right]^2}}$$

Curves labeled $\zeta=0$, $\zeta=0.55$, $\zeta=0.65$, $\zeta=0.70$, $\zeta=0.80$.

The x-axis is labeled $\dfrac{\omega}{\omega_n}$.
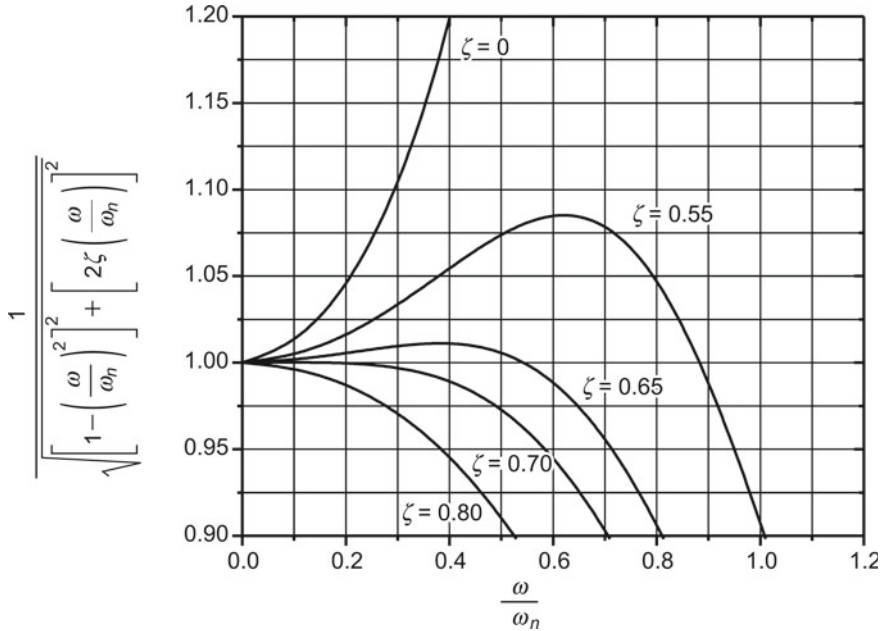
Fig. 3.5 Acceleration error versus frequency

$$z(t) = \left(\frac{\omega_1^2}{\omega_n^2}\right) Y_1 \sin \omega_1 \left(t - \frac{\pi}{2\omega_n}\right) + \left(\frac{\omega_2^2}{\omega_n^2}\right) Y_2 \sin \omega_2 \left(t - \frac{\pi}{2\omega_n}\right)$$

$$+ \left(\frac{\omega_3^2}{\omega_n^2}\right) Y_3 \sin \omega_3 \left(t - \frac{\pi}{2\omega_n}\right) + \cdots \tag{3.13b}$$

Such linear phase leads to the same time lag $\pi/2\omega_n$ for all frequency components, ensuring that there is no distortion.

3.2.2 Seismic Displacement Transducer of the Inductive Type

The inductive pickup or LVDT is a passive transducer consisting essentially of a spring–mass–damper system as shown in Fig. 3.6. This transducer can be used for the measurement of medium-to-large displacements at low frequencies.

The housing of the pickup is fixed to the object whose vibrations are to be sensed. The larger LVDTs (mass > 1 kg) are simply made to rest on the vibrating object after appropriate levelling, while the plungers of the smaller ones may be kept pressing against the vibrating object after the transducer is fixed to a stationary stand. The pickup is so designed that the mass remains more or less fixed in space for the specified frequency range of operation and hence the core connected to the mass
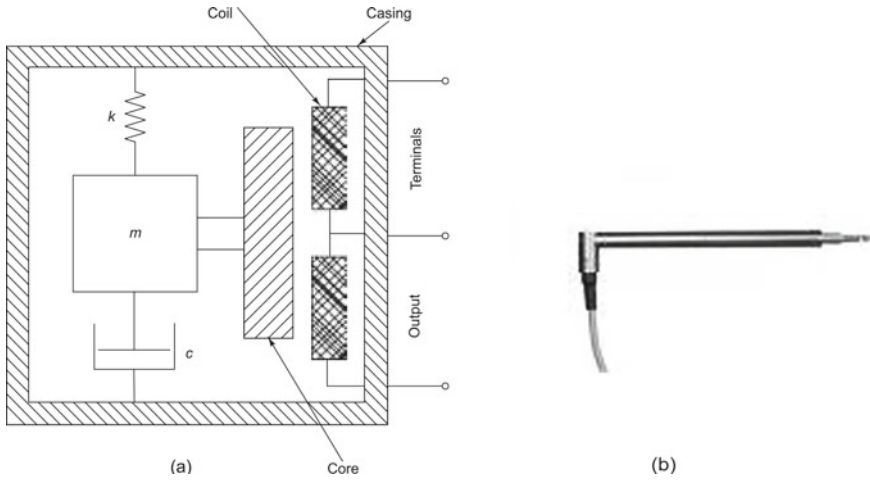
Fig. 3.6 Inductive transducer: **a** schematic, **b** photograph (Courtesy of HBM, Germany)

serves as a satisfactory stationary reference. The core also provides the magnetic flux path linking the two inductive coils which are connected to the housing of the pickup and follow its motion. The two coils are connected in phase opposition to the centre, or null position, so that the output is the difference of these voltages. Hence, the vibration to be measured is sensed as a relative motion between the core and the coils and is proportional to the change in inductance due to the relative motion. The null position of an LVDT is extremely stable and repeatable.

By designing the transducer to have a large mass and a soft spring, the change in inductance may be made proportional to the displacement of the vibrating object. The two inductive coils constitute two arms of an AC Wheatstone bridge. The remaining two arms, the required AC excitation voltage, signal conditioning amplifier, and phase sensitive detector are provided by a carrier frequency amplifier. Initially, the bridge is balanced to give zero output. Any vibration will now cause a change in inductance and hence bridge balance gets affected. The voltage corresponding to the amount of bridge unbalance can be measured on an oscilloscope or recorder. The inductive LVDT has a low natural frequency ω_n and can be used in the frequency range ω such that $1.414 \leq \omega/\omega_n \leq \infty$. The transducer is provided with a damping factor between 0.6 and 0.7 so that the measured relative motion is equal to the displacement of the vibrating object in the frequency range mentioned.

Specifications of Typical LVDTs

	Large	Small
Natural frequency:	1 Hz	3 Hz
Frequency range:	1–200 Hz	5–65 Hz
Vibration amplitude:	±25 mm	±5 mm
Nominal sensitivity:	80 mV/V	80 mV/V

Advantages

   (i) Relatively low cost.
  (ii) Solid and robust, capable of working in a wide variety of environments.
 (iii) No frictional resistance, since the iron core does not contact the transformer
       coils, resulting in a very long service life.
  (iv) High signal-to-noise ratio (SNR) and low output impedance.
   (v) Negligible hysteresis.
  (vi) Excellent resolution limited only by the resolution of the amplifiers.
 (vii) Short response time, limited by the inertia of the iron core and the rise time of
       the amplifiers.

Disadvantages Readings affected by extraneous magnetic fields.

Applications

  (i) Vibration monitoring of huge structures, buildings, railway carriages, bridges,
      machine foundations, and ship hull vibration.
 (ii) Low-frequency measurements.

3.2.3 Seismic Velocity Transducer: Electrodynamic Pickup

We have seen in Sect. 3.2.1 that the relative displacement sensed by a seismic trans-
ducer is related to the actual input displacement by Eq. (3.9). Since the amplitude
of the input velocity is ωY and the amplitude of the relative velocity sensed by the
transducer is ωZ, it follows that

$$\left|\frac{\omega Z}{\omega Y}\right| = \left|\frac{Z}{Y}\right| = \frac{\eta^2}{\sqrt{(1-\eta^2)^2 + (2\zeta\eta)^2}} \tag{3.14}$$

The electrodynamic velocity transducer is a moving coil or moving magnet pickup
of the seismic type. Also called a velometer, this transducer works on the principle
of magnetic induction. If an electrical conductor is moved in a magnetic field so as to
change the magnetic flux through the conductor, a voltage is induced in the conductor
which is proportional to the rate of change of the magnetic flux. Conversely, if a cur-
rent is passed through the same conductor, the latter experiences a mechanical force,
proportional to the current and the magnetic field. The conversion from electrical to
mechanical energy or vice versa depends on the polarities of voltage and current, or
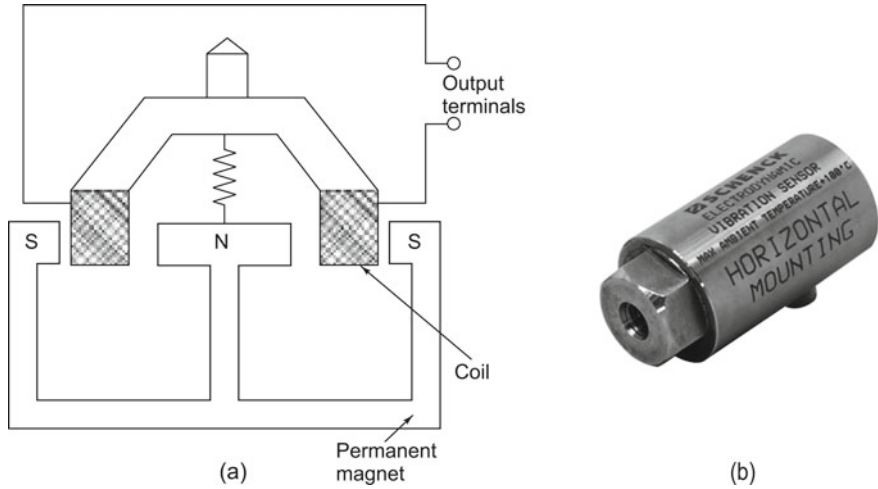
Fig. 3.7 Electrodynamic transducer: **a** schematic, **b** photograph (Courtesy of www.schenck-india. com)

the directions of force and motion, respectively. All electrical motors and generators and many other devices, such as solenoids and loudspeakers, utilize this principle.

In the electrodynamic vibration transducer, the relative motion Z in Eqs. (3.9) and (3.14) is usually converted to an electric voltage by making the seismic mass a magnet moving relative to coils fixed in the case as shown in Fig. 3.7.

When the pin fixed to the coil moves up and down within the magnetic field, a voltage is generated across the coil and this is proportional to the rate of cutting of the magnetic field or the velocity experienced by the vibrating body. The output velocity can be differentiated or integrated to get acceleration or displacement, respectively. The output of the instrument, i.e., the induced voltage e is as shown in Eq. (3.15a).

$$e = Blv \tag{3.15a}$$

where B is the effective flux density in the winding in Wb/m^2 or Tesla, v is the relative velocity between the coil and magnet in m/s, e is the induced open-circuit electromotive force (emf) in V, and l is the length of conductor cutting the flux in m.

The springs are initially compressed so that the vibrating object has to exert additional force to lower the pin into free motion in the magnetic path. The sensitivity (S_t) is given by

$$S_t = \frac{\mathrm{d}e}{\mathrm{d}v} = Bl \tag{3.15b}$$

To increase the sensitivity, B and l or l should be increased. The value of B depends on the material of the permanent magnet. Further increase in sensitivity can be obtained only by increasing the length of the coil. But increase in length results in increase in

weight of the coil, which may give rise to loading effect. Greater length with reduced loading effect can be obtained for a given mass of the coil by reducing its cross section. This in turn results in a higher output resistance of the coil, necessitating the use of a voltmeter with a much higher resistance for reading the output voltage of the pickup. Thus, increasing the sensitivity is possible to a limited extent only.

Advantages This is one of the most versatile transducers for machinery condition monitoring, for the simple reason that it costs less than other vibration transducers and can be easily installed on machines. It does not need any additional signal conditioning amplifiers and its output may be directly fed to an oscilloscope/recorder/signal analyser. Besides, it gives vibration levels in terms of velocity, which is the parameter, in general, used to specify vibration severity criteria. Most velocity pickups have a frequency response in the range of 10–5000 Hz.

Disadvantage This transducer is affected by extraneous magnetic fields.

Typical Specifications

Sensitivity: 100 mV/(cm/s).
Compression force: 8 N.
Maximum acceleration: 20.
Displacement: 2 mm peak to peak.
Frequency range: 10–1000 Hz.

Applications

 (i) Machinery vibration diagnostics.
(ii) Modal analysis.

3.2.4   Seismic Inductive Accelerometer

The inductive accelerometer (Fig. 3.8) is similar in construction to the LVDT as shown in Fig. 3.6a, the major difference being that the transducer is designed to have a small mass and a stiff spring such that the change in inductance is proportional to the acceleration of the vibrating object. It is a passive transducer consisting essentially of a spring–mass–damper system. This pickup is of the contact type and may be stud-mounted or fixed to the vibrating object using a magnetic base. The inductive accelerometer has a high natural frequency ω_n and can be used in the frequency range ω such that $0 \leq \omega/\omega_n \leq 0.7$. The measured relative motion for a seismic accelerometer is given by Eqs. (3.9)–(3.11). The transducer is provided with a damping ratio between 0.6 and 0.7 so that the measured relative motion is proportional to the acceleration of the vibrating object in the desired frequency range. In contrast to the piezoelectric accelerometer which is essentially meant for high-frequency applications,

Fig. 3.8 Photograph of inductive accelerometer (Courtesy of HBM, Germany)

this accelerometer is ideally suited for acceleration measurements at low frequencies. It may be recalled that generally displacement transducers are suggested for low-frequency measurements.

Advantages and Disadvantages: This transducer has to be used in conjunction with a carrier frequency amplifier, as in the case of the LVDT. It responds to static and dynamic measurements and has continuous resolution and a fairly high output. An accelerometer has the advantage that its output can be integrated once to get velocity and twice to get displacement. Its disadvantage is that external magnetic fields may cause erratic performance.

Specifications of Inductive Accelerometer

Natural Frequency: 500 Hz.
Nominal acceleration: ±1000 m/s^2.
Nominal output signal: ±80 mV/V.

Applications

 (i) Measurement of ground motion.
(ii) Measurement of low-frequency accelerations (<1000 Hz) for modal analysis or machinery diagnostics.

3.3   Displacement Transducers

DID YOU KNOW that the earliest vibration measuring instruments were displacement transducers that evolved from seismographic devices? By the thirteenth century, such devices were used in the Maragheh Observatory in Persia. In 1703, the French physicist Jean de Hautefeuille built one. After 1880, most seismometers were based on those developed by John Milne, James Alfred Ewing, and Thomas Gray who worked as advisors in Japan. These seismometers used damped horizontal pendulums. In later seismographs, a large mass was suspended from a frame by a spring or a set of levers that significantly magnified the motion of a pen connected to the mass. It was only long after that devices like the linear variable differential transformer or transducers working as capacitive or Eddy current-based devices were developed.

Displacement measurements go hand-in-hand with pressure, force, and vibration measurements. Displacement sensors are those that sense the variation of the position of a body. For vibration measurements, the vibration magnitude is measured in terms of the displacement of a machine or structure and is displayed in terms of millimetres (mm) or μm of displacement. There are a large number of displacement measuring devices (vibrometers) working on different principles. Among the electrical devices, all the three passive elements, namely, resistance, inductance, and capacitance, are used for transducing displacement signals. The passive type of displacement transducer which is sensitive to changes in inductance is called the LVDT and has already been discussed in Sect. 3.2.2.

3.3.1 Eddy Current Transducer

This is a passive, non-contact type of transducer and uses the effect of Eddy (circular) currents to sense the proximity of non-magnetic but conductive materials. An Eddy current is a local electric current induced in a conductive material by the magnetic field produced by an active coil. This Eddy current induces a magnetic field with opposite direction to the one from the active coil and reduces the inductance. When the distance between the target and the transducer changes, the impedance of the coil changes accordingly and can be detected by a bridge circuit. Eddy currents are confined to small depths below the conductive target surface.

Their effective depth is given by

$$\delta = \frac{1}{\sqrt{\pi f \mu \sigma}} \tag{3.16}$$

where f is the excitation frequency of the circuit, μ is the magnetic permeability of the target material, and σ its conductivity. The thickness of the target material should be at least thrice the effective depth of the Eddy currents.

The pickup consists of a ferromagnetic core on which two coils are wound: one (active) main coil, which senses the motion of a conducting target and a second (balance) coil, which serves to complete a bridge circuit and provide temperature compensation. Bridge excitation is a high-frequency (about 1 MHz) alternating current. Magnetic flux lines from the active coil cut the conductive target surface, producing Eddy currents in the target; the density of the currents is greatest at the surface and becomes negligibly small about three 'skin depths' below the surface. When the target comes closer to the probe, the Eddy currents become stronger, changing the impedance of the active coil and causing a bridge unbalance which is a function of target position. This unbalance voltage is demodulated, low-pass filtered (and sometimes linearized) to produce a DC output proportional to target displacement. The high excitation frequency not only allows the use of thin targets, but also provides good system frequency response (up to 100 kHz). Figure 3.9 shows a schematic of the transducer with the bridge circuit arrangement and a photograph.
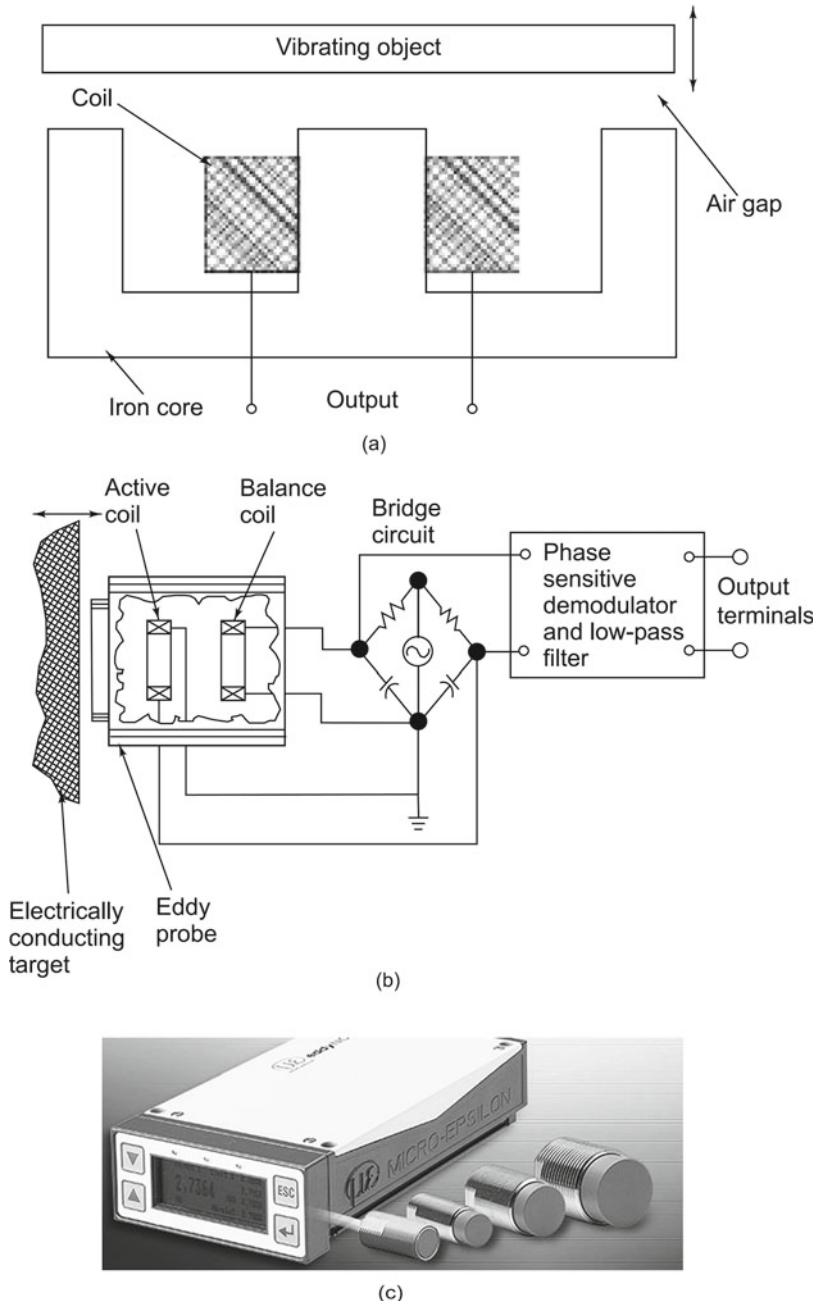
Fig. 3.9 Eddy current transducer (proximity probe): **a** schematic of transducer, **b** transducer with bridge circuit, **c** photograph (Courtesy of https://www.micro-epsilon.co.uk)

Probes are commercially available with full-scale ranges from about 0.25 to 30 mm (probe diameter 2–76 mm), non-linearity of 0.5%, and a maximum resolution of 0.0001 mm. Targets are not supplied with the probes since majority of applications involve non-contact measurement of existing machine parts, the part itself serving as target. Since target material, shape, etc. influence the output, it is necessary to statically calibrate the system every time with the specific target to be used. For non-conductive targets, one must fasten a piece of conductive material of sufficient thickness to the surface. Commercially available adhesive-backed aluminium foil tape is convenient for this purpose. The recommended measuring range of a given proximity probe begins at a 'stand-off' distance equal to about 20% of the probe's specified range, i.e., a probe rated at 0–1 mm range should be used at target to probe distances of 0.2–1.2 mm.

Advantages

  (i) High sensitivity.
 (ii) Flexibility to be used for static and dynamic measurement.
(iii) Non-contacting measurement.
(iv) High resolution.
  (v) High-frequency response.

Disadvantages

  (i) Effective distance is limited to a small range.
 (ii) The relationship between the distance and the impedance of the coil is non-linear and temperature dependent. Fortunately, it is possible to use a balance coil to compensate for the temperature effect.
(iii) This pickup works only on conductive materials with sufficient thickness. It cannot be used for detecting the motion of non-conductive specimens or thin metalized films. A piece of conductive material with sufficient thickness can be fixed on non-conductive objects to overcome this drawback.
(iv) Calibration is generally required, since the shape and conductivity of the target material can affect the sensitivity.

Typical Specifications

Size: About 2–75 mm in diameter, 20–40 mm long.
Range: 0.25–30 mm.
Resolution: Up to 0.1 μm.
Non-linearity: 0.5%.
Bridge circuit frequency: 50 kHz to 10 MHz.

Applications For measuring orbital motions of rotating shafts and various centring and alignment operations, special four-probe systems are available.
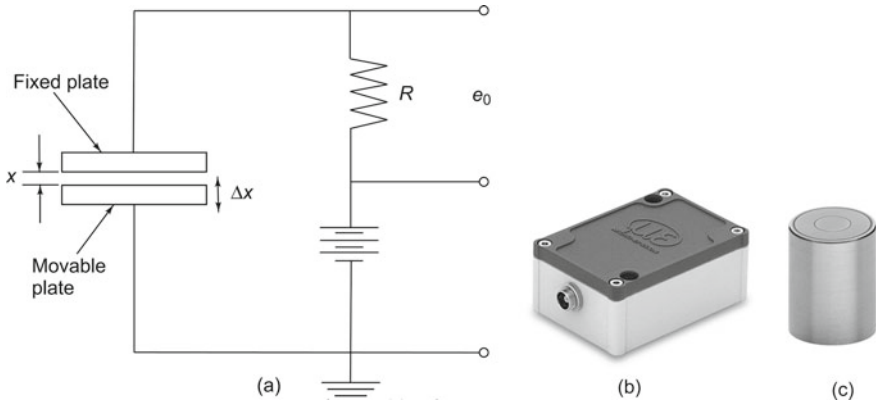
Fig. 3.10 Capacitance transducer: **a** schematic, **b** photograph controller (measuring unit), **c** close up of transducer (Courtesy of http://www.micro-epsilon.co.uk)

3.3.2 Capacitance Pickup/Condenser Vibrometer

The capacitance transducer is a passive, non-contact type of vibrometer. This parallel plate transducer (Fig. 3.10) works on the principle that the reactance of an ideal plate capacitor is inversely proportional to the distance between the plates. Its capacitance C in pF is given as follows:

$$C = 0.00886\kappa A/d \tag{3.17}$$

where d is the distance between plates (mm), A is the overlapping area (mm^2), and κ is the dielectric constant (1.0006 for air).

From this equation, it is seen that the capacitance is dependent on the three quantities mentioned above. By varying any of these quantities and measuring the capacitance, we can find out the effect of change of that quantity. Variation in spacing of the parallel plates is often used for measuring displacements if the change in spacing is less than the electrode size and gives a large measurable value of capacitance at small spacing. Signal conditioning is required to compensate for the parabolic capacitance–displacement relationship and this is easily done by measuring impedance rather than capacitance. The capacitance transducer is generally used for measurements in air and special care should be taken for measurements in liquids.

Advantages

 (i) It has excellent linearity over the entire dynamic range when area is changed.
 (ii) The technology is low cost, is stable, and uses simple conditioning circuits.
(iii) It offers freedom of electrode materials and geometry for demanding environments and applications.
(iv) Small changes in capacitance can be made large.
 (v) Capacitive sensors can be made to respond to uni-directional displacements.
(vi) Capacitors are noiseless and hence excellent SNR can be obtained.

Disadvantage Special care and calibration is required for dielectric media other than what is given in the manual.

Typical Specifications

Measuring range: 0–1 mm.
Frequency response: 0–1 kHz.
Sensitivity: 10 V/mm.

Typical Applications For non-contact measurements in lightweight and rotating machinery.

3.3.3   Fibre Optic Probe

The intensity-modulated fibre optic displacement transducer is extensively being used for displacement measurements in vibration monitoring, what with its attractive features like simplicity, good performance, versatility, and low cost. This sensor is based on the principle that, all other parameters being held constant, the light reflected back from a target surface varies with the distance that the transmitted light travels between the light source and the detector. Figure 3.11 shows a schematic diagram of this probe.

The basic principle employed here is the use of a pair of fibre optic elements, one for carrying light from a source to a target whose displacement is to be measured and the other for receiving the light reflected from the target and carrying it back to a photo-sensitive detector. By comparing the intensity of the transmitted light with that of the reflected light, the distance between the sensor and the target surface
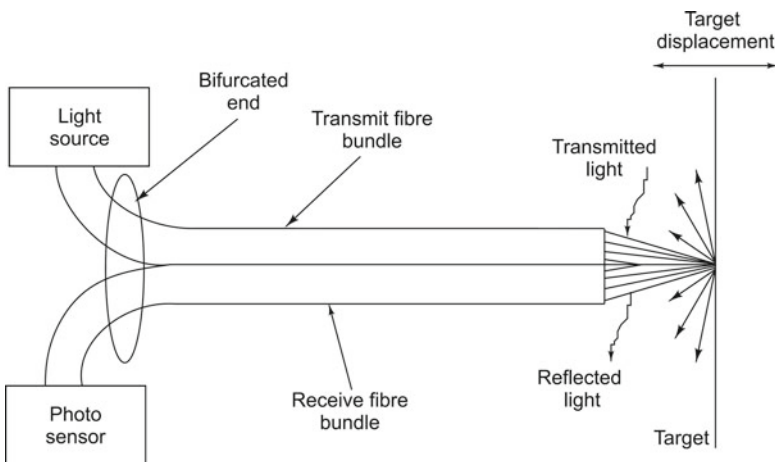
Fig. 3.11 Schematic of an external intensity-modulated fibre optic displacement sensor

Fig. 3.12 Fibre optic probe characteristics
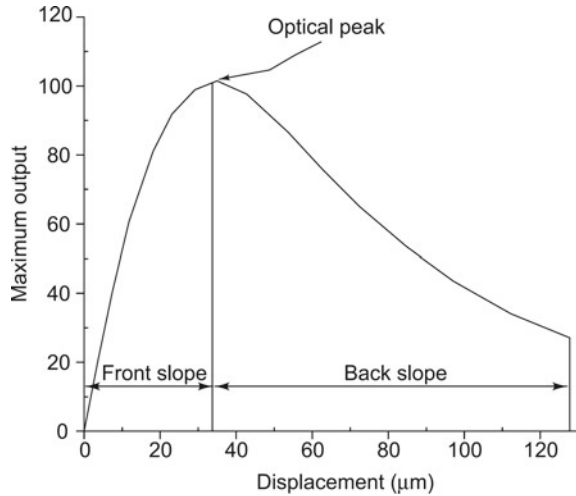
is determined. Figure 3.12 shows the relationship between the sensor output and distance from the target surface. This curve exhibits the typical 'front slope' and 'back slope' behaviour of such photo-detectors. For small sensor to target distances, more of the reflected light is captured by the receiving fibres producing the highly sensitive 'front slope'. As the target displacement increases beyond the 'optical peak', the intensity of the recaptured light decreases and produces a relatively wider 'back slope' in the output curve. Both regions are used in the calibration of the meter, the former for small gaps and the latter for large gaps. The gap at which the optical peak, or zero slope occurs, provides a calibration reference position at which the output signal can be normalized to obtain a constant sensitivity, independent of the colour or finish of the surface of the target.

In these sensors, light propagates by total internal reflection within the glass fibre, which is a flexible glass or plastic strand, capable of transmitting light along its length by total internal reflection of the light impinging at its input end. The most commonly used fibres are of the 'step index' type and consist of an inner core of high refractive index (N_1) to carry the light flux and a surrounding concentric cladding of lower refractive index (N_2) such that total internal reflection occurs. Individual fibres usually fall in the range of about 25 μm diameter to 250 μm diameter, although recent advances in the fibre optic manufacturing technology have extended the size up to about 1.5 mm. Transmission efficiency is dependent upon the composition and purity of the glass used in the core and cladding as well as the quality of the optical finish on the end surfaces of the fibres. Most commercial sensors use multiple transmit and receive fibres to obtain the high levels of intensity at the photo-detectors needed to ensure acceptable levels of performance. Fibre optic transducers are available in a great variety of sizes, configurations, and fibre distribution patterns. This offers the user a broad choice of displacement range, resolution, frequency response, and
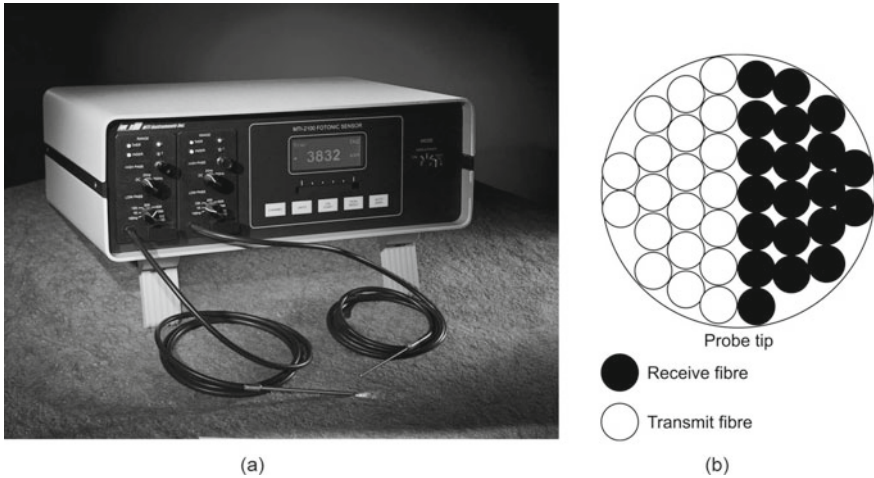
Fig. 3.13 Fibre optic probe: **a** photograph of sensor and indicator (Courtesy of https://mtiinstruments.com), **b** probe tip

physical shape. Figure 3.13a shows the photograph of a typical fibre optic meter and Fig. 3.13b the probe tip.

Advantages

(i) A major advantage of the fibre optic transducer is its ability to operate directly with a large variety of surfaces and materials from conductors to insulators.
(ii) It has inherent simplicity, versatility, and ease of use, small physical size, fast response, absence of mass loading due to its non-contacting nature, and immunity to electromagnetic interference.
(iii) It possesses many outstanding characteristics, such as high-frequency response, high resolution, high sensitivity, small measurement footprint, and reasonable cost.

Disadvantages

(i) The operating principles of these sensors make the experimental data susceptible to various measurement errors due to improper orientation of the probe, poor finish of the surface of the target, vibration of the probe clamp, etc.; particular attention has to be paid to these aspects.
(ii) These measurement errors result in varying sensitivity and affect the linearity between measurement points, which translates into inconsistencies in the measured dynamic data.
(iii) The use of these sensors in the laboratory is somewhat more complicated than that of other transducer systems and extra care must be exercised in the experimental setup and data interpretation.

(iv) The fibre optic transducer is to be calibrated for use in different media such as air, water, and motor oil.

Applications

 (i) Modal analysis of small, lightweight parts or mechanisms such as read/write head of hard disc drive.
 (ii) Condition monitoring of rolling element bearings.
(iii) Measurement of very high-frequency, small amplitude vibrations in ultrasonic devices.
(iv) Repeatability, hysteresis, and response time measurement of precision mechanisms or piezoelectric micropositioners.
 (v) Non-intrusive and remote vibration testing, especially in the presence of electric or magnetic fields.

Typical Specifications

Resolution: 0.3–12 μm.
Sensitivity: 0.012–0.75 μm/mV.
Stand-off: 0.05–5 mm.
Target footprint: 0.17–2.2 mm.
Linearity: 0.003–0.015.
Frequency response: 0–70 kHz.

3.3.4 Holography

Holography is a laser-based interferometric technique for non-contact displacement measurements. It is a relatively new measurement procedure, exploiting the characteristics of the coherent, monochromatic light beam created by a laser, for extremely accurate and effective non-intrusive measurement and for response measurement in experimental modal analysis. Holography is essentially a process which enables the recording and storage of a light wave and later the reconstruction of that light wave in complete detail. In this method, a deformed surface is illuminated with a coherent light source and the light scattered by the surface is recorded, and using this a complete three-dimensional image of the deformed object can be reconstructed later.

Figure 3.14a shows the basic holographic setup for recording. The coherent laser beam is split by a beam splitter into two beams: the reference beam and the illumination beam for the object. The reference beam is reflected onto a holographic plate (recording medium), while the illumination beam is scattered by the object surface to produce the object beam, this beam also being recorded on the holographic plate. Thus, at any instant of time, the combination of differing path lengths and incident angles between the reference beam and the object beam will cause constructive and
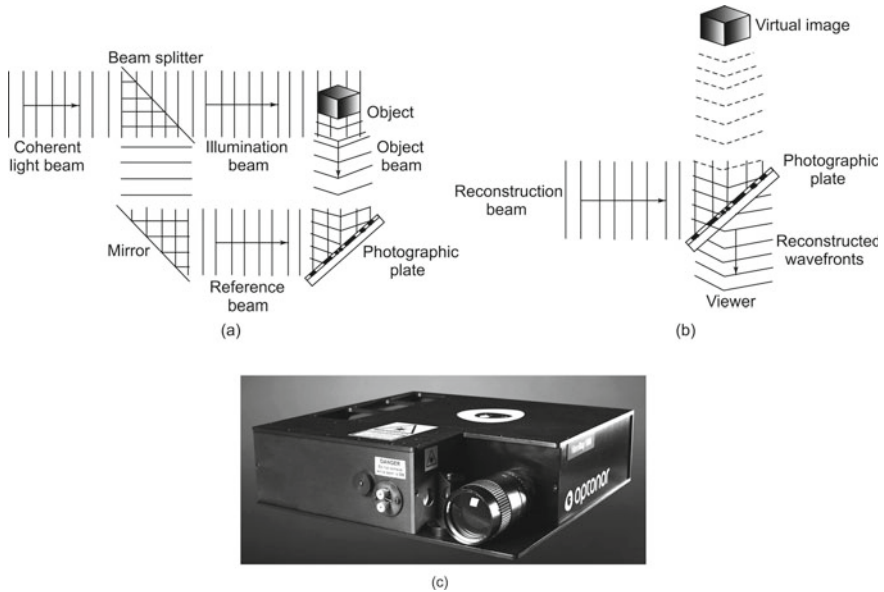
Fig. 3.14 Basic holographic process: **a** recording process, **b** reconstruction process, **c** holography equipment (Courtesy of https://optonor.com)

destructive optical interference fringes. This relative interference between the two beams is recorded to produce a photographic image called the hologram. By varying the configuration of the setup so as to vary the angle between the two beams, the distance between the fringes and thus the displacement resolution can be changed. A time-varying fringe pattern is produced when one of the beams is reflected from a vibrating target. The fringe pattern is sampled at different frequencies to obtain the dynamic target surface displacement. On re-illumination of the processed holographic plate with a reconstruction beam (reference beam similar to the original beam), diffraction from the fringe pattern on the film reconstructs the original object beam in both intensity and phase as depicted in Fig. 3.14b. Figure 3.14c shows a photograph of the holography equipment.

Of late, electronic-speckle pattern interferometry (ESPI), which is a method for increasing the speed of the basic holographic technique, is gaining attention. This is a state-of-the-art technology which is well suited for non-intrusive measurement of vibration modes in experimental modal analysis. It uses real-time video recording and displays over and above the basic interference techniques of holography. The systems presently available are able to produce resonant mode shapes only, the displacement information at a specific point on a target not being easily obtainable.

Typical Specifications

- Laser: 100 mW.
- Frequency range: 20 Hz to 50 kHz.
- Object size: 1 cm^2 to 4 m^2.
- Amplitude resolution: <1 nm.
- Maximum amplitude: 10 μm.

Advantages Holography offers the distinct advantage of being capable of measuring the entire three-dimensional vibrating surface at one time. It offers extremely high resolution, up to 2×10^{-6} mm.

Disadvantages

 (i) The system is very expensive.
 (ii) Vibration free table is required for mounting the equipment.
 (iii) Standard holography techniques suffer from slow, complex post-experimental development of the actual three-dimensional displacement diagrams.
 (iv) Numerical displacement values which are required for modal parameter extraction are difficult to obtain from these holograms.

Applications For non-destructive testing in the aerospace, automotive, and other defence related industries.

3.4 Velocity Transducers

DID YOU KNOW that the electrodynamic velocity transducer is one of the most popular vibration pickups? Electrodynamics became a field to be considered seriously following the discoveries of magnetism and electromagnetism. The accredited father of electricity and magnetism is the English scientist, William Gilbert, who in 1600 first used the terms 'electric force', 'magnetic pole', and 'electric attraction'. In 1819, Hans Christian Oersted discovered a connection between electricity and magnetism and demonstrated that an electric current caused a torque on a compass needle nearby. A number of inventions followed this, including moving coil meters to measure current and voltage. British electrician William Sturgeon in 1825 invented the first electromagnet which was a horseshoe-shaped piece of iron with a loosely wound coil of several turns wrapped around it. When a current was passed through the coil, the horseshoe became magnetized. This led to the use of electrical energy for making useful and controllable machines, loudspeakers, permanent magnet direct current (DC) motors, and many more devices based on the principle

of conversion of electrical energy into mechanical energy. These discoveries made the ingenious British inventor Michael Faraday believe that the converse was possible—that an electric current could be produced by a magnetic field, a speculation which was proved right in 1831. Thus was born the subject of electromagnetism, leading to the invention of the electric power generator. Joseph Henry of America independently discovered electromagnetic self-induction in 1831; however, the credit goes to Michael Faraday for being the first. Later, James Clerk Maxwell, Nicola Tesla, Jacques-Arsène d'Arsonval, and others contributed tremendously to this field.

Electrodynamic velocity transducers are very popular and come with most commercial vibration meters for the reasons that velocity is used for specifying vibration severity criteria and that these transducers cover a fairly wide range of frequencies. Besides, they are also suitable when the measurement frequency range is not known a priori.

3.4.1   Electromagnetic Transducer

This is a very inexpensive and useful vibration pickup of the non-contact type. It is an active transducer. In the electromagnetic transducer, a coil is wound directly on the core of a permanent magnet as shown in Fig. 3.15a. Figure 3.15b shows a photograph of the same. When a ferromagnetic body placed opposite to one of the poles of the magnet is moved with respect to the magnet, a change in the flux Φ linking the coil is obtained. This causes a voltage proportional to $\mathrm{d}\Phi/\mathrm{d}t$ to be generated in the coil as per Faraday's law. The voltage may be expressed as

$$e = k_0 \frac{\mathrm{d}\Phi}{\mathrm{d}t} \tag{3.18a}$$

$$= k_0 \frac{\mathrm{d}\Phi}{\mathrm{d}y} \frac{\mathrm{d}y}{\mathrm{d}t} \tag{3.18b}$$

$$= k_0 \frac{\mathrm{d}\Phi}{\mathrm{d}y} \frac{\mathrm{d}y(y_0 + x)}{\mathrm{d}t} \tag{3.18c}$$

$$= k_0 \frac{\mathrm{d}\Phi}{\mathrm{d}y} \frac{\mathrm{d}x}{\mathrm{d}t} \tag{3.18d}$$

where y is the instantaneous gap, y_0 the average gap between the face of the magnet and the ferromagnetic body, and k_0 is a proportionality constant. As can be seen from Eqs. (3.18a) to (3.18d), the induced voltage is directly proportional not only to the velocity of vibration, but is also dependent on the gap y. When the pickup is located near a vibrating ferromagnetic body, a voltage proportional to the velocity of vibration will be induced and as the average distance increases, sensitivity reduces.
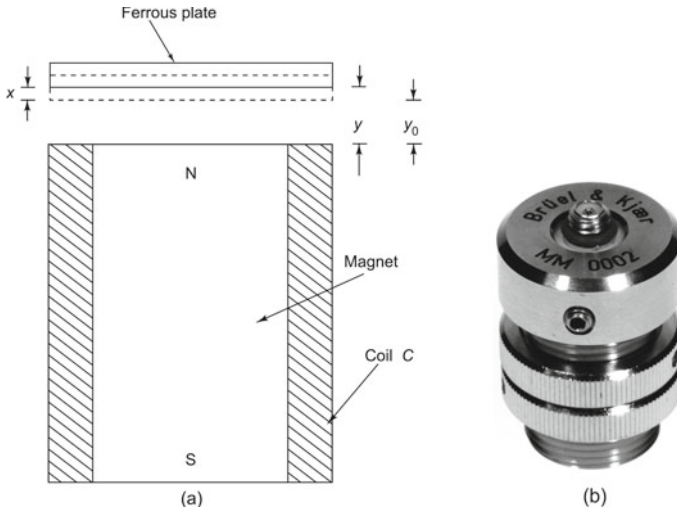
Fig. 3.15 Electromagnetic pickup: **a** schematic, **b** photograph   (Courtesy of Brüel & Kjær, Denmark)

Thus $d\phi/dy$ forms the sensitivity factor indicating that the sensitivity is inversely proportional to the gap. In principle, this method can be employed to measure absolute or relative velocities. The air gap should be at least five times the vibration amplitude for distortion to be less than 2%.

Advantages

 (i) They require no external supplies since they are of the self-generating (active) type.
(ii) High sensitivity of the order of 200 mV/(cm/s) is obtainable, without amplification.

Disadvantages Strong external magnetic fields can adversely affect the performance of the pickup.

Typical Specifications

- Sensitivity: 20 mV/(cm/s) at 1 mm air gap.
- Frequency range: DC to 1000 Hz.
- Weight: 0.5 N.

3.4.2   Laser Doppler Vibrometer

A Laser Doppler Vibrometer (LDV) is a non-contact velocity transducer based on the principle of the detection of the Doppler shift of a coherent laser light that is scattered

Fig. 3.16 Michelson
interferometry
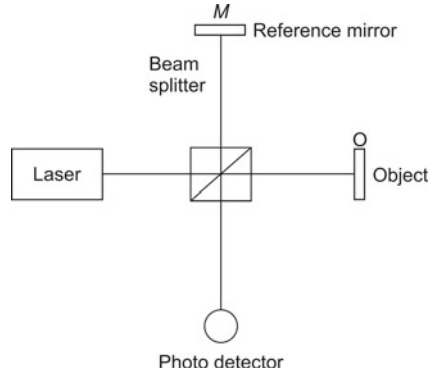
from a moving object. Light from a laser beam is made to fall on a vibrating object,
the surface vibrations of which result in scatter or reflection of the light, inducing a
Doppler frequency shift on the laser beam. This shift is directly proportional to the
component of velocity which lies along the axis of the laser beam and is measured
using an interferometer. The electronic circuitry converts the Doppler signal to an
analogue voltage proportional to the instantaneous velocity of the target. Two mov-
ing mirrors driven by galvanometric actuators are used to direct the laser beam on
to the desired measurement points. Such an instrument is called a Scanning Laser
Doppler Vibrometer (SLDV) and can quickly perform velocity measurements on a
grid of points over the structure under test.

   The main problem in this method is that the Doppler shifts obtained are usually
very small when compared to the very high laser fundamental frequency (approx-
imately 4.74×10^{14} Hz), typically 1 part out of 108. The only way to handle such
small quantities is to use interferometry. An optical interferometer is therefore used
to mix the scattered light coherently with a reference beam; the high-frequency oscil-
lations are thus combined and reduced to much lower values that can be dealt with
by standard electronics. Such an arrangement can be a Michelson interferometer as
shown in Fig. 3.16.

   In this arrangement, a laser beam is divided by a beam splitter into a measurement
beam and a reference beam. Let the distances travelled by the light from the beam
splitter to the reflectors be x_R and x_M for the reference mirror M and object O,
respectively. Let the optical phase of the reference beam in the interferometer be
$\phi_R = 2kx_R$ and that of the measurement beam be $\phi_M = 2kx_M$ where $k = 2\pi/\lambda$.
The time-varying relative phase is then $\phi(t) = \phi_R - \phi_M = 2\pi \Delta l/\lambda$ where Δl is the
vibrational displacement of the object and λ the wavelength of the laser light. The
rate of change of phase is proportional to the rate of change of position, i.e., the
vibrational velocity v of the surface. Let $I(t)$ denote the time-dependent intensity
at the point where the measurement and reference beams interfere. If Δl changes
periodically, the light intensity $I(t)$ also varies in a periodic manner. A phase change
ϕ of 2π corresponds to a displacement Δl of $\lambda/2$ giving rise to a Doppler frequency

$f_D = 2v/\lambda$. Vibrometers usually employ He–Ne lasers, giving rise to a Doppler frequency shift f_D of about 3.16 kHz for each mm/s.

The optical signal is converted to an electrical signal by a photo-detector which measures the intensity of the mixed light and the beat frequency which is equal to the difference in frequencies between the reference and measurement beams. The intensity $I(t)$ is

$$I(t) = I_R + I_M + 2K\sqrt{I_R I_M R}\cos(2\pi f_D t + \phi) \tag{3.19}$$

where I_R and I_M are the intensities of the reference and measurement beams, K is a mixing efficiency coefficient, and R is the effective reflectivity of the surface.

It is seen that $I(t)$ is a frequency-modulated (FM) signal, and, by demodulating this signal, the amplitude of v may be obtained; however, information regarding the direction of the surface velocity is missing because of the cosine function. The most common solution to obtain directional sensitivity is by introduction of an optical frequency shift into one arm of the interferometer to obtain a virtual velocity offset. Figure 3.17 shows the schematic of an LDV setup where an acousto-optic modulator (Bragg cell) is incorporated into one arm of the interferometer.

The Bragg cell is driven at frequencies of 40 MHz or higher and generates a carrier signal at the drive frequency. The frequency of the motion of the object modulates the carrier signal. The object velocity (vector) determines the direction and amount of frequency deviation with respect to the centre frequency f_B.

With the introduction of a shift frequency f_B, the intensity at the detector changes to

$$I(t) = I_R + I_M + 2K\sqrt{I_R I_M R}\cos(2\pi(f_B - f_D)t + \phi) \tag{3.20}$$

For signal decoding, one can process the phase to produce a displacement output or carry out an FM demodulation to provide the vibrational velocity. Figure 3.18 shows a photograph of a commercially available LDV.

Typical Specifications

- Laser power: <1 mW.
- Maximum velocity: 1–10 m/s.
- Upper frequency limit: 200 kHz.
- Resolution: 1 μm/s.
- Maximum stand-off distance: 10s of m.

Applications When mounted on to a tripod, it can be easily pointed at a vibrating object. This technique is effectively used in structural dynamic testing, online monitoring of industrial plants, fault detection, etc. Besides, laser vibrometers coupled with scanning systems open up new possibilities, such as in the field of measurements in tracking mode on moving objects.
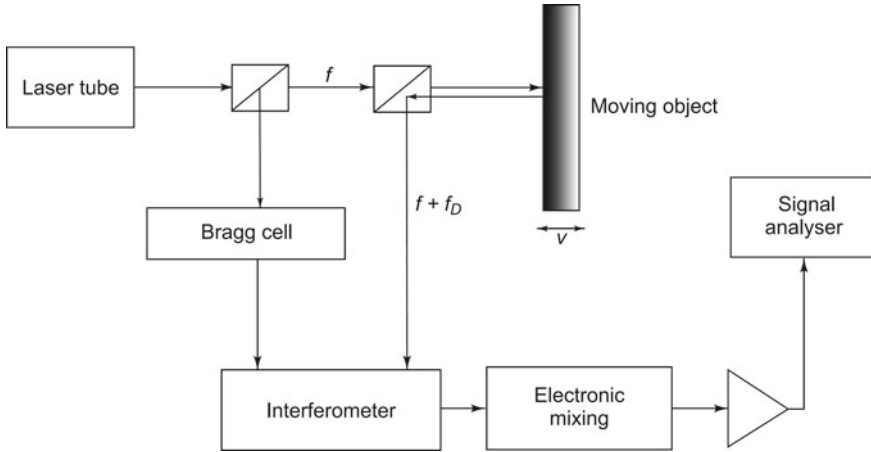
Fig. 3.17 Schematic of an LDV

Fig. 3.18 Laser Doppler vibrometer (Courtesy of Brüel & Kjær, Denmark)

Advantages

(i) The single-point LDV can measure velocity and displacement of vibrating struc-
tures completely without contact.

(ii) Most LDVs have low power levels so that no special safety measures are required.

(iii) Results are storable in digital formats like BMP and JPG images or AVI movies
or TXT text data files.

3.5   Acceleration Transducers

INTERESTING FACTS: Accelerometers are transducers with an unimaginable variety of applications. In mechanical engineering, they are used for monitoring vibrations in rotating machinery such as turbines, pumps, fans, compressors, etc.

Highly sensitive accelerometers are used in inertial navigation systems in aircrafts, missiles, and space vehicles. Accelerometers are also used in drones for flight stabilization. They also find application in the measurement of vibration of automobiles and also for suspension control, brake control, fuel cut-off, and engine knock off. They are also used for measuring the vibration of structures like buildings and bridges and in experimental modal analysis. They are used in smart phones, tablet computers, and digital cameras so that images on screens can be rotated based on the way the device is being held. Accelerometers that measure gravity are called gravimeters. High-frequency recordings of acceleration allow tracking of behavioural patterns of land and marine animals in biological applications. MEMs based accelerometers are extensively used in airbag deployment systems for modern automobiles and in electronic stability control systems. Accelerometers are increasingly being used in portable electronic devices and video game controllers. They are used for the measurement of the motion of patients with diseases such as Parkinson's and in sports watches for runners to help determine the speed and distance of the runner. They can also be used to measure seismic activity.

3.5.1   Piezoelectric Accelerometer

Piezoelectric transducers are the most widely used in the world of shock and vibration, the reason being their very wide frequency response to amplitude and phase. These pickups involve a class of materials which, when mechanically deformed, produce an electric charge. They are very small (extending down to a fraction of a gram) and are of the active type, producing charge variations as output. Unlike other sensors, piezoelectric transducers have a reversible effect, deflecting mechanically when subjected to an applied voltage.

The piezoelectric transducer uses barium titanate or lead zirconate titanate (PZT) or lead niobate or lithium niobate or other suitable piezoelectric crystals as the basic element. This crystal is placed inside a casing with a hard spring pressed against it via a block of mass (Fig. 3.19). This stiff spring-light mass combination has a high natural frequency, around 20,000 Hz and is used as an accelerometer. When the whole assembly is subjected to vibration, the mass exerts a variable force on
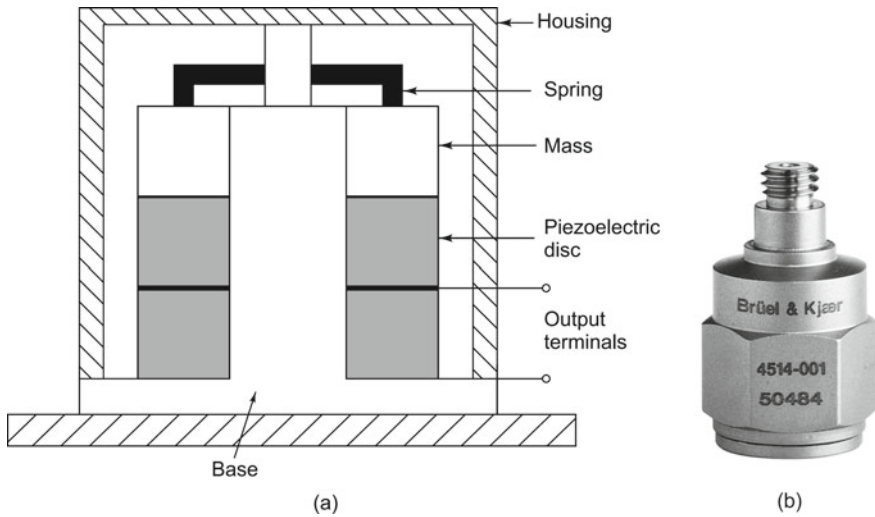
Fig. 3.19 Piezoelectric transducer: **a** schematic, **b** photograph   (Courtesy of Brüel & Kjær, Denmark)

the piezoelectric crystal and the latter in turn develops a voltage change across its faces. For a wide range of frequencies, the spring-mass system develops a voltage or charge proportional to the acceleration to which the transducer is subjected. As the transducer has high impedance, a charge amplifier would be essential to use it to its full capability. The piezo pickup is a contact type of absolute measuring device. It is generally fixed by stud mounting or with a magnetic base. The smaller devices are fixed using bee's wax.

The sensitivity of the crystal accelerometer is given either in terms of charge (pC) per g or in terms of voltage per g (mV/g). Since the voltage V and charge Q are related by the equation $V = (Q/C)$, the capacitance of the crystal C, including the shunt capacitance of the connecting cable, must be specified. Typical sensitivity of a crystal accelerometer is 20 pC/g, with a crystal capacitance of 500 pF. The equation $V = (Q/C)$ then gives the sensitivity in terms of voltage as $20/500 = 0.04$ V/g or 40 mV/g. This severe loss of signal due to cable shunt capacitance can be avoided by using a charge amplifier.

Today we have commercially available 'ICP' accelerometers where ICP is PCB's registered trademark standing for 'Integrated Circuit—Piezoelectric'. This class of transducers has built-in signal conditioning electronics, which converts the high-impedance charge signal produced by the piezoelectric crystal into a usable low-impedance voltage signal. Such a signal can be easily transmitted over ordinary two-wire or coaxial cables over long distances, to any voltage measuring or recording device. The special electronics, typically, also has incorporated in it other signal conditioning features such as gain, filtering, and self-test features. Since these ICP

sensors are simple to use, have high accuracy, wide frequency range, and low cost, they are very popular in many shock and vibration applications.

Advantages

 (i) One of the most popular transducers since it can ideally be used with portable instruments.
 (ii) Wide frequency range.
(iii) Very low mass loading due to small weight.
(iv) Readings not affected by extraneous magnetic fields.
 (v) Very useful for experimental modal analysis since this throws open the roving pickup option.

Disadvantages

 (i) Not suitable for low frequencies.
 (ii) Not meant for high-temperature applications (for this special probes are required).
(iii) Requires charge amplifier.

Typical Specifications

Voltage sensitivity: 50 mV/g.
Charge sensitivity: 55 pC/g.
Resonance frequency: 30 kHz.
Transverse sensitivity: <3%.
Dynamic range: ±3000 g.
Mass: 20 gm.

Typical Applications For general-purpose shock and vibration measurements, modal analysis, machinery diagnostics.

3.5.1.1 Universal Serial Bus (USB)-Based Accelerometer

There is a new range of accelerometers available in the market with special USB cables for fully integrated digital data acquisition. These are typically high-resolution, broadband piezoelectric accelerometers. They are found to be useful in predictive maintenance and condition monitoring applications. These are designed to provide high-quality, reliable vibration measurements and can be directly connected to a personal computer (PC), smartphone, or tablet as shown in Fig. 3.20. The process of vibration recording is as simple as recording audio input in a PC or phone using a simple USB connection. These accelerometers are packaged in stainless steel, hermetically sealed housings with integral cables. Some accelerometers that come with USB adaptors are also of the MEMS type.

Fig. 3.20 USB cable-based accelerometer (Courtesy of www.digiducer.com)
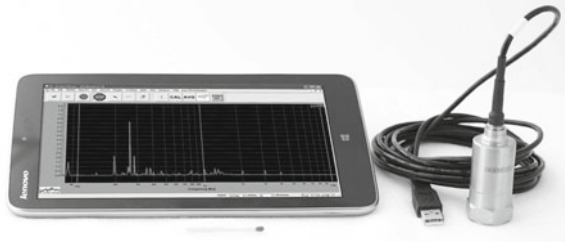
Advantages

- Can be used with a wide variety of software.
- Vibration data can be recorded using a laptop, phone, or tablet.
- They work with any modern operating system including Windows, Linux, Android, iOS, and MAC OS.
- No data acquisition box is needed.

Disadvantages

- Same as for any piezoelectric accelerometer.

Typical Specifications

- Various sensitivities available: 100, 50, and 25 mV/g.
- Measurement range: ±20 g.
- Frequency range (±3 dB): 1.6–10000 Hz.
- Resonant frequency: 25000 Hz.
- Temperature range: −65 to +250°F.

3.5.2 Microelectro-Mechanical Systems (MEMS)-Based Accelerometers

Modern machinery diagnostic systems employ a new family of sensors called micro-machined accelerometers. Micromachining refers to the technique of manufacturing tiny moving mechanical structures on a single wafer of, say, silicon. Techniques developed from the manufacture of large-scale integrated (LSI) semi-conductor electronic components such as deposition, photolithography, and etching are employed for making micromachined devices. The new accelerometers include both the signal conditioning circuitry and the sensor, fabricated together on a single monolithic chip at a very low cost with high reliability and make them the perfect candidates for modal testing applications. Previously, modal testing was limited by the cost of the accelerometers and of the associated signal conditioning circuitry, such as charge amplifiers and the high cost of the setup. Apart from the sensors, the cost of calibration was also responsible for the increased cost of testing. The mass of the standard
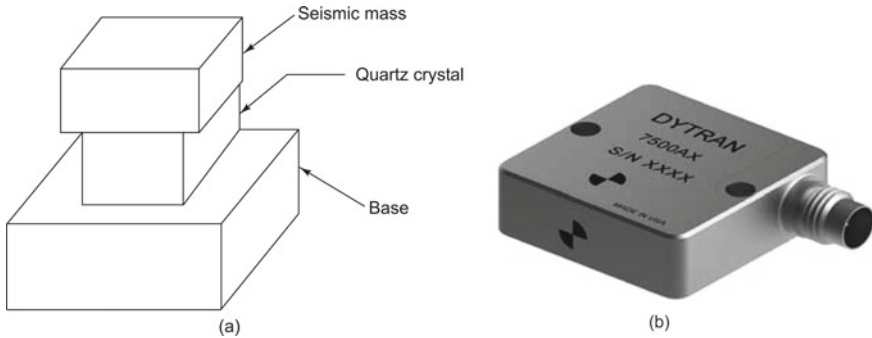
Fig. 3.21 MEMS accelerometer: **a** schematic, **b** photograph (Courtesy of https://www.dytran.com)

accelerometers, i.e., 100–200 gm often produced mass loading effects. Reduced cost of modal testing with improved accuracy is now possible due to MEMS accelerometers, since a larger number of sensors with a lower cost can be used. The sensitivity of the micromachined accelerometers is within 1% of the full scale and its shock survival is approximately of the order of 2000 g.

A piezoelectric MEMS sensor (Fig. 3.21a) is very similar in construction to a conventional piezoelectric accelerometer, except that the fabrication is by an etching process. In one design of a MEMS single-axis accelerometer, a proof mass is supported on a frame, by means of one or more compliant beams. When acceleration is applied to the frame, each beam is deformed by the force required to accelerate the proof mass, causing it to move relative to the frame. The motion of the proof mass is controlled by the resilient nature of the beams, which apply a restoring force on the mass. Acceleration can be measured by sensing the strain in each beam that supports the proof mass, typically using either piezoelectric or piezoresistive sensors present on each beam. An alternative design of MEMS accelerometer uses a capacitive measurement technique. Here a proof mass is supported by one or more resiliently deformable beams, the mass carrying one plate of a capacitor and the frame carrying the other plate. Acceleration is sensed by measuring the change in capacitance caused by the deflection of the inertial mass which changes the capacitance between the two plates. A third form of MEMS accelerometer uses a torsion member to constrain a proof mass and a capacitive or servo-capacitive arrangement is used to measure the displacement of the mass when the accelerometer is subjected to acceleration. The torsional stiffness of the support member controls the displacement of the proof mass or electrostatic forces generated by the servo-capacitors control that displacement. Figure 3.21b shows the photograph of an MEMS accelerometer.

Conventional triaxial accelerometers usually consist of an assembly of three single-axis accelerometers, arranged with their sensing axes orthogonal to each other, making the resultant three-axis accelerometer significantly larger and heavier than desirable. MEMS technologies have led to the production of devices which are extremely small and yet are capable of giving very accurate and reliable indications of

acceleration in three dimensions. The required alignment accuracy can be achieved using lithographic etching processes, and no subsequent assembly processes are required to complete the basic structure of the three-axis accelerometer.

Broadly, there are three types of triaxial MEMS accelerometers. In the first arrangement, three separate single-axis MEMS accelerometers are mounted onto three faces of a cube, to measure accelerations in three directions. The whole assembly of the individual wafers and the mounting cube significantly increase the weight of the complete three-axis accelerometer. Further, difficulties in aligning the three accelerometers with great accuracy result in significant manufacturing difficulties and so there is a cost penalty. In the second arrangement, an MEMS accelerometer with a single mass is used to sense acceleration in three orthogonal directions. Ideally, the sensitivities in all directions should be equal, but in practice the out-of-plane response (with respect to the wafer) is usually several times larger than the in-plane response. Isolation of the individual signals for each direction is limited by the accuracy of manufacture of the device and the requirement for equal signals from each axis, leading to cross-axis signals. The performance of such a device is consequently compromised. In the third case, three identical single-axis MEMS devices are produced in a single wafer, to sense accelerations in three directions, but a typical three-axis accelerometer manufactured thus cannot produce exactly the same strain distribution in the support beams for the proof mass in response to in-plane and out-of-plane accelerations.

Advantages

 (i) These are high-performance, high-accuracy, high-reliability, and low-cost accelerometers with low power consumption.
 (ii) They have very low mass, giving rise to very low mass loading.
(iii) They are suitable for measurements in harsh vibration environments, having high shock withstand capability and high-temperature operation.
(iv) These include both the signal conditioning circuitry with low noise electronics and the sensor, fabricated together on a single monolithic chip at a very low cost.
 (v) They are suitable for measurements at low frequencies down to DC.
(vi) Besides, they are immune to electromagnetic interference.

Typical Applications These accelerometers are the perfect candidates for modal testing involving roving transducer applications. Other applications include crash testing, air bag testing, robotics, seismic monitoring, and tilt measurement.

Specifications

Acceleration range: ± 50 g.
Output voltage: ± 2.0 V DC @ ± 50 g.
Sensitivity: 40 mV/g, $\pm 5\%$.
Bandwidth: DC–1000 Hz.
Cross-axis sensitivity: $<2\%$.
Shock: >1000 g.

3.5.2.1   Wireless Accelerometers

There are many measurement applications in engine testing, landing gear, and in-flight status monitoring that require radio-frequency (RF) wireless connections instead of physical wiring. These applications require continuous acquisition, transmission, and processing of acceleration data for further analysis. Often the sensor is required to be installed at remote, difficult-to-access, or hostile locations and clearly physical cables become impossible. Most of the current-day wireless accelerometers use MEMS and have simplified installation. However, these MEMS accelerometers have low measurement range, low sensitivity, large size, and susceptibility to noise. Besides there are aspects to be considered like limited frequency bandwidth, environmental conditions like high temperature, power for the device as well as for wireless transmission and communication for dynamic signal measurement. They are designed such that a number of devices may be networked together, each operating as a measurement node having its own processor, memory, and radio. The sensor network systems also have their gateways and software programs with simple user interfaces.

The signals are typically transmitted in packets by a radio-frequency transmitter to minimize the risk of data loss. The radio device is generally a low-energy module with high-speed data transfer, very low power consumption, and extremely robust connection. The wireless sensor will have provisions for the operator to initiate and terminate the transfer of data, check connectivity, and to do basic on-board diagnosis, in addition to an alarm system for any connection loss. Power consumption is a critical challenge while using wireless sensors. Since the wireless accelerometer could be located very remotely, battery life is very important. Power consumption is highest while data is being transmitted; hence, to extend battery life, management of wireless communication is important. Researchers are also considering techniques such as energy harvesting to make the wireless sensors more efficient and usable. Figure 3.22 shows a typical wireless accelerometer.

Advantages

- Measurement in remote or hazardous environment.
- A number of devices may be networked together in the form of nodes.

Fig. 3.22 Wireless accelerometer (Courtesy https://wireless-sensors.co.uk)

Disadvantages

- Batteries may have to be changed frequently or alternate power sources have to be identified.

Typical Specifications

- Sensitivity: ($\pm 10\%$) 50 mV/g.
- Frequency range: 0–800 Hz.
- Sensing element: Ceramic.
- Measurement range: ± 8 g.
- Weight: 100 gm.
- Board storage of 2 MB.
- Transmission frequency: 900 MHz.
- Battery: 1–2 years.
- Range: 100 m.

Typical Applications In engine testing, and in condition monitoring at remote locations.

3.5.3 Servo Accelerometer

These devices were originally designed for use in inertial guidance systems in aircraft and missiles, but have subsequently found application in the measurement of vibrations of large structures which poses special problems. Such structures have low natural frequencies and the associated accelerations are very small. This demands that the accelerometer be very sensitive and suitable for use right from near-zero frequency. But unlike many other applications, the motion of large structures is relatively immune to the loading of the accelerometer mass, allowing the use of accelerometers which are heavier than piezoelectric devices. The servo accelerometer or force-balance accelerometer is ideally suited for such applications.

A servo accelerometer measures the acceleration of the structure on which it is mounted by measuring the force required to prevent the proof mass from moving relative to the instrument frame under acceleration. A schematic drawing of the servo accelerometer is shown in Fig. 3.23a. Figure 3.23b shows the photograph of a commercially available servo accelerometer. A displacement transducer is used to sense the relative movement of the mass with respect to the instrument frame. This signal is amplified and fed back as direct current to the force coil suspended in a magnetic field, generating the required restoring force for equilibrium. Hence, the current required to prevent the mass from moving is a measure of the input acceleration (on the frame) along the direction in which the mass is free to move.

With respect to Fig. 3.23a, we have

$$m\ddot{x}_0 = kd + F_0 i_c + c\dot{d} \tag{3.21}$$
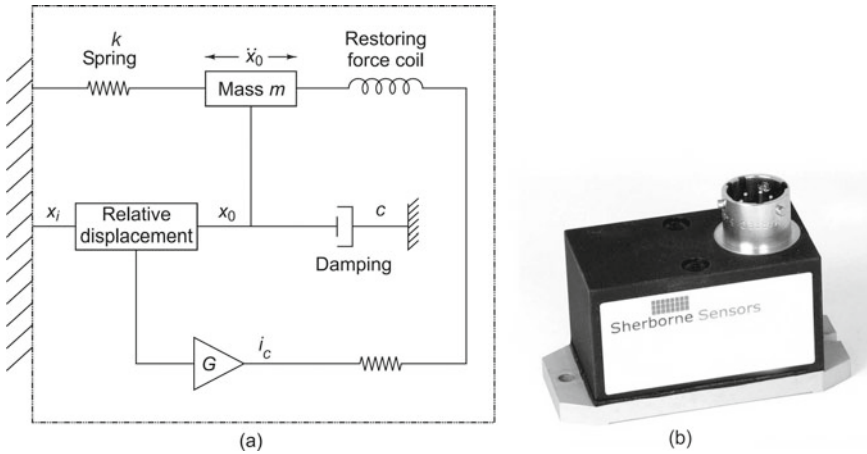
Fig. 3.23 Servo accelerometer: **a** schematic arrangement, **b** photograph (Courtesy of https://www.sherbornesensors.com)

Here k is the linear spring rate, F_0 is the force produced on the mass per unit current, c is the viscous damping coefficient, and i_c the current. x_i is the displacement of the casing, x_0 is that of the mass, and d is the relative displacement given by

$$d = x_i - x_0 \tag{3.22}$$

Therefore,

$$m\ddot{x}_i = kd + F_0 i_c + c\dot{d} + m\ddot{d} \tag{3.23}$$

For the condition that $i_c = Gd$, where G is the gain of the amplifier,

$$m\ddot{x}_i = \frac{m}{G}\frac{d^2 i_c}{dt^2} + \frac{c}{G}\frac{di_c}{dt} + \left(\frac{k}{G} + F_0\right)i_c \tag{3.24}$$

For $k \gg F_0 G$, i.e., for large values of G, it can be proved that the sensitivity is given by

$$\frac{i_c}{\ddot{x}_1} = \frac{m}{F_0} \tag{3.25}$$

Advantages

(i) It has high accuracy and reasonable frequency response.
(ii) A flat frequency response up to about 2000 Hz can be obtained in these devices with high sensitivities.
(iii) It can be used for low-frequency measurements.

Disadvantages

 (i) Extremely high cost as compared to conventional open-loop accelerometers.
(ii) It has limited dynamic range.

Typical Specifications

Frequency range: 0–1000 Hz.
Acceleration range: ±500 g.
Sensitivity: 0.1–10 mV/g.
Linearity: 0.02%.

Applications For vibration measurements of huge structures, typically in guidance and navigation systems.

3.6 Rotary Vibration Transducers

DID YOU KNOW that torsional vibrations are angular vibrations of an object, say a shaft about its axis of rotation? These vibrations are sensed mainly as the variation in rotational speed within one cycle of rotation. These speed variations are generally caused by a varying driving torque or by a varying load in a machine. The drive line may then amplify and transfer these vibrations to other structural parts, causing issues with respect to comfort, durability, or efficiency. The amplitude of torsional vibration is influenced by factors such as material properties, as well as operating conditions like temperature, load, and speed. Torsional vibration can be represented as angular displacement, velocity, or acceleration. All these are related to each other and one can be obtained from the other by integration or differentiation. There exist a number of transducers for measurement of torsional vibration. They are based on incremental encoders, optical devices, and magnetic devices including inductive sensors and Hall effect sensors.

3.6.1 Rotary Variable Differential Transformer (RVDT)/Torsional Vibration Pickup

This is a seismic type of transducer. The principle of operation is similar to that of the LVDT or inductive accelerometer, except that it is used to measure angular displacements. Whereas the LVDT uses a cylindrical iron core, the torsional vibration pickup uses a rotary ferromagnetic core and instead of a linear spring it uses a torsional spring. The shaft, the torsional vibrations of which are to be measured, is attached to the core and the magnet (which is similar in function to the seismic mass) is free
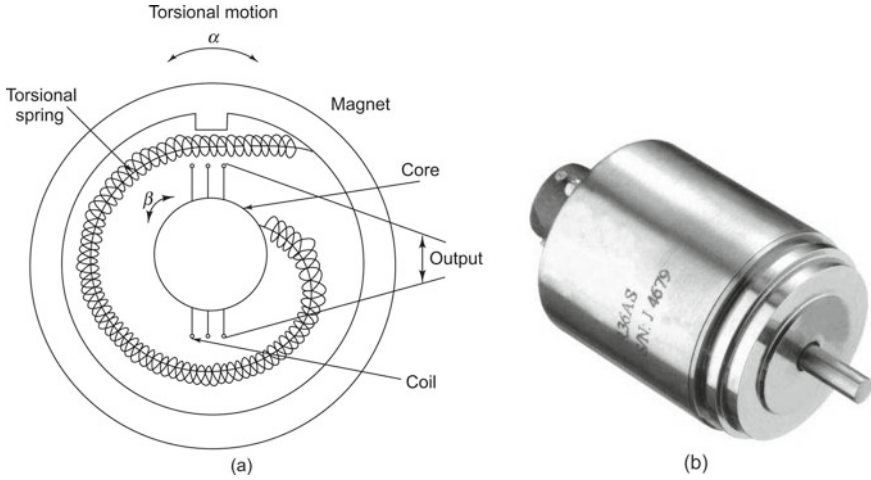
Fig. 3.24 RVDT vibration pickup: **a** schematic, **b** photograph (Courtesy of https://www.te.com)

to oscillate. The relative angular motion of the core with respect to the magnet gives an output signal proportional to torsional vibration velocity. Figure 3.24 shows the principle of operation and photograph of this transducer.

Let the angular motions of the magnet and the core be α and β, respectively. Let $\gamma = \alpha - \beta$ be the relative motion between the core and the magnet. The equation of motion of the system shown in Fig. 3.24 may be written as

$$J\ddot{\alpha} = -s(\dot{\alpha} - \dot{\beta}) - q(\alpha - \beta) \tag{3.26}$$

where J, s, and q are the mass moment of inertia of the shaft, the torsional damping, and the torsional stiffness, respectively. Assuming sinusoidal motion $\beta = B \sin \omega t$ for the torsional motion, we obtain the equation

$$J\ddot{\gamma} + s\dot{\gamma} + q\gamma = J\omega^2 B \sin \omega t \tag{3.27}$$

The steady-state solution is
$$\gamma = \Gamma \sin(\omega t - \phi) \tag{3.28}$$

The frequency response function Γ/B is a complex quantity with magnitude and phase lag ϕ as shown below:

$$\frac{\Gamma}{B} = \frac{\eta^2}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \tag{3.29a}$$

$$\tan \phi = \frac{2\zeta(\omega/\omega_n)}{1 - (\omega/\omega_n)^2} \tag{3.29b}$$

where $\zeta = \omega/\omega_n$ is the frequency ratio and ζ is the damping ratio.

Equations (3.29a) and (3.29b) are depicted by Fig. 3.4 as a function of η. ζ of the order of 0.65–0.7 is chosen as before.

Advantages

  (i) It is of relatively low cost.
 (ii) It is robust and is capable of working in a wide variety of environments.
(iii) It has a high signal-to-noise ratio.
(iv) It has negligible hysteresis.
 (v) Its theoretical resolution is infinitesimal, being limited only by the resolution of the amplifiers and voltmeters used to process the output signal.

Disadvantages

  (i) The fixing and calibration of this torsional transducer requires substantial machinery downtime.
 (ii) It is affected by extraneous magnetic fields.

Applications This transducer is ideal for the following:

  (i) Industries such as the automotive, aerospace, petrochemical, marine, and power industries, where even small speed variations can lead to poor machine performance, premature failure, and human discomfort.
 (ii) General torsional vibration measurements and analysis on rotating components in engines, motors, prime movers, pumps, compressors, couplings, shafts, and dampers, marine propulsion systems, and structural analysis on rotating components.
(iii) Relative torsional vibration measurements during coast-up and coast-down measurements, order analysis, dynamic torque-loading influence measurements, and shaft speed measurements.

3.6.2 Laser-Based Torsional Vibration Transducer

A laser-based torsional vibration transducer is an easy to use, highly accurate and reliable vibration measuring system. A non-contact torsional vibration transducer using the state-of-the-art laser technology may be used for making torsional vibration measurements where it is not feasible to mount a transducer onto a rotating object. Using a dual beam laser transducer, the instantaneous changes in angular velocity, and, by integration, in angular displacement, of a rotating component may be found from the frequency difference of the retro-reflected, Doppler-shifted beams. A schematic of such an instrument is shown in Fig. 3.25a and a photograph in Fig. 3.25b. At the heart of the system is typically a low power (<1.5 mW), Ga-Al-As laser. The laser beam is split into two equal-intensity parallel beams separated by a distance, d given by

$$d = R_A \cos \alpha_A + R_B \cos \alpha_B \tag{3.30}$$
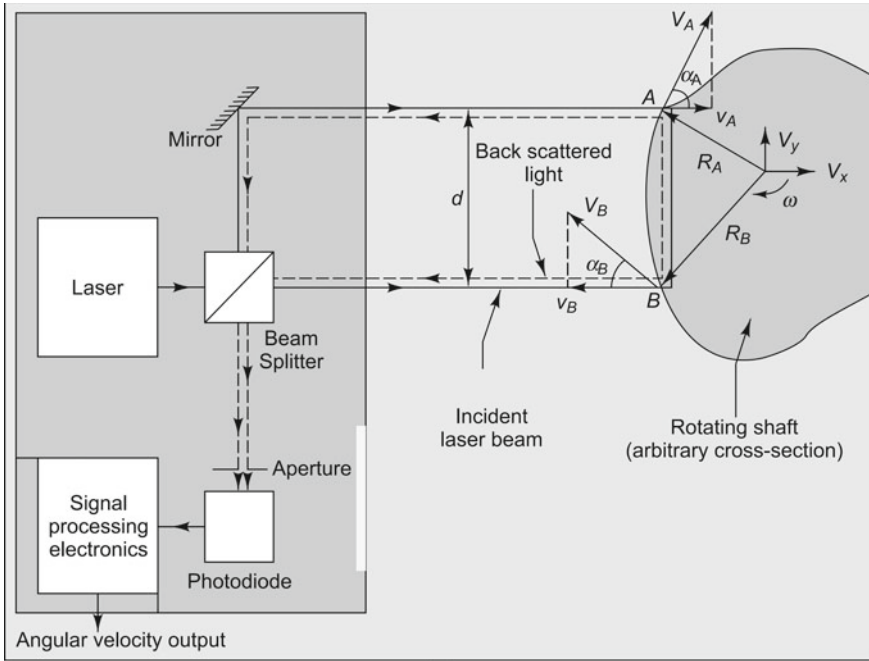
Fig. 3.25 Laser torsional vibration transducer: **a** schematic, **b** photograph (Courtesy of Brüel & Kjær, Denmark)

The beams strike the surface of the rotating shaft at points A and B which have velocities V_A and V_B, respectively. Each beam sees only the velocity in the x-direction.

$$v_A = -V_A \cos \alpha_A - V_x = -\omega R_A \cos \alpha_A - V_x$$
$$v_B = -V_B \cos \alpha_B - V_x = -\omega R_B \cos \alpha_B - V_x \tag{3.31}$$

and is frequency shifted as shown below:

$$f_A = 2\frac{v_A}{\lambda} \quad \text{and} \quad f_B = 2\frac{v_B}{\lambda} \tag{3.32}$$

The reflected beams heterodyne, giving an output current modulated at the beat frequency which is the difference between the frequencies of the Doppler-shifted beams as shown below:

$$f_D = f_B - f_A = \frac{2}{\lambda}(v_B - v_A) = \frac{2\omega d}{\lambda} \tag{3.33}$$

The equation indicates that the beat frequency is directly proportional to the shaft speed (ω) and inversely proportional to wavelength λ; it is independent of any rigid body motion of the shaft. These equations hold good as long as the plane of the laser beams is perpendicular to the axis of the shaft.

In practice, the laser transducer is mounted on its tripod or held in the hand. A strip of reflective tape is wound around the target. The laser is made to point at the tape and readings are taken or the output is sent to a frequency analyser, together with a tacho signal, to perform order analysis. The optimum measuring distance from the laser transducer to the target is 5–50 cm.

Advantages

 (i) Measurements are independent of target cross section.
 (ii) Remote measurements up to 0.5 m are possible.
(iii) It is simple and quick to use.
(iv) It is compact, lightweight, and portable.
 (v) Battery-powered, hand-held operation is possible.

Specifications

  Angular vibration velocity: 0.3–10,000 °/s (peak).
  Angular vibration displacement: 0.01–17′.
  Rotational speed: 30–7200 RPM.
  Rotational vibration displacement: 0.3–12′.
  Frequency range: 0.3–1000 Hz.
  Laser: Ga-Al-As diode producing 780 nm light (Class 3B Laser).
  Output power: <1.5 mW.
  Operating distance: 5–50 cm.
  Laser spot diameter: Less than 1 mm.

Applications The laser-based torsional vibration transducer is useful for all the cases mentioned in Sect. 3.6.1. Besides, it is ideal for non-contact angular vibration velocity and angular vibration displacement measurements.

3.7 Smart Sensors (Plug-and-Play Devices) and Transducer Electronic Data Sheets (TEDS)

DID YOU KNOW that certain types of transducers are classified as smart sensors? They contain a smart transducer interface module (STIM), which, in turn, contains an on-board Electrically Erasable Programmable Read-Only Memory (EEPROM) IC, called Transducer Electronic Data Sheet (TEDS), using which the sensor can identify and describe itself to the network and/or to a smart signal conditioning device, thereby facilitating automatic system configuration. TEDS, or Transducer Electronic Data Sheet, is a set of electronic data in a standardized format defined within the IEEE 1451 standard giving these sensors the name 'Plug-and-Play' sensors. TEDS electronically stores information about the transducer's characteristics such as type of device, manufacturer, model number, serial number, calibration date, sensitivity, reference frequency calibration coefficients, and frequency response for a transducer, in addition to interface details such as bridge type, excitation, etc. in terms of a table or an algorithm. It also provides read-and-write functions for accessing TEDS and transducer data. The specification also defines the number of samples acquired for one command, which varies from 0 to 65,535 samples per set. The signal conditioners and other interface hardware used with these smart sensors must provide for an option or include circuits for TEDS communications.

TEDS capability was originally intended for piezoelectric sensors such as accelerometers and pressure sensors, but it now includes all common analogue sensors and actuators, such as MEMS accelerometers, pressure transducers, and temperature sensors with two-wire (analogue) and mixed-mode (analogue and digital) input/output. For two-wire analogue sensors and actuators (called Class 1), the output signal is generally coupled to the signal conditioner, while the TEDS are enabled and read out with a DC bias voltage applied to the same two wires. The transducers with mixed-mode capability (called Class 2) also communicate digitally with the TEDS memory. The TEDS file may be contained on board the sensor in the EEPROM, or off board in a reserved file in the data acquisition system. There are services offering conversion of non-TEDS sensors to TEDS 'Plug-and-Play' sensors using a connector, recalibration, and a programmed EEPROM. Figure 3.26 shows a smart TEDS sensor configuration.
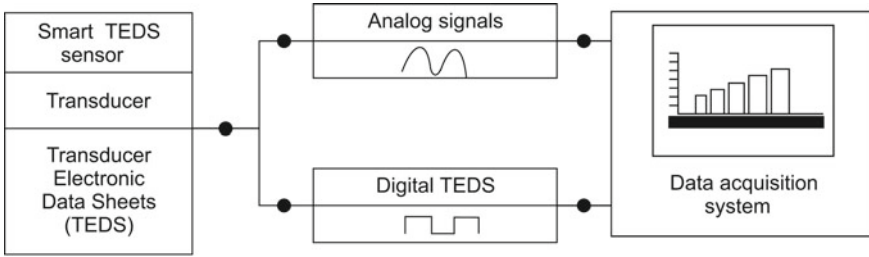
Fig. 3.26 Configuration of a smart TEDS transducer

Advantages of TEDS

  (i) Ease of troubleshooting.
  (ii) Reduced safety risks: it is no longer necessary to 'physically' access a machine for connection verification.
 (iii) Reduced costs for setup and tear-down.
 (iv) No recalibration is needed when replacing sensors.
  (v) Capability of system of recalibrating itself.
 (vi) Integration with wireless transceivers.
(vii) Feasibility of adding intelligence to sensors.

Parameters Typically Included on a TEDS Chip

Manufacturer ID	Model number	Version
Serial number	Transducer	Electrical signal type
Minimum physical value	Maximum physical value	Mapping method
Minimum electrical value	Maximum electrical value	Bridge type
Impedance of each bridge element	Response time	Measurement location
Excitation level (nominal)	Excitation level (minimum)	Excitation level (maximum)
Calibration date	Calibration initials	Calibration period (days)

3.8 Comparison of Transducers

We have seen a wide range of transducers working on various principles. Non-contact transducers are preferred for many of the vibration measurement and modal analysis applications since they cause minimal mass loading of the structure. There exists a wide range of transducers well suited for non-contact vibration response measurement as has been discussed in the earlier sections. Table 3.2 gives an assessment of transducers for non-intrusive measurement.

Table 3.2 Summary of non-intrusive vibration transducers

Characteristics	Proximity probes				Laser techniques	
	Intensity-modulated fibre optics	Electromagnetic	Eddy current	Capacitive	Holography	LDV
Frequency range	DC–70 kHz	DC–2 kHz	DC–100s of kHz	DC–50 kHz	DC–10 MHz	<1 Hz–1 MHz
Working or stand-off range	0.05–1.3 mm	0.05–2 mm	0.5–60 mm	10^{-7}–1000 mm	–	10 mm–18 m
Resolution (at optimum stand-off)	2.5×10^{-4} mm	0.1 mm/s	10^{-5} mm	10^{-4} mm	1 mm	10^{-4} mm
Non-linearity	±1%	±5%	±0.05%	±1%	–	±1%
Minimum target footprint	0.15 mm	2 mm	2 mm	1 mm	N/A	As small as 0.05 mm
Affected by environment	Dust, shock and vibration	Stray magnetic fields	Stray magnetic fields	Grease or dirt, humidity	Dust, shock, and vibration	Dust, shock, and vibration

There are also many transducers for contact measurement which can be used for general machinery diagnostics and non-critical applications. Table 3.3 gives a comparison of the performance characteristics of contact type of pickups.

3.9   Introduction to Vibration Excitation Techniques

DID YOU KNOW that there is a large variety of excitation techniques which can be used for obtaining structural response in vibration/modal testing? The choice of excitation technique depends on the size and fixity condition of the structure, the excitation signal to be imparted, the required frequency range, the sensing mechanism, and the data analysis procedure. Some procedures require specific excitation characteristics and the induced vibration level may dictate the use of specific transducers. Whether we are concerned with printed circuit boards or suspension bridges, high-speed printer mechanisms or satellite launchers, good dynamic response is important for sustained and satisfactory operation. Vibration exciters are used for studying the effects of vibration on a structure or for evaluating the physical properties of materials or structures. They may be used in any of the following phases: product development, simulation, production, dynamic testing, or life evaluation.

3.9.1   Choice of Excitation Technique

Basically excitation techniques can be divided into two categories, contacting (intrusive) and non-contacting (non-intrusive). The former type involves connection of the exciter to the test structure throughout the test. The range of exciters available includes shakers of hydraulic and electrodynamic type and unbalanced mass exciters. There are also mechanical shakers of the displacement type, as well as inertial exciters and electromagnetic shakers, especially for rotating structures. While the electrodynamic or electrohydraulic shakers are usually used to directly transmit force to the structures through pushrods or stingers, bulk-fixing of items to these shakers is also possible. In this case, the test item totally sits on the shaker table and this type of testing is more applicable to fatigue life and resonance behaviour studies of the system and/subsystems fixed to the base of the exciter. Fully active actuators, such as electromagnetic shakers, piezoceramics and films, and magnetostrictive and electro-

Table 3.3 Summary of contact transducers

Characteristics	LVDT	Electrodynamic pickup	Inductive accelerometer	Piezoelectric accelerometer	MEMS accelerometer	Servo accelerometer
Frequency range	DC–200 Hz	5 Hz–10 kHz	DC–5 kHz	5 Hz–50 kHz	DC–1000 Hz	DC–2 kHz
Working range	±50 mm	±50 mm/s	±1000 m/s^2	±5000 g	±100 g	±500 g
Resolution	10^{-3} mm	10^{-2} mm/s	10^{-3} g	10^{-3} g	10^{-2} g	10^{-3} g
Non-linearity	±0.05%	±1%	±0.05%	±1%	±0.5%	±0.02%
Minimum target footprint	0.2 mm	2 mm	5 mm	2 mm	2 mm	1 cm
Affected by environment	Stray magnetic fields	Stray magnetic fields	Stray magnetic fields	Temperature	–	Dust

hydraulic devices, can also be used to excite a structure. The measurement of forces transmitted to the structures by these shakers also requires careful consideration.

In the non-contacting type of excitation technique, the exciter is not in contact with the test structure during the test or is in contact for a very short duration only, as in the case of an impact hammer. Most of the non-contact structural excitation methods currently used are typically extensions of traditional modal excitation procedures such as operational, step relaxation, hammer impact, and acoustic methods; they do not possess the ability to apply a force input with a controlled time history to a point on the structure, which is a drawback.

The choice of excitation can make the difference between a good measurement and a bad one. Selection of excitation should be approached from both the type of excitation techniques available and best suited for the application (for example, shaker excitation) and the type of function desired (for example, sinusoidal or random noise) because they are interrelated. The excitation technique is the physical mechanism used to provide the force input to the structure and the excitation function is the mathematical signal used for the input.

3.10   Vibration Exciters of the Contact Type

INTERESTING   FACTS: The history of vibration shakers is relatively new. In the late 1920s, German engineers developed vibration machines to analyse the elastic properties of structures. These machines were mounted on a structure and used a series of rotating weights driven by a motor to produce vibration. By 1935 reaction type of exciters were regularly used for prototype testing of civil engineering structures. During World War II, there was a lot of demand for moulded rubber vibration mounts and vibration-isolating engine mounts for aircraft engines. This necessitated the need to create, measure, and control vibration so that the vibration created by an engine could be simulated and the isolation properties of its mounts could be tested. Electrodynamic shakers for generating controllable forces were built around this time and testing on a shaker became the preferred method to qualify vibration environments.

3.10.1   Electrodynamic Vibration Shaker

This is the most common type of exciter and is extensively used for the study of structural dynamics, which is essential for understanding and evaluating the performance
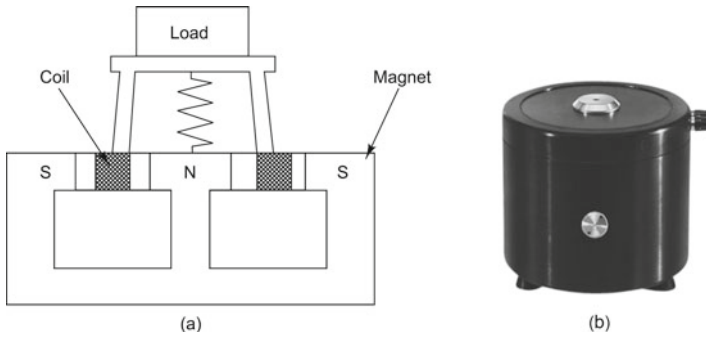
Fig. 3.27 Electrodynamic shaker: **a** schematic drawing, **b** photograph (Courtesy of Brüel & Kjær, Denmark)

of any engineering product. The electrodynamic exciter is essentially the electrodynamic transducer used in the reverse direction. The popularity of these shakers in vibration testing is due to the fact that they offer a wide range of forces—from less than 5 to 20,000 N and a broad range of frequencies—from near 0 to over 10 kHz. These exciters may be given sinusoidal, random, or transient signals for excitation and are invariably driven by an audio frequency power amplifier.

In principle, the electrodynamic exciter is similar to a common loudspeaker, where the motion is produced by a current passing through a coil in a magnetic field. But, in general, shaker coils use heavier conductors than speakers and are more robust to allow flow of heavier currents. Hence, shaker coil resistance is generally lower than that of common speakers. Just as in the case of an electrodynamic pickup, the springs are initially compressed, forcing the vibrating object to exert additional force to lower the pin into free motion in the magnetic path. Figure 3.27a shows the principle of operation of the shaker and Fig. 3.27b the photograph of a typical exciter. At the heart of the shaker is the coil of wire, suspended in a radial magnetic field. When a current is passed through this coil, a longitudinal force which resists the motion of the conductor is produced. This is proportional to the magnetic flux passing through the coil, to the current flowing through it and to the number of coil turns within the flux field. The force provided by the shaker is given by the equation

$$F = Bi_c l \tag{3.34}$$

where F is the force produced in N, B is the magnetic flux density in Wb/m^2, i_c is the current through the coil in A, and l is the length of conductor cutting the flux in m.

This force is transmitted through the spring to a table structure to which the test specimen may be fixed. The radial magnetic field is obtained by building a magnetic circuit of permeable iron or steel around an axially polarized cylindrical magnet. An inner pole piece transmits flux from one end of the magnet, say the north face to the outer south face, resulting in a radial flux field through the gap between these ferrous

parts. In order to ensure that the coil moves axially, but is restrained from all other motions, it must be accurately centred in the annular gap between the inner and outer poles. The coil is wound around the outer diameter of a stiff thin-walled tube called the coil form.

The larger electrodynamic shakers, though very similar in operation, use DC-excited electromagnets. These coils may require cooling jackets or blowers, especially those in exciters with high force ratings. Care should be taken to ensure that the shaker (i) is used within the specified maximum table acceleration velocity and stroke limits, (ii) delivers no more than the rated maximum force, (iii) consumes root mean square (RMS) current less than the continuous duty rating, and (iv) avoids eccentricity of loading beyond the rated moment restraint of the machine. Generally, the larger the shaker, the greater is the force which is available for exciting the structure, but this is at the expense of a reduced upper frequency limit. The acceleration level which can be obtained depends on the maximum current and the load. At low frequencies, however, this acceleration level is low. Resonances in the moving element decide the upper frequency limit.

The shaker is usually attached to the structure using a stinger (long, stiff, and slender rod) so that the shaker will impart a force to the structure only along the axis of the stinger, which is the axis of force measurement. A load cell is generally connected between the structure and the stinger to measure the excitation force.

Advantages

 (i) These exciters are relatively inexpensive, easy to control, simple to interface, and quite linear in their behaviour (if used within their specified force and motion limits).
(ii) The frequency and amplitude of excitation can be controlled independent of each other using a function generator, allowing a lot of operational flexibility and this is found to be useful, especially around resonances.

Disadvantages

 (i) The performance of the exciter is affected by stray extraneous magnetic fields.
(ii) It has a low specific force, i.e., output force-to-exciter weight ratio.

Typical Specifications

Maximum dynamic force: 1700 N.
Maximum acceleration (peak): 140 g.
Maximum velocity (peak): 120 cm/s.
Maximum displacement (peak to peak): 18 mm.
Mass of moving element: 1 kg.
Suspension stiffness: 300 N/cm.
Frequency range of operation: 5–5000 Hz.
Input current: 0.1 A/N.

Applications This shaker is very popular for modal analysis (finding natural frequencies, mode shapes, and damping ratios), product evaluation, stress screening, squeak and rattle testing, testing of aircraft, automobiles, etc.

3.10.1.1   Indirect Measurement of Excitation Force

It is often required to measure the force imparted by the exciter for computation of the frequency response function of a vibrating structure. The force may be computed from the expression given in Eq. (3.34) where it is assumed that the magnetic flux density B, the length of conductor in the flux field l, and the number of coil turns N in the field are constant throughout the full range of travel of the exciter. Thus, instead of using a force transducer to measure the force input into a structure, the force can be calculated from the measured current fed to the coil. As mentioned earlier, a flexible push rod or stinger links the exciter to the structure directly. Though by definition, the force generated in the coil is directly proportional to the coil current and the two are in phase, it may be inappropriate to deduce the force by measuring the current passing through the shaker coil, since this does not measure the force applied to the structure itself, but to the assembly of the structure and the shaker's drive. The moving components of the exciter and the stinger have a dynamic modifying effect on the force. It may appear that the difference between the force applied to the structure and that generated within the shaker is small, but it should be remembered that in reality, especially around resonant frequencies, the applied excitation force becomes very small, even without alteration of the settings on the power amplifier or signal generator, nevertheless producing a large response and making the frequency response function vulnerable to noise or distortion in this frequency range.

The true force applied to the structure is actually the difference between the force generated in the shaker and the inertia force required to move the stinger and the shaker table, and is, in fact, much smaller than either. Suppose that the mass of the moving part of the shaker and its connection to the structure is m_s, the force generated in the shaker is f_s and the force actually applied to the structure, i.e., the force to be measured is f_p, then we have the relation $f_s - f_p = m\ddot{x}$ where $\ddot{x}$ is the acceleration of the structure. With lightweight structures the errors can be significant, and corrections in the direction in which the force is being sensed and in perpendicular directions will be required. It is clear that there is a need for direct measurement of the force applied to a structure in order to obtain an accurate and reliable indication of the force and theorem the frequency response function properties.

3.10.1.2   Direct Approach to Force Measurement

In this method, a load cell/force transducer is attached between the structure and the stinger to measure the excitation force. Any force transducer is sensitive to (i) bending moments and transverse movements which result in translational motion of the structure in any direction other than along the driven axis, (ii) rotational motion of the test structure at the location of the force transducer, and (iii) rotational or translational motion of the exciter on its stand or suspension system. The function of the stinger is to minimize these effects. A conventional test setup using a force transducer can be seen in Fig. 3.28a. The force transducer is attached with its 'base' towards the test structure and the 'top' is connected to the shaker via the stinger.
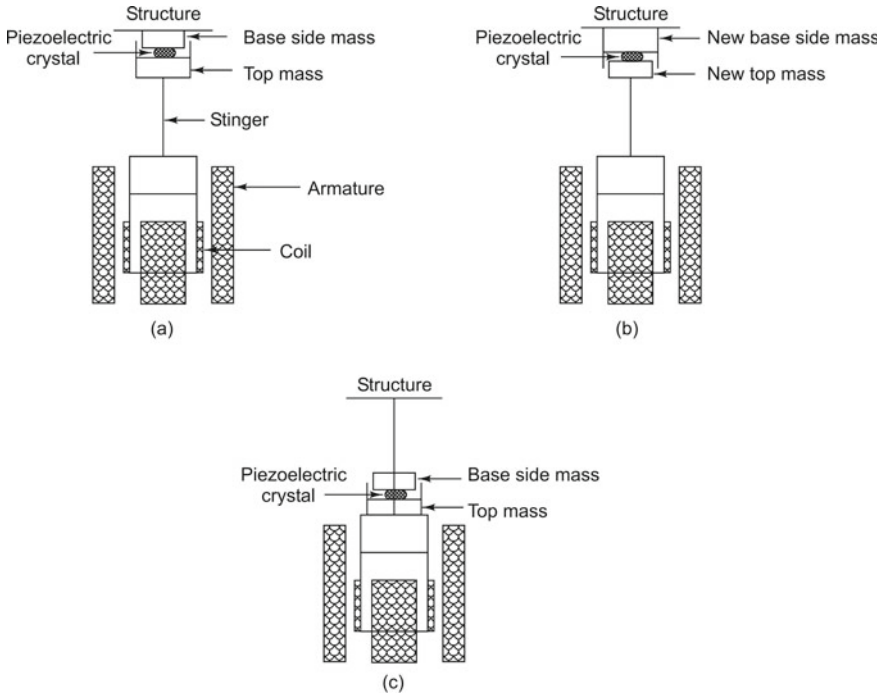
Fig. 3.28 Force transducer mounting: **a** conventional mounting, **b** wrong mounting, **c** alternate correct mounting arrangement

The 'live-side' (base-side) mass of the force transducer is kept small in an effort to minimize the modification to the structure. However, what is actually minimized is the mass modification to the structure only in the direction in which force is being sensed. The full mass of the force transducer modifies the structure in directions perpendicular to the direction of force measurement. Thus, there is a large difference between the apparent mass of the force transducer, as seen by the structure, in the direction in which force is being sensed and perpendicular to it. For a typical commercial force transducer, the 'base-side' mass is only of the order of say, 3 gm, and this is the mass seen by the structure in the direction of measurement. The total transducer mass may, however, be 20 gm, say, and this is the mass seen by the test structure in directions perpendicular to the one in which the force is being sensed.

   If the force transducer were to be mounted on the structure the 'wrong' way up (Fig. 3.28b), then there would be less difference in the apparent loading on the structure in the measurement direction and those directions perpendicular to it. However, the 'live-side' mass as seen by the structure in the sensing direction would now be 17 gm (for the force transducer mentioned earlier), causing mass loading. The mass seen by the structure in directions perpendicular to this would remain unchanged at 20 gm, i.e., the total mass of the force transducer.

A third way of fixing the force transducer is to mount it directly on the shaker platform, with the 'base' arranged towards the structure and the 'top' towards the shaker as shown in Fig. 3.28c, with the stinger connecting the force transducer to the structure. The difference between the mass added in the sensing and perpendicular directions is thus reduced. Only the small 'base-side' mass of the transducer plus the stinger modifies the structure in the perpendicular directions. Quite often, since the stinger mass is smaller than the total mass of the force transducer, the overall modifying effect will be less than for the conventional arrangement. Besides, there will be a much smaller difference in the effect of modification between the measurement direction and perpendicular directions. An advantage of this arrangement is that the force transducer cable (which can be relatively stiff and massive with indeterminate effect) may also have less mass loading on a lightweight structure.

Only the translational effects of the force transducer have been considered so far. It is to be noted that the force transducer will modify the structure in all six degrees-of-freedom (DOF) and the effects of the rotational degrees-of-freedom should be considered for accurate measurements. In the case of multipoint excitation with relative phases between the different inputs, the phase shift between the drive signal and the force must also be taken into account. This can be done using the transfer function between each drive signal and its associated force. The simpler but more laborious approach is to control the phase iteratively.

3.10.1.3 Slip Tables

For very large and heavy test objects, electrodynamic generators can no longer be used. Slip table systems help simulate practical conditions and enable structures under test to be excited in their working position. A linear guide system ensures high stiffness of the slip plate and minimizes cross vibration if the test setup is not symmetrical. Systems with horizontal slip table and vertically guided load bearing platform (Fig. 3.29) make it possible to test extremely large and heavy loads in x-, y-, and z-directions.

Fig. 3.29 Photograph of slip table (Courtesy of Saraswati Dynamics Private Ltd., India)

Guided oil film slip tables typically consist of a precision ground and lapped natural granite block with a magnesium plate that slides on an oil film to provide a highly damped horizontal slip surface and simultaneously eliminates overturning moments. The slip table generally comes with a precision granite surface, internal oil reservoir, filter, regulator, a magnesium slip plate, and a magnesium drive bar. Depending upon the application, the thickness of the slip plate typically varies from 40 to 75 mm. Low-pressure bearings designed for testing of heavy loads and voluminous objects help restrain pitch and roll moments. An oil film is automatically created through a closed-loop system, when the oil pump coupled with the motor is started and oil supply is provided for table breakaway friction. The oil supply returns to the reservoir through a flexible pipe under gravity and is adjustable for various table loads. The oil used in the table is generally a standard hydraulic fluid. The important factor for operation of this table is viscosity of the oil. When operating at higher temperatures, it may be necessary to use a larger supply or oil of higher viscosity, or both.

Commercially available models come in two versions: (i) stand-alone models designed for retrofitting existing vibration generators along with the slip table frame on a base plate and (ii) slip table integrated in a steel frame, together with the vibration generator, allowing quick alignment and precise coupling of the generator to the sliding table. These are available with standard working areas of a maximum of 4 m^2.

Driver bars provide the link between the shaker and slip plate and are generally made of magnesium. Their design is such as to enable perfect force transmission while minimizing the moving mass. Pneumatic isolation elements are also available and these make it possible to install slip tables without expensive foundations. Due to the low natural frequency of these isolators (3–5 Hz), a wide test spectrum may be applied. Hydrostatically guided slip tables are also available; these incorporate hydrostatic slide bearings, making it possible to apply the high yaw, roll and pitch moments as they appear with heavy test items or very large loads which may have a high centre of gravity above the slip plate. These tables use high-pressure bearings with a separate hydraulic supply unit.

Typical Specifications of Electrodynamic Shaker with Slip Table

Frequency range: 5–2500 Hz.
Maximum dynamic force: 80 kN (peak).
Shock load (peak): 260 kN.
Maximum acceleration: 100 g peak (for continuous operation).
Maximum velocity: 200 cm/s peak (for continuous operation).
Maximum displacement: 40 mm peak–peak (for continuous operation).

Table for Vertical Excitation

Diameter: 430 mm.
Mass of moving armature: 59 kg.
Mass of mounting plate: 8 kg.
Suspension stiffness: 600 N/cm.

Table for Horizontal Excitation (Slide Table)

Size: 900×900 mm.
Mass of moving elements: 170 kg.
Maximum payload: 7000 kg.

3.10.1.4 Vibration/Climatic Test Systems

Today industrial equipment are expected to work in all kinds of environment covering a wide range of temperature, vibration, and high humidity. Combined vibration/climatic test systems can simulate the effect of ambient conditions on components. This helps to identify weak points at an early stage, preventing expensive downtime. There are complete solutions available with software for controlling vibration testing with required test profiles in climatic chambers. Such multi-test systems are very useful for quality assurance, research, and development. Figure 3.30 shows such a climatic chamber.

Specifications

Test space: $140 \times 95 \times 115$ cm^3.
Maximum table size: 130×90 cm^2.
Temperature range: -70 to $+175\,°$C.
Temperature change rate: $5\,°$C/min.
Temperature control: $\pm 1\,°$C.
Humidity range: 20–95% R.H.
Controllable programmer: controls for temperature, humidity, etc.

Fig. 3.30 Climatic chamber
(Courtesy of Saraswati
Dynamics Private Ltd.,
India)

3.10.2 Modal Thrusters

From the discussion on electrodynamic exciters and force measurement in Sects. 3.10.1.1 and 3.10.1.2, it is evident that the measurement of input force in a modal test is very important for successful experimental modal analysis. Modal thrusters/exciters are specifically designed with the idea of obtaining consistent frequency response function (FRF) measurements. They are also designed for long-time operation and to ensure the highest possible quality and accuracy with minimum setup time.

A modal exciter is essentially an electrodynamic generator which is supported in a rugged revolving frame permitting excitation in the vertical or horizontal direction (Fig. 3.31).

There are many test structures which require an accurately controlled static preload in order to take up the slack in bearings, gears, or joints and this can easily be set in a modal exciter. These exciters are available in the market with forces from 100 N up to 15 kN. The modal generator is precisely coupled to the test structure using a stinger which faithfully transmits axial forces, and also helps isolate and decouple the force transducer/impedance head from side loads and bending moments. Modal exciters sometimes have built-in electrical stinger pre-tensioning systems, enabling exciter positioning and orientation with complete flexibility. The axial stiffness can also be adjusted electronically in many systems. The exciters come with compatible power amplifiers which can be run in current or voltage mode. Vibration transfer to the floor can be reduced by a swivel frame which allows many coupling options. The frame also has vibration isolators as a standard feature so that no extra foundation is needed.

Present state-of-the-art stinger designs are all based on the tension wire concept. A tension wire test setup utilizing a thin metal (piano) wire with zero compression and negligible bending stiffness is arranged in such a way that there is a constant static tension in the wire, upon which the dynamic oscillatory force is superimposed. In practice, the tension wire technology is implemented by using a 'through-hole'

Fig. 3.31 Modal exciter
(Courtesy of Brüel & Kjær,
Denmark)

design of the modal exciter and exciter fixture, such that the exciter can be 'slid' along the wire. A complete range of dedicated accessories including force transducers, impedance heads, exciter stands for easy horizontal positioning, as well as a family of stingers is generally provided with these exciters. The generator is cooled by a fan/blower, with cooling air entering through a filter assembly. The blower can also be installed outdoors.

Modal exciters are characterized by

- High excitation force per unit cost of exciter.
- High force-to-weight ratio due to unique rare earth magnet technology.
- High cross-axial stiffness.
- Large displacements of the order of 20–100 mm (peak–peak) allowing good low-frequency performance.
- Low mass and high rigidity of armature (moving element).
- Low weight of exciter allowing ease of positioning and orientation relative to test object.
- Minimum preparation for attachment and pre-tensioning of stinger wire.
- Wide frequency range.
- Special 'through-hole' magnesium armature for tension wire stinger technology.

Typical Specifications

Rated force: 100 N (peak sine)/70 N (RMS random).
Frequency range: 2–5000 Hz.
Maximum travel: 25 mm.
Rated current: 5 A.
Natural frequency: 6 kHz.
Spring stiffness: 4 N/mm.

3.10.3 Mechanical Exciters of the Direct Drive Type

The direct drive mechanical exciters consist of vibrating tables which are guided to have rectilinear motion and are driven by some appropriate mechanisms. They are generally driven by DC or AC motors with variable speed drives, typically from 100 to 6000 RPM (1.67–100 Hz), so as to be able to control the frequency of excitation. The displacement type of exciter is capable of producing excitation at a predetermined frequency with a predetermined magnitude of displacement. It produces a constant displacement irrespective of the speed. The displacement type of exciter consists of a rotating eccentric disc or cam driving a positive linkage mechanism between the base of the exciter and the table. A schematic drawing of two such exciters is shown in Fig. 3.32.

The mechanism may be (i) a connecting rod, (ii) a cam and roller follower or (iii) a cam and flat face follower. The amplitude of vibration is dictated purely by the crank length in the first case and by the eccentricity setting of the cam in the second
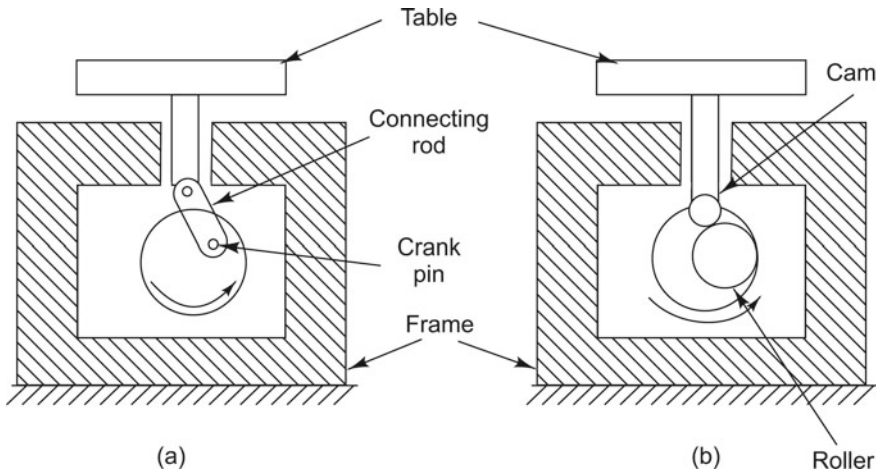
Fig. 3.32 Direct drive mechanical vibration exciter: **a** eccentric and connecting link mechanism, **b** cam and follower arrangement

case and is independent of the frequency of operation. The first mechanism will not give rise to pure simple harmonic motion, but will contain higher harmonics also. In the case of the cam mechanism, variable types of motion, dependent on the cam design employed, can be obtained. The flat face follower, on the other hand, ideally produces pure harmonic motion, but with wear over a period of time, performance may degrade. The acceleration obtained on the table is proportional to the square of the speed and hence for the same setting of the eccentric, maximum dynamic force is obtained for the highest speed of operation. With an eccentricity of around 5 mm, a maximum acceleration of 50 g is obtained at a speed of 3000 RPM.

Advantages It can produce a constant vibration displacement irrespective of frequency of operation.

Disadvantages It is generally meant for small test loads of the order of a few hundreds of newtons.

Typical Specifications

   Amplitude: ±0.1–1 mm vertical and horizontal.
   Frequency range: 0–200 Hz.
   Test mass: 8 kg.

Applications These exciters are used for evaluating the physical properties of structures, for studying the effects of vibration on components and for fatigue tests on specimens with masses of the order of 20–40 kg.

3.10.4 Reaction-Type Exciters

Reaction types of exciters produce excitation with a predetermined magnitude of dynamic force at any chosen frequency. A reaction exciter develops its excitation force through an inertial loading which is caused by the acceleration of a reaction mass. In operation, the exciter's base is attached to the structure under test. A double mass reaction exciter consists of two rotating unbalanced masses of equal eccentricity rotating in opposite directions and phased such that the unbalanced forces add up to a sinusoidal force acting in a plane at right angles to the two axes of rotation. The total force produced by the exciter is

$$F = 2me\omega^2 \sin \omega t \tag{3.35}$$

where m is the mass of one rotor, e is the eccentricity, and ω is the angular speed of the rotor in rad/s.

It is seen that the reaction type of exciter produces an exciting force proportional to the square of the speed for a given setting. Figure 3.33a shows schematically the principle of operation of this exciter. For the arrangement shown in this figure, vibratory forces cancel out in the horizontal direction. Vibration in the vertical direction is facilitated by fixing the exciter base 'a' and the test structure on a test platform resting on vertical springs as shown in the figure. By mounting the exciter on face 'b', excitation in the horizontal direction can be obtained. The photograph of a heavy duty reaction exciter is shown in Fig. 3.33b. Figure 3.33c, d indicates the procedure for changing the eccentricity in steps using the principle of double eccentric. The eccentric mass consists of an outer hollow shaft 2 and an inner solid shaft 1. When the unit is stationary, eccentric mass 2 can be rotated relative to eccentric mass 1 and fixed, as shown in the figure to get different settings of eccentricity.

Advantages

 (i) This exciter has a very high force-to-exciter weight ratio and is capable of generating large amounts of forces at higher frequencies.
(ii) The magnitude of excitation force can be calculated fairly accurately knowing m and e and no further measurement of force may be required.

Disadvantages

  (i) This exciter produces very small forces at low frequencies because of speed-squared dependence.
 (ii) There is relatively little flexibility or control over the generated forces as compared to electrodynamic, electrohydraulic, or other exciters since the magnitude of force is dictated by eccentricity e, which can be changed only in discrete steps and not continuously.
(iii) In practice, it is not possible to change the magnitude of force at a particular speed while the exciter is running, since this involves changing e and this can be done only when the exciter is stationary.
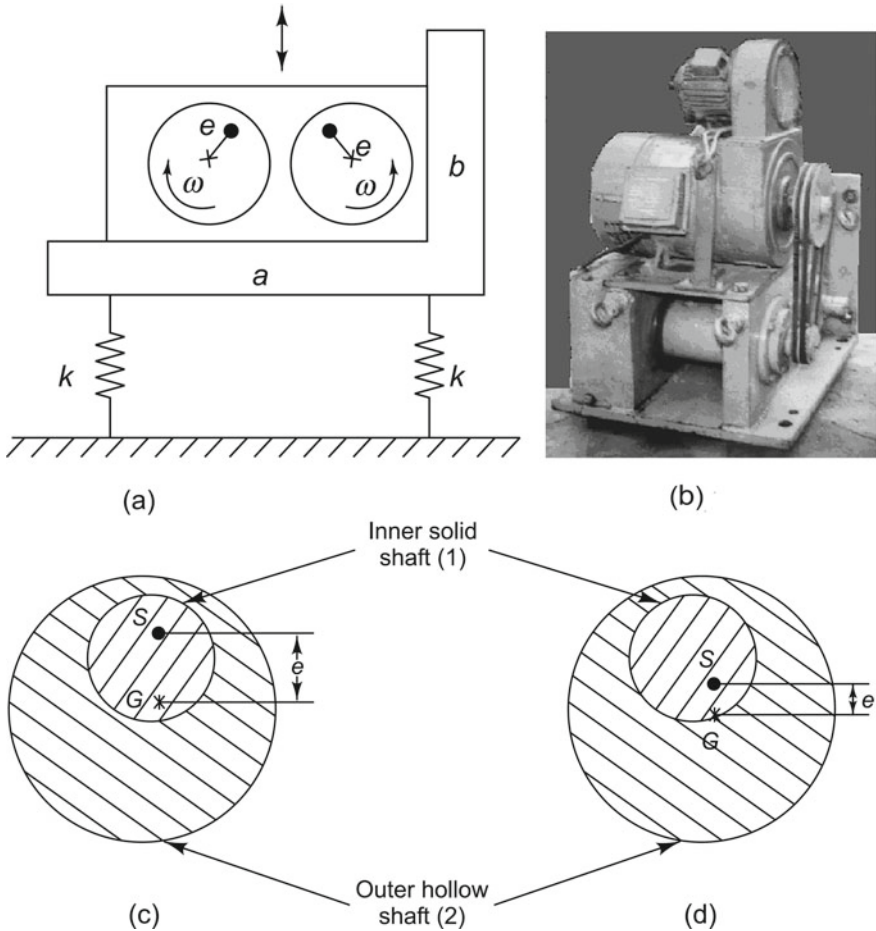
Fig. 3.33 Double mass mechanical exciter: **a** schematic of exciter, **b** photograph (Courtesy of IIT Madras), **c** maximum eccentricity setting, **d** minimum eccentricity setting. S—centre of rotation, G—centre of mass, e—eccentricity

(iv) Since the exciter is driven by a motor, its upper frequency limit of operation is typically restricted to around 100 Hz.

(v) This exciter can be used only as a reaction exciter.

Applications

(i) These exciters can be used to determine the natural frequencies of heavy structures like foundations, bridges, masts, towers, etc., which require large forces and the shaker itself offers negligible weight compared to the test item.

(ii) It can also be used to perform simulated seismic tests on specimens which are required to undergo seismic qualification tests like bus ducts, transformers, pan-

els, and other parts of power plants which are to be erected in earthquake prone zones.

Typical Specifications

Speed range: 100–4200 rpm.
Drive: 2.5 kW DC motor with speed control unit.
Exciting force: 30,000 N at 3000 rpm.

3.10.5   Inertial Systems

An interesting development in exciters is that of the smaller inertial shakers (reaction mass shakers) used extensively in modal analysis. Figure 3.34 shows the photograph of an inertial exciter.

Inertial exciters can efficiently apply dynamic forces to large structures and have applications in common production, aerospace and civil engineering, and shipbuilding. The inertial mass modal exciter allows the testing of structures where backup fixturing is not available or possible. The inertial forces are generated by accelerating an inertial mass with a closed-loop hydraulic excitation system. In some systems, excitation is caused by permanent magnets, and a special spring system provides optimal guidance so that the full body mass of the exciter can impact the structure. The controller helps to maintain a constant force over the required frequency range by continuously altering the drive signal to maintain the required dynamic force level. Inertial systems are available in the force range from 100 to 5000 N. These systems can be bolted directly to the structure and aligned at any angle within 360° on the test structure. The generators have an excellent lateral and axial stiffness. They are fan-cooled, with cooling air entering through a filter assembly. They often come with load cells and LVDTs for force and displacement measurements.

Fig. 3.34   Inertial exciter (Courtesy of https://www.xcitesystems.com)
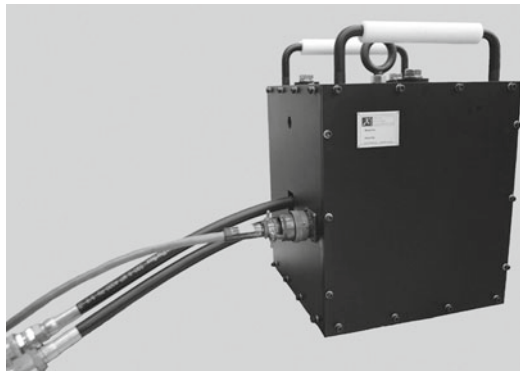
Advantages The ease with which this exciter may be positioned and oriented on the test structure makes it highly suitable for field applications.

Typical Specifications

Dynamic force: 4450 N.
Stroke: 25 mm.
Moving mass: 25 kg.
Total mass: 61 kg.
Frequency range: 30–300 Hz.

Applications The inertial shaker is ideal for extracting the response of a ship hull or car body which requires fixing the exciter in an arbitrary fashion; this type of excitation requires the least test preparation. This exciter is useful in testing electrical transmission towers, turbine rotors, stators and bearing housings, generator armatures, stators and windings, diesel engines, and motor–generator sets. The ease with which these systems can be fixed allows for testing ship bulkheads, propeller shafts, and propellers, as well as structure-borne noise isolation systems in submarines.

3.10.6 *Electrohydraulic/Hydraulic Exciters*

Electrohydraulic exciters are highly suitable for applications requiring large amounts of forces, typically of the order of tonnes. These exciters can provide relatively long strokes, allowing excitation of structures at large amplitudes, a feature not available in electrodynamic exciters of comparable sizes. The power amplification required for the generation of large forces is achieved through the use of hydraulics. These exciters are, in general, more complex and expensive than electrodynamic and mechanical exciters. One main advantage of these exciters is their ability to apply simultaneously a static load with a superimposed dynamic load. This is useful while testing structures with a major static load, which may change the dynamic properties, or even geometry of the structure. They exhibit high force capability, high force per unit weight and volume, good mechanical stiffness, and high dynamic response. There are also commercially available units which produce triaxial forces. These exciters are suitable for low-frequency operation; at higher frequencies, the stroke gets reduced considerably. The frequency range of these exciters is typically 0–100 Hz, the exact frequency range depending on a number of factors such as pump and flow rate capacity. The upper usable frequencies of the electrohydraulic exciters available today are considerably higher than what they were earlier; very few operate above 1 kHz though.

The hydraulic vibration machine, the schematic of which is shown in Fig. 3.35a, transforms power in the form of high-pressure fluid flow to the reciprocating motion of the table attached to the stem of an actuator. The major components of this exciter are the system controller, the servo valve, the hydraulic pump (power pack), a four-way control valve, the hydraulic actuator, and the test table. The power pack typi-
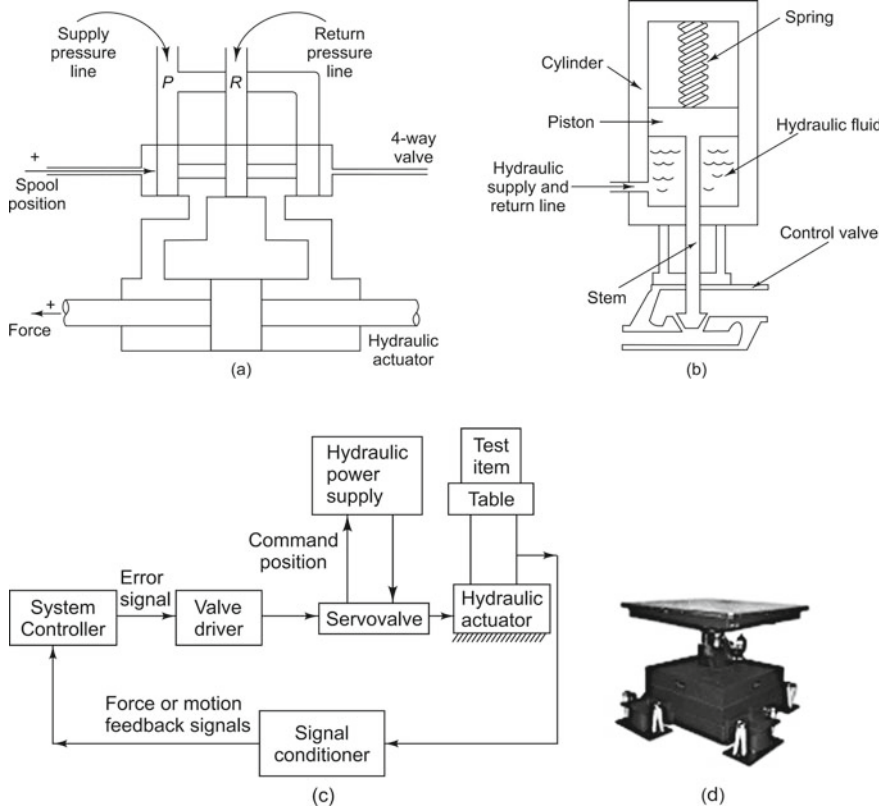
Fig. 3.35   Hydraulic shaker: **a** schematic, **b** piston-type hydraulic actuator, **c** closed-loop hydraulic vibration exciter, **d** photograph (Courtesy of https://www.mbdynamics.com)

cally consists of a high-performance hydraulic pump with appropriate pressure relief valves, oil coolers, etc. and produces the required high pressure of the oil. The four-way control valve regulates the flow of oil to either side of the working actuator and is in many systems actuated by a double eccentric and connecting rod mechanism. The double eccentric is driven by a variable speed motor.

In the figure, the four-way valve is shown in its mid-position blocking both high-pressure line P and return low-pressure line R. The position of the servo valve controls the oil flow from the pump's high-pressure line P to the actuator piston and from the piston low-pressure side to the pump's reservoir line R. If the servo valve position causes the spool to be displaced to the right, high-pressure oil enters the right side of the actuator piston and the low-pressure oil is drained to the reservoir on the left-hand side of the piston. This causes the piston to move to the left so as to develop a force on the test table. As there is no feedback arrangement in this setup, the motion of the valve spool should be symmetric with reference to the mid-position closing the control ports. Hence, a reciprocating motion which can be approximated to simple

harmonic motion is provided by the double eccentric and connecting rod mechanism, by making the connecting rod sufficiently longer than the crank.

In some simpler designs of the piston-type hydraulic actuator shown in Fig. 3.35b, one side of the piston contains a spring. The vibration of the test table is produced by allowing alternating flows of the high-pressure hydraulic fluid by means of the control valve into the actuator cylinder. The actuator has a test table attached to its piston rod. Generally, there is provision for clamping the cylinder so that the table vibrates either in a horizontal or vertical plane. The stem transmits the motion of the piston to a valve. By regulating the amount of oil supplied to or drained from the actuator cylinder, the valve can be positioned between fully open and fully closed conditions.

Figure 3.35c shows a block diagram of a typical closed-loop hydraulic vibration exciter. The test item is fixed on the table driven by the hydraulic actuator. The valve driver which is an electrohydraulic device is driven by the error signal between the required output motion and the input force and/or motion of the hydraulic actuator, which is used as a feedback signal. This error signal actuates the larger servo valve, which in turn controls the amount of oil that is forced into the hydraulic actuator to obtain the desired motion. Figure 3.35d shows a photograph of this type of exciter.

The flow rate performance of such a system is frequency dependent, the high-frequency behaviour being essentially controlled by the servovalve, hydraulic actuators, and test system dynamics. Important specifications for electrohydraulic and hydraulic valve actuators are actuation time and hydraulic fluid supply pressure range. Devices that have linear motion valves vary in terms of valve stem stroke length and actuator force. For both electrohydraulic and hydraulic valve actuators, acting type should also be specified. In single-acting devices, fluid pressure actuates the valve in one direction, while a compressed spring actuates the valve in the other. In double-acting devices, fluid pressure actuates the valve in both directions. Most of these exciters come with the capability of sine, random, sine on random, random on random, or resonant search, and dwell testing features.

Advantages

  (i) They provide relatively long strokes, allowing excitation at large amplitudes.
 (ii) They can apply simultaneously a static load with a superimposed dynamic load.
(iii) They have high force capability, high force per unit weight and volume, good mechanical stiffness, and high dynamic response.
(iv) Triaxial forces can be produced by these exciters.
 (v) They are highly suitable for low-frequency operation.

Disadvantages These exciters are, in general, more complex and expensive than electrodynamic and mechanical exciters.

Typical Specifications

Maximum stroke (peak–peak): 50 mm.
Maximum dynamic force (peak): 15 kN.
Frequency range: 0–800 Hz.

Oil-hydraulic pump system:
Maximum flow capacity: 180 L/min.
Maximum pressure: 25 MPa (250 bar).

Applications These exciters were originally developed for materials testing applications. Present-day applications, however, include fatigue testing, fluid–structure interaction studies, and earthquake simulation studies. They find wide use in the aeronautical and automobile industry, where they are placed under the wheels of a vehicle to simulate the condition of a road or runway to test the suspension system of automobiles or aircrafts or to conduct seat vibration tests on a human body.

3.10.7 Lead Zirconate Titanate Actuators

Devices such as electromagnetic shakers, piezoceramics and films, and magnetostrictive and electrohydraulic devices work as fully active actuators supplying mechanical power to a system. They generate secondary response in a linear mechanical system, which interferes with the original response of the system caused by the primary disturbance and reduces overall response. Semi-active actuators behave basically as passive elements, capable of storing or dissipating energy. These actuators are adaptive since their passive mechanical properties can be adjusted by applying control signals. Electrorheological fluids, magnetostrictive actuators, electrostatic devices, or shape memory alloys can be used in the fabrication of these semi-active actuators. Among these different options, piezoelectric transducers offer several distinctive advantages in active control applications because they are lightweight, can be of any shape, and can even be microsized and are easily driven by a voltage. These actuators are flat, thin piezoceramic devices bonded to the surface of a structure. When used as strain actuators, they create linear motion and extend with applied voltage. When used as bimorph actuators, they bend with applied voltage while clamped firmly at one end. They have a wide frequency response, are capable of delivering considerable amount of forces with very little input power, and are well suited for high-precision applications. They can be directly embedded or bonded to the structure.

The best known and most commonly used piezoceramic is lead zirconate titanate or PZT. Many of the newer devices do not have the problems traditionally associated with using piezoelectric materials, i.e., no wires are required to be soldered. They come in a protective skin which provides electrical insulation and protection against humidity and harsh contaminants. They also have pre-attached electrical leads, producing a highly reliable component. The new piezoelectrics are easy to integrate into volume manufactured products. Figure 3.36 shows a photograph of a PZT patch.

When piezoelectric materials are subjected to stress, a charge proportional to the applied stress is obtained and hence gives rise to an electric field. Likewise, when these materials are exposed to an electric field, the intermolecular distances change. This gives rise to a strain field that is proportional to the strength and polarity of the electric field. These two effects are represented in strain-charge tensor form as
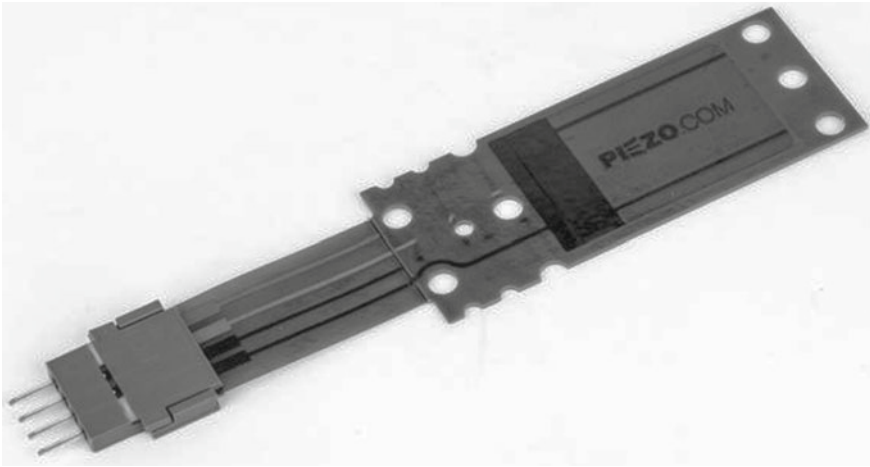
Fig. 3.36 Photograph of PZT patch (Courtesy of https://www.mide.com)

$$S = s^E \cdot T + \mathrm{d} \cdot E$$
$$D = \mathrm{d} \cdot T + \varepsilon^T \cdot E \tag{3.36}$$

where S is the strain in m/m, s is the compliance with unit 1/Pa, T is the stress in Pa, d is the piezoelectric coefficient in C/N, E is the electric field in V/m or N/C, D is the dielectric displacement in C/m^2, and ε is the permittivity in F/m or C/Vm. The superscript E on compliance matrix s^E indicates that the compliance data was measured under at least a constant and preferably a zero electric field. Likewise superscript T on the permittivity matrix ε^T implies that the permittivity data was measured under at least a constant, and preferably a zero stress field.

Advantages

 (i) They are light and extremely small and are well suited for high-precision applications.
(ii) They are easily driven by a voltage and can deliver large amounts of forces with very little power input.
(iii) They have a wide frequency response.
(iv) Besides, they can be used as both sensors and actuators.

Disadvantages

 (i) They are consumables, in the sense that once bonded to a particular location, they cannot be removed and reused.
(ii) They are not as rugged as other exciters.
(iii) The force obtained, as compared to electrodynamic or mechanical exciters, is very small.

Typical Specifications of Actuator Patches

Application type: Strain actuator.
Device size (cm): $10 \times 2.5 \times 0.08$.
Device mass (gm): 10.
Active elements: 2 stacks of 2 piezos.
Piezo wafer size (cm): $4.6 \times 2.1 \times 0.025$.
Device capacitance: (μF): 0.26.
Full-scale voltage range (V): ± 200.
Power output (W): 20.

3.10.8   Summary of Intrusive Excitation Techniques

Table 3.4 summarizes the features of various contact types of exciters in terms of force ratings, frequency ranges, applications, etc. These are indicative values only and will vary even within the same class of exciters.

3.11   Non-intrusive (Non-contact) Excitation Techniques

INTERESTING FACTS: Frequency response measurements of medium and heavyweight structures are possible through the use of either intrusive or non-intrusive (non-contact) forms of excitation and vibration response transducers, since they do not alter system dynamics significantly by causing mass loading effects for the extraction of modal parameters. However, in the case of rotating as well as lightweight delicate mechanical structures, the use of intrusive excitation techniques, as well as contact type of transducers, causes considerable mass loading so as to significantly alter the system dynamics, preventing accurate modal parameter extraction. Similarly, the structural impedance matching and mass loading effects between a modal shaker and a lightweight test structure may also alter the dynamics. This necessitates the use of non-intrusive excitation and sensing mechanisms.

The major requirements of an ideal non-intrusive structural excitation system are as follows:

- The system should apply a force of sufficient magnitude to excite all the structural modes.
- Direct force measurement should be possible to obtain calibrated input/output data, and hence FRFs and modal parameters.

Table 3.4 Summary of intrusive exciters

Type	Force rating	Frequency range	Application	Nature of force signal	Maximum acceleration (g)
Electrodynamic shaker	2 kN	DC–5 kHz	Product testing/resonance search/fatigue testing/modal testing	Sine/random/ sine sweep/shock	150
Electrodynamic modal exciter	0.25 kN	DC–2 kHz	Modal testing	Sine/random/ sweep burst	150
Mechanical direct drive exciter	0.2 kN	5–100 Hz	Fatigue testing/calibration of transducers	Sinusoidal, variable speed	70
Unbalanced Mass exciters	20 kN	1–100 Hz	Modal testing of large structures/seismic testing	Sinusoidal, variable speed	10
Electrohydraulic shakers	250 kN	0.1–100 Hz	Automobile testing, human body response, seismic simulation	Sine/random	5
Piezoactuator	10s of N	0–5 kHz	Active vibration control	Sine/random/ sine sweep	0.1

- It should be possible to position the exciter and the transducer at various locations since existing modal analysis procedures require the estimation of input/output data in the form of point FRFs and transfer FRFs (discussed in Sect. 9.4).
- It is desirable to have the ability to control the force amplitude and time waveform characteristics to make the system compatible with standard testing procedures.
- The exciter must be able to deliver excitation forces in a broad frequency range, being equally effective at very low and high frequencies so that the system is applicable to different types of structures.

3.11.1 Near Non-intrusive Excitation Using Electrodynamic Shaker

Structural excitation with a shaker system is highly dependent on the test setup and is essentially intrusive in nature as we have seen in the earlier sections. For effective shaker excitation of a lightweight structure, the shaker should be attached

to the structure in such a way that the dynamics of the structure is not altered. One way of doing this is by mounting the entire test structure directly on the shaker system. This technique is attractive since it provides a means to use the shaker and its inherent control of the excitation signal. The drawback of this approach arises from the difficulty of separating the dynamics of the exciter and test stand from that of the structure under study. The main body of the shaker must be isolated from the structure to prevent any reaction forces from being transmitted back to the structure and this is done using a stinger as has been discussed. Besides, mounting the entire test stand on a shaker and exciting it to acceptable measurement levels may be expensive/impossible for some applications. Another potential problem that arises with electrodynamic shakers is the impedance mismatch that can exist between the structure and the shaker coil. The electrical impedance of the shaker also varies with the amplitude of motion of the coil. As discussed earlier in Sect. 3.10.1.1, at resonance with a small effective mass, very little excitation force is required to produce a response. This can result in a drop in the force spectrum around resonance, causing the force measurement to be susceptible to noise. The problem can be corrected by moving to a point with a larger effective mass, by using shakers with coils of different sizes or driving the shaker with a constant current-type amplifier.

3.11.2 Operational Excitation/Ambient Testing

For very large structures, such as dams, large bridges, and buildings, ambient testing is used. Here it is important not to interfere with the functioning of the structure. For example, in the testing of bridges, interruptions of traffic flow are undesirable. Any ambient testing must be done in the presence of the inherent operating forcing functions. Also, the applied force functions are often immeasurable and of distributed nature. Since the actual force cannot be measured, only response and not FRF data can be got and the point-to-point structural dynamics are often estimated through the use of a reference and a moving transducer. This requires special procedures and prevents the use of reliable forced vibration studies with data of this type.

Operational excitation provides a viable technique for lightweight structures also. Since no external connections are required, external loading effects are eliminated. However, since operational excitation may typically be harmonic or periodic in nature, all the structural modes may not be excited and hence may not be present in the vibration database. In addition, this method is applicable so long as the test structure does not undergo any gross motion during its work cycle, necessitating delicate positioning of the non-contact transducers.

3.11.3 Step Relaxation

Dynamic testing of large structures, such as bridges and windmills, presents a variety of measurement challenges. Foremost among these challenges is a means of supplying an impact with sufficient energy to excite the test structure uniformly. Since majority of the dynamic characteristics of interest in large structures are of low frequency, a means of restricting the input spectrum of the force is an important consideration. Step relaxation/free decay testing is one of the most common forms of structural excitation in which the system is given an initial condition (displacement, velocity, etc.) and the free decay response of the system is measured. The technique entails preloading the structure at a point by a flexible tendon. When either the tensile load in the tendon or the displacement reaches a specified value, the load is suddenly released. This imparts an impulsive step function to the structure. Typically, a load cell is attached to the tendon anchor to measure the force time history. The resulting free decay can be processed using any of the parameter estimation algorithms available for processing frequency response or unit impulse response measurements. For the dynamic testing of bridges, the step relaxation technique has been used very successfully with the load provided by hanging a large mass from the structure and then suddenly releasing it. The technique is especially useful for low frequencies.

The technique is easy to use and requires minimal equipment. The difficulties in this technique are related to the tedious time-consuming setup procedures required to produce repeatable impulses and the application of the tendon fracture mechanism. Also, because the technique produces an impulsive excitation, it generates a poor signal-to-noise ratio when the time window is much larger than the actual impulsive signal, as is the case with FFT zoom analysis. Finally, a potentially time-consuming trial-and-error process may be required to obtain an appropriate input, since the technique has little direct control over the excitation levels.

3.11.4 Impact Testing Using Modal Hammers

Hammer impact testing is one of the most widely used and popular excitation techniques for modal analysis due to its ease of use and ability to adapt to difficult testing environments. The impact/modal hammer is simply a hammer with an integral force transducer built into its head. This transducer detects the magnitude of the force felt by the impactor, which is equal and opposite to that experienced by the structure. It is generally used in conjunction with an accelerometer on the component being tested to obtain FRFs by processing impact force and acceleration response data simultaneously. When the hammer tip impacts the specimen, the latter experiences a force which is that of a half sine pulse. Although the hammer contacts the test structure, the contact time is usually quite short and does not affect the structure's vibration response. The short impulsive force produces desirable broadband excitation. Since the excitation is impulsive, it suffers from the same poor signal-to-noise
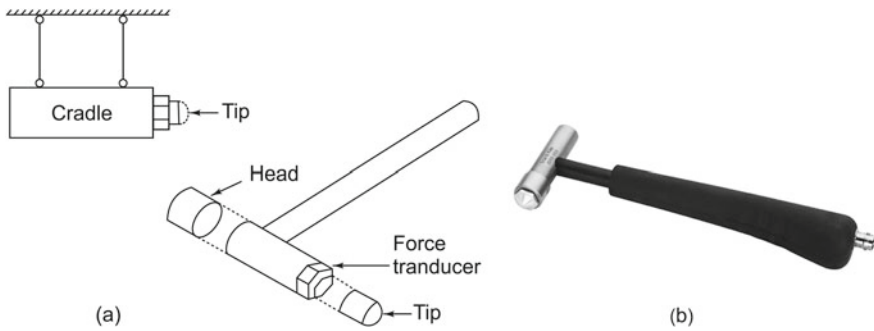
Fig. 3.37 Impact hammer: **a** different arrangements, **b** photograph (Courtesy of Brüel & Kjær, Denmark)

ratio and poor load control limitations as the step relaxation technique, making it difficult to obtain consistent results. However, its low cost and ease of operation make it an attractive alternative for some applications. This technique can be used with lightweight structures, provided the structure is sufficiently stiff to produce a useful force level. Testing on miniature structures requires a precise positioning mechanism to ensure that the impact is applied consistently at the same point. Impact hammers can be controlled either manually or remotely. That is, the device could be an actual hand-held hammer to be used by manually swinging the impactor or in a cradle arrangement to enhance the repeatability of the impulse as shown in Fig. 3.37.

Selection of the hammer involves choosing the appropriate size and mass of the hammer, which will impart the required force amplitude and spectral content for excitation of the test structure. Each hammer comes with a number of tips and an extender mass which allows further tuning by concentrating more energy at lower frequencies. The striking tips are usually of various materials and hardness (typically soft plastic, rubber, aluminium, or steel stud), allowing for different force level ranges (energies for excitation) and different impact durations (measured in milliseconds) which control the frequency range of excitation. In general, harder tips will deform less than softer tips during impact and will therefore result in shorter pulse durations. The frequency range effectively excited is controlled by the stiffness (not hardness) of the contacting surfaces and the mass of the impactor head and there is a system resonance at a frequency ω_c equal to $\sqrt{\text{contact stiffness/impactor mass}}$, above which it is difficult to deliver energy into the test structure. Typically, the stiffest tips are used to measure response at the highest frequencies. Figure 3.38 illustrates the impulse response and force spectra for different hammer heads. Table 3.5 gives typical force values and frequency ranges for the different tips.

A number of hammer kits are commercially available; they consist of matched components which are tuned for testing structures of certain size and weight combination. The manufacturer's data sheets indicate the frequency content of the force impulse that can be obtained using the various tips. Low mass hammers with hard tips provide short duration impacts, while softer tips and mass extenders are used
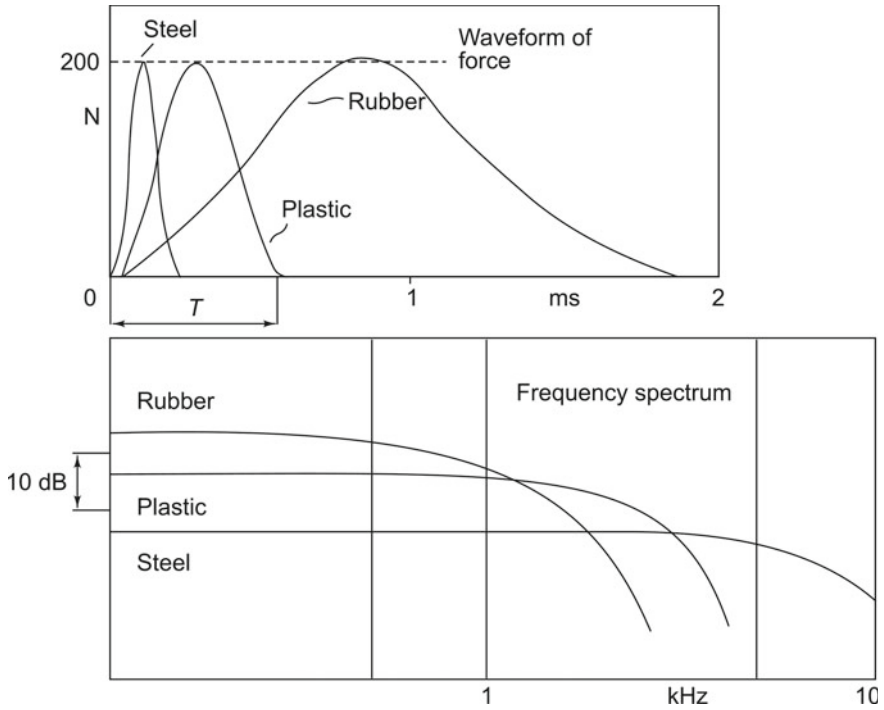
Fig. 3.38 Typical impact time history and force spectrum

Table 3.5 Force and frequency characteristics of different tips

Tip	Obtainable force (N)	Frequency range (Hz)	Duration of impulse (ms)
Rubber	500	~400	1.3–1.8
Plastic	1500	~2000	0.3–0.5
Steel	5000	~6000	0.1–0.2

to broaden the impulse. The kits also sometimes include response accelerometers, signal conditioners, cables, and all other accessories. PCB's Modally Tuned® Impact Hammer eliminates hammer resonances from the test results, reducing double impacts and resulting in more accurate test results. Such hammers have been refined, through the selection of their materials of construction, to deliver consistent, accurate results. The frequency spectrum obtained is nearly flat over a broad, yet specified frequency range and is similar to broadband random excitation. Thus a simple impact can excite all the structural resonances within the frequency range. These hammers have found use in such applications as automotive design, bridge health assessment, and aerospace vehicle development.

3.11.4.1 Punch Impactor

Remotely operated impact hammers are also available and these are more like an automated piece of machinery akin to a hand-held power tool that is operated by actuating a switch; here the operator can control the force to be applied. These electric hammers minimize errors due to overloads, mis-hits, and multiple hits, in addition to improving the consistency of impulse excitations. They also avoid difficulties of a hand-controlled hammer due to multiple impacts and poor repeatability of impact conditions, namely, the force amplitude and angle of force detection.

Structures often exhibit geometries which make the impact excitation of certain locations difficult or impossible with a standard impact hammer. The impacting force provided by the hammer is not equal to the exciting force transmitted to the structure due to the inertial mass of the punch. A device called the modal punch or instrumented punch or modal sledge hammer consisting of a force transducer attached to the end of a variable length shaft was designed to eliminate the inertial mass of the shaft by positioning the force transducer to be in contact with the test object. The main use of a punch impactor is to impact locations that cannot be accessed by a conventional impact hammer. It can also be used to precisely locate impacts to minimize the variation in the impact location and to impact in a skewed direction (non-orthogonal to the global coordinates). The punch impactor should be fitted with a hard tip. A tip of metal or hard plastic tip is recommended. The punch impactor should be calibrated in the configuration in which it is to be tested since the sensitivity is dependent on the tip and specimen combination. It can be swung manually or be used in a cradle arrangement to enhance the repeatability of the impulse.

3.11.4.2 Problems in Impact Testing

Impact testing has its own disadvantages, limitations, and potential problems like double hits, local deformation of a structure, skewed inputs, etc. It also poses problems when used on non-linear structures.

Double Hits Multiple impacts occur if the surface is not impacted sharply or if the amplitude of vibration is so large as to cause the rebounding surface to contact the hammer tip before the hammer rebounds away from the surface. This happens mainly in very lightly damped systems, in which case a light hammer should be used to reduce the possibility of multiple impacts. Multiple impacts should be avoided whenever possible, since they produce substantial measurement errors, especially in FRF measurements.

Local Deformation of Structure Impacting highly flexible areas of a structure, such as a panel, can cause large, local deformation which effectively lengthens the duration of the impact, thereby reducing the usable frequency range of the input spectrum. In such a case, it is not possible to impart energy at the higher frequencies, leading to poor estimates of the frequency response function. Besides, care must be taken not to dent the surface being impacted.

Use on Non-linear Structures An impact force is a poor input for exciting a system because it is impulsive and non-deterministic, both being undesirable for measuring the frequency response of a non-linear system when modal parameters are to be estimated. The large amplitude can overdrive the system and exaggerate its non-linear response. To reduce the effects of non-linearities which are inherent to all physical structures, it is desirable to apply a series of impacts to the test structure and take the average of the responses over a large number of records. Because of inconsistencies in the duration and force of the blows, the level of energy in impact excitation applied with a conventional hammer can be difficult to control. To ensure repeatability in averaged impact measurement, an electric hammer may be used.

Skewed Inputs and Immeasurable DOFs Obtaining impacts that are consistent in magnitude and alignment is an important concern. It is important to impact at the same point on the surface and normal to the surface, for each average at an impact location giving rise to FRF measurements showing accurate reciprocity between the input and output DOFs.

Advantages

 (i) Low cost and ease of operation make it an attractive technique for many applications.
 (ii) No special fixtures are required even in difficult testing environments.
(iii) Self-windowing potential for minimal signal processing errors makes it attractive.

Typical specifications	For medium structures	For very heavy structures
Frequency range:	8 kHz	0.5 kHz
Maximum force output:	2200 N	22,000 N
Sensitivity:	2 mV/N	0.2 mV/N
Resonant frequency:	30 kHz	3 kHz
Hammer mass:	150 gm	5 kg
Tip diameter:	0.5 cm	8 cm
Handle length:	20 cm	100 cm

3.11.5 Acoustic Excitation

Acoustic excitation of a structure is typically used for qualification tests on space vehicles and requires placing the structure in an artificially generated sound field. These vehicles and their accessories are subjected to high noise levels of the order of 140–150 dB transmitted through the surrounding air during the launch phase. This acoustic loading is particularly critical for lightweight structures with large surfaces such as solar panels. It is preferable to have these structures tested in reverberant chambers with a diffuse sound field of comparable power distribution over
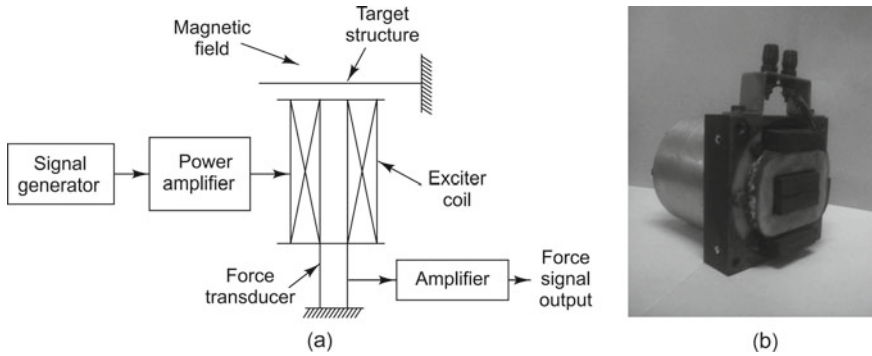
Fig. 3.39 Electromagnetic exciter: **a** schematic, **b** photograph (Courtesy of IIT Madras)

the frequency range generated under actual conditions. Acoustic excitation could be generated either by a single loudspeaker or an array of loudspeakers or by using an electropneumatic transducer (Sect. 5.3.6). The sound wave from the sound source is coupled to the object under test by means of an acoustic horn which is typically made of titanium or steel and is coupled to the object through a softer buffer material that serves both to transmit the sound and to prevent direct contact between the horn and the structure under test. Although, in some cases, the horn is directly coupled to the part with no intermediate buffer, this might produce damage to the structure at the point of contact. It is possible to input a wide variety of excitation signals (random, swept sine, pseudo-random, etc.) which can greatly enhance the quality of the acquired data. The most important feature of this technique is that it is truly a non-intrusive excitation technique. The major drawback with this technique lies in the fact that it is a distributed and not point excitation technique. This fact, combined with the problem of obtaining calibrated structural force data, makes it difficult to apply structural forced response studies to modal models developed with acoustic excitation impracticable.

3.11.6   Electromagnetic Exciter

The electromagnetic exciter is a non-contact exciter and is extremely suitable for applications, where exciter and transducer loading are particularly important. This exciter is also useful for transmitting forces to rotating objects, enabling analysis and control of vibrations of rotating machines and calculation of stiffness and damping of bearings. The exciter is also useful for quantitative evaluation of the dynamic response of impellers, circular saws, rotor discs, etc. This exciter essentially works in the reverse direction as the non-contacting electromagnetic pickup discussed in Sect. 3.4.1. Figure 3.39a shows a generalized block diagram of the test setup for electromagnetic excitation and Fig. 3.39b a shows a photograph of the exciter.

The current through the coil produces electrical energy which is stored in the electromagnetic field. This energy is transferred in a relatively complex way to the target structure. This process involves the interaction of the electrical inputs with not only the electrical parameters of the exciter coil, but also with the properties of the exciter coil, core material, and the target itself. The driving source input current produces a magnetic field in and around the exciter coil, which interacts with the ferromagnetic target structure to produce a mechanical force input for structural excitation. For non-ferromagnetic materials, it is necessary to attach a small ferromagnetic target or magnet to the target surface.

The induced force from the magnet is given by the equation

$$F = \frac{PA}{2\mu_0}[B \sin \omega t]^2 = \frac{PA}{2\mu_0}\left[B^2\left(\frac{1 - \cos 2\omega t}{2}\right)\right] = \frac{PAB^2}{4\mu_0}[1 - \cos 2\omega t] \, N$$
(3.37)

where F is the total induced force of attraction at the air gap (N), P is the number of air gaps assumed to have the same length and cross section, A is the cross-sectional area of magnetic core at air gap (m^2), B is the magnetic flux density through the core cross section (T or Wb/m^2), μ_0 is the permeability of free space (H/m), and ω is the angular frequency of the exciter driving current (rad/s).

Equation (3.37) shows that the input force from an electromagnetic exciter is a pull-release force with a DC effect, and not a typical push–pull force like that produced by a standard modal shaker. This effectively produces a preload on the target structure, which may not affect the modal parameters for a linear structure, but may cause some problems with the response transducer positioning or with the analysis of non-linear structures.

When the current in a ferromagnetic material is increased, the magnetic flux of the material increases along the idealized magnetization curve as shown in Fig. 3.40. A sinusoidal input current is distorted by the magnetization curve, producing a distorted magnetic field in the exciter, leading to significant harmonics in the force. When a pure sinusoidal excitation is required, the experimentalist has to use a lower amplitude of input current so as to operate along only a small part of the magnetization curve, thus linearizing the system by reducing the overall distortion caused by the magnetization curve. Another option is to generate the magnetization curve for the exciter system and to use this data to develop the required exciter coil input current signal to produce a pure sinusoidal magnetic field. The design of these systems is relatively straightforward. However, the key component in calibrating electromagnetic shakers is the force transducer, which directly measures the force transmitted to the target structure.

This exciter can be fed with a variety of force input time histories, in addition to the pure sinusoidal forcing function. A force transducer may be mounted between the exciter coil and the ground to measure the dynamic force on the target structure. The magnetic material properties of the core and the target material have a distorting effect on the magnetic field setup due to the current flowing in the exciter coil windings. Typically, ferrite is used as the exciter core and target material to overcome the hysteresis phenomenon, resulting in electrical energy losses within the system. This

Fig. 3.40 Idealized
magnetization curve
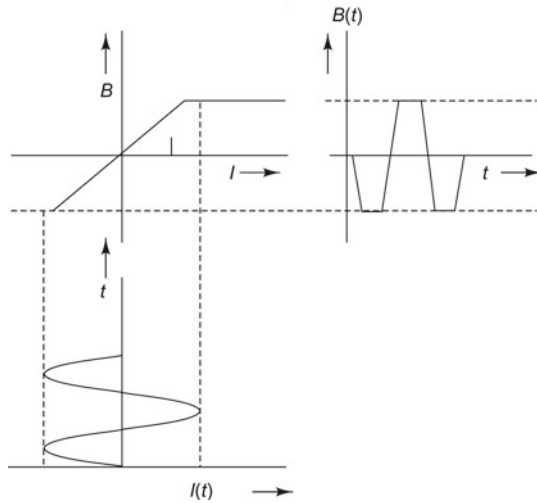
is especially important at the low flux densities required to excite small, light, and
ultra-lightweight structures.

Typical Specifications

Dynamic force: ± 300 N.
Force constant: 100 N/A.
Frequency range: 20–300 Hz.
Output impedance: 2.5 Ω with 20 mH in series.
Weight: 50 N.

Advantages This excitation technique possesses many of the desirable capabilities
required for modal testing such as non-contact excitation, easily controllable excita-
tion level, and small target area on the structure.

Disadvantages

 (i) This exciter is affected by stray magnetic fields.
(ii) Excitation current versus magnetic field is not linear, posing calibration prob-
     lems.

3.11.7 Assessment of Non-intrusive Excitation Techniques

To examine the relative merits and demerits of the various non-contact excitation
techniques discussed, five characteristics of the excitation systems such as frequency

range, control of signal or waveform type, signal level or amplitude control, application footprint or size of target area, and special test fixture requirements are assessed. All the ratings of these techniques have been made on a qualitative basis in Table 3.6.

The usable frequency range is highly dependent on the test setup. Only the acoustic and electromagnetic systems inherently possess the ability to impart an excitation over a specific frequency range independent of the experimental setup, since they are fully of the non-contact type. In the case of hammer excitation, the frequency range depends on the contact stiffness which in turn depends on the tip. As far as the ability to impart a specific excitation signal/waveform/forcing function to the test structure is concerned, the shaker and electromagnetic systems have the best ability. The acoustic approach may be limited by the ability of the structure to respond to the applied acoustic field in a specific fashion. The remaining approaches have no control over the waveform imparted to the structure. This may be a serious limitation for multiple-input multiple-output (MIMO) testing, where specific forcing functions may be required for excitation.

Considering the ability of each technique to control the forcing function level/ amplitude imparted to the structure, the shaker and electromagnetic systems have excellent control over the input levels. The acoustic system has a reasonable degree of control, while the other techniques have very little control and hence repeatability. As far as target footprint area is concerned, the shaker, step relaxation, impact, and electromagnetic systems are capable of imparting essentially point excitation to the structure. This feature is important since most modal parameter extraction procedures require the estimation of frequency response functions at distinct points on a structure. The operational excitation technique has the disadvantage that the forcing function is not measurable and is often distributed over the entire test system. Acoustic excitation is also distributed over the entire structure.

Regarding specific laboratory procedures to be implemented to utilize the excitation procedures, the shaker approach may require significant amount of special hardware to be used, while the operational technique requires no special testing hardware. The step relaxation technique may require some hardware for repeated application of the impulse. The impact technique necessitates a positioning scheme for precise alignment of the hammer for multiple impacts. The acoustic system requires very little special hardware to be implemented. The electromagnetic system may require a moderate amount of hardware since positioning and mounting of the exciter are dependent on the test stand.

A detailed description of the various excitation signals that may be used in vibration/modal testing is given in Sect. 9.2.5 as it is felt that it will be easier to understand this topic after signal processing concepts have been dealt with.

Table 3.6 Summary of non-contact excitation techniques

Excitation technique	Frequency response	Excitation signal control	Signal amplitude control and level	Specimen size	Target area	Special test fixture requirement	Structure
Non-intrusive shaker	Wide; system dependent	Excellent	Excellent; large	Large	Point	Moderate	Rotating or lightweight
Operational	System dependent	None	None; large	Very large	Distributed	None	Bridges, dams, buildings
Step relaxation	System dependent	None	Poor; large	Very large	Point	Moderate	Bridges, windmills, towers
Hammer impact	System and tip dependent	None	Poor; very small	Small	Point	A little	Small or lightweight
Acoustic	Excellent; wide	Good	Good; not very large	Large	Distributed	Moderate	Space vehicles, solar panels
Electromagnetic	Excellent; wide	Excellent	Excellent; not very large	Large	Point	Moderate	Rotating or lightweight

3.12 Signal Conditioning Amplifiers

> **INTERESTING FACTS:** Signal conditioning devices are typically used to convert the output of a transducer to a form (convenient voltage level) that is easily measured or recorded or analysed by the system.
>
> In the absence of these conditioners, some signals cannot be measured at all. The operations that a signal conditioning device performs are amplification of small signals, attenuation of large signals, signal shaping, and isolation of signals from transducers before they are sent to the recording, measurement, or analysis hardware. Other functions include thermocouple compensation for temperature measurements, current sourcing for two-wire and four-wire resistance measurements, filtering to remove system noise and fixing shunt resistors for current measurements or for calibration. Some data acquisition systems have built-in signal conditioning components.

We have seen in this chapter that the principle of operation of each transducer is unique and hence the signal conditioning of each transducer is also unique. A signal conditioner leads to precise measurements. This instrument also performs the functions listed below:

- **Signal conversion:** This process involves data acquisition and conversion of the input electrical or mechanical signal into another easy-to-read electrical output signal which is compatible with data acquisition systems. The output of any transducer may be converted to any standard signal.
- **Linearization:** Many signal conditioners linearize the output signal when a sensor output does not vary linearly with its input physical quantity. The output signal is interpreted using some software; this is true for capacitive and thermocouple signals.
- **Amplification:** Most signal conditioners use signal amplification; this is done by increasing the resolution of the input signal, or by increasing the signal-to-noise ratio. Many different types of amplifiers are used. The commonly used ones are instrumentation amplifiers, which are optimized for use with DC signals, and isolation amplifiers, which isolate high DC levels while passing small AC signals.
- **Filtering:** The signal conditioner filters the signal so as to result in the required frequency range, while removing noise. The filter may be a passive device incorporating capacitors, resistors, and inductors with a maximum gain of one. Alternately, it could be an active device using additional active components such as operational amplifiers and integrated circuits (ICs). Many of the modern signal conditioners use software-based digital filters in which cut-off frequency and gain can be easily adjusted without any hardware.

- **Evaluation and warnings:** Most modern signal conditioners have extra functions for signal evaluation and data preprocessing. They also give warnings and alarms directly via an electrical switching output.
- **Providing interfaces:** Modern signal conditioners need to adhere to standard analogue or digital interfaces and protocols for signal transmission. The present-day digital interfaces are essentially Ethernet-based bus interfaces which allow the connection of several components with only one wire. They also allow additional information to be transmitted, such as diagnostic information of the components.

The signal conditioners for commonly used pickups are shown in Table 3.7. Other transducers may have their own specific signal conditioning units.

Chapter 10 describes many simple laboratory experiments which may be conducted to understand vibration concepts. Many of the pickups and exciters discussed here will come in handy.

List of Symbols

Symbol	Meaning	SI unit
A	area between plates	m^2
B	flux density	T or Wb/m^2
B	amplitude of torsional motion	rad
c	viscous damping coefficient	N.s/m
C	capacitance	F
d	distance between plates or beams/relative displacement	m
d	piezoelectric coefficient	C/N
D	dielectric displacement	C/m^2
e	induced voltage	V
e	eccentricity	m
E	electric field	V/m
f	frequency	Hz
f_A, f_B	frequencies of beams A and B	Hz
f_B	centre frequency of Bragg cell	Hz
f_D	Doppler frequency	Hz
f_P	force to be measured	N
f_s	force generated in shaker	N
$f(t)$	dynamic force	N
F	force	N
F_0	force produced per unit current on mass	N/A
G	amplifier gain	

Table 3.7 Signal conditioners

No.	Transducer	Output level	Signal conditioner
1	Electrodynamic pickup	Low-level voltage (50 mV)	Broadband amplifier
2	Electromagnetic pickup	Low-level voltage (50 mV)	Broadband amplifier
3	Piezoelectric pickup for acceleration/force/pressure	Low-level, high-impedance charge	Charge amplifier
4	Inductive transducer for displacement (LVDT)/angular motion (torsional vibration)/acceleration/pressure	Change in inductance	Carrier frequency amplifier with constant voltage supply
5	Capacitive pickup for displacement, pressure or force	Change in capacitance	Closed-loop AC servo bridge
6	Eddy current probe	Change in inductance	High-frequency carrier amplifier
7	Resistance strain gauge transducer for stress/force/torque	Change in resistance	Carrier frequency amplifier with constant voltage supply
8	Fibre optic transducer	Low-level voltage (50 mV)	Broadband amplifier
9	Semi-conductor strain gauge	Large change in resistance	Constant current bridge
10	Electrodynamic/electromagnetic exciters and PZT exciters	Low-level current	Linear power amplifier

i_c	current	A
$I(t)$, I_R, I_M	light intensity	cd
J	mass moment of inertia	kg.m^2
k	stiffness or spring rate	N/m
k_0	proportionality constant	
K	mixing efficiency coefficient	
l	length of conductor	m
m	vibrating mass	kg
m_s	mass of moving part of shaker	kg
N	number of turns of coil	
P	number of air gaps	
q	torsional stiffness	N.m/rad
Q	charge	C
R	reflectivity of surface	
R_A, R_B	distances of points A and B from shaft centre	m
s	torsional damping	N.m.s/rad
s	compliance	1/Pa
S_t	sensitivity of transducer	
S	strain in piezoelectric material	m/m
t	time	s
T	stress in piezoelectric material	Pa
V	voltage	V
υ	velocity	m/s
v_A, v_B	velocities of points A and B in x-direction	m/s
V_A, V_B	velocities of points A and B	m/s
$x(t)$	vibration displacement	m
x_i, x_0	displacement of casing and mass	m
x_R, x_M	distances travelled by light	m
x, y, z	coordinates, displacements	m
X	amplitude of $x(t)$	m
y	base motion	m
y	instantaneous gap	m
y_0	average gap	m
Y	amplitude of base motion $y(t)$	m
Y_1, Y_2, Y_3, ...	amplitudes of components of base motion $y(t)$	m
z	relative displacement	m
Z	amplitude of relative displacement $z(t)$	m
α, β	angular motions of magnet and core	rad
γ	relative angular motion	rad
Γ	amplitude of γ	rad
δ	depth of eddy current	m
Δl	vibrational displacement	m
ε	permittivity	F/m
ζ	viscous damping ratio	
η	frequency ratio $= \omega/\omega_n$	

Symbol	Meaning	SI unit
κ	dielectric constant	
λ	wavelength	m
μ_0	permeability of free space	H/m or (N·A^{-2})
σ	conductivity	S/m
ϕ	phase lag, relative phase	rad
Φ	magnetic flux	Wb
ϕ_R, ϕ_M	optical phase	rad
ω	angular frequency	rad/s
$\omega_1, \omega_2, \omega_3, \ldots$	frequencies of components of base motion $y(t)$	m
ω_c	cut-off frequency	rad/s
ω_n	natural frequency	rad/s

Abbreviations

AC	alternating current
DC	direct current
DOF	degrees-of-freedom
EEPROM	erasable programmable read-only memory
ESPI	electronic-speckle pattern interferometry
FM	frequency modulated
FRF	frequency response function
IC	integrated circuit
ICP	integrated circuit-piezoelectric
LDV	laser Doppler vibrometer
LSI	large-scale integrated
LVDT	linear variable differential transformer
MEMS	microelectro-mechanical systems
PC	personal computer
PZT	Lead Zirconate Titanate
RMS	root mean square
RPM	revolutions per minute
RVDT	rotary variable differential transformer
SLDV	scanning laser Doppler vibrometer
SNR	signal-to-noise ratio
STIM	smart transducer interface module
TEDS	transducer electronic data sheet
USB	universal serial bus

Questions

1. Name four factors to be considered in the choice of a transducer.
2. Define the term 'sensitivity' of an instrument.
3. What do you mean by dynamic range of a transducer?
4. What is the advantage of a non-contact transducer in vibration measurement?
5. What is the difference between relative and absolute measuring instruments?
6. What is an active transducer? Give two examples.
7. What is a passive transducer? Give two examples.
8. Give examples of two transduction principles which are reversible.
9. What is meant by phase distortion? How can you minimize it?
10. Give two examples of non-contact vibration transducers.
11. Give two examples of contact type of vibration transducers.
12. How do you prevent mass loading?
13. What do you mean by a seismic transducer? Give an example.
14. How is phase distortion minimized in a seismic transducer of the displacement type?
15. How can the sensitivity of an electrodynamic transducer be increased?
16. Can the same signal conditioning amplifier be used for all transducers?
17. How is phase distortion minimized in seismic accelerometers?
18. What signal conditioning amplifier is to be used with LVDTs?
19. What is the principle of operation of a capacitive displacement transducer?
20. If the phase angle of the output with respect to the input in a transducer is proportional to the measured frequency, what is the advantage?
21. What are the applications of an eddy current transducer?
22. What signal conditioning amplifier is to be used with a piezoelectric accelerometer?
23. What is meant by 'skin depth' in an eddy current probe?
24. What transducer is used to measure torsional vibration?
25. Name two devices/equipment which facilitate measurements on rotating machines.
26. What is a signal conditioner? Give an example.
27. Name two transducers which are extremely popular for machinery diagnostics.
28. Sketch a typical experimental setup for vibration measurement in a moving vehicle.
29. Differentiate between bandpass filters used for machinery fault diagnostics and human ride comfort.
30. For vibration measurements in the high-frequency range, which transducer is preferred? Why?
31. What unique benefits does a fibre optic probe offer?
32. What are the limitations in the use of a fibre optic probe?
33. Why are seismic transducers provided with a damping ratio between 0.6 and 0.7?

34. Why is vibration velocity preferred to displacement or acceleration for specifying vibration severity?
35. Give two examples of non-contact vibration exciters.
36. What is the major advantage of a laser Doppler vibrometer?
37. What is the reason for the popularity of the piezoelectric accelerometer?
38. Where is an RVDT used?
39. What are TEDS-based transducers?
40. What do you mean by target footprint?
41. What do you mean by intrusive and non-intrusive excitation techniques?
42. Name an exciter which can be used for testing huge structures like bridges, foundations, etc.
43. Which shakers are best for fatigue testing?
44. State a drawback of a mechanical type of exciter.
45. Give examples of two non-intrusive excitation techniques.
46. How is eccentricity changed in a double mass reaction exciter?
47. Name two exciters which are good for large test loads.
48. What is a drawback of a double mass reaction exciter?
49. Which exciters give a constant displacement irrespective of frequency?
50. What is meant by an electrodynamic transducer?
51. Where are hydraulic/electrohydraulic exciters used?
52. Name some applications requiring triaxial vibration measurement.
53. Which exciters are best suited to apply a static load with a superimposed dynamic load?
54. What is meant by a modal shaker?
55. What is a double mass reaction exciter?
56. Why are stingers sometimes pre-tensioned?
57. Which exciters have a high-frequency range of operation?
58. Name a shaker that gives a large excitation force for a small weight of the shaker.
59. Which exciters give a large stroke?
60. What is meant by a seismic mass?
61. What is meant by the FRF of a transducer?
62. What are PZT actuators?
63. What environmental factors are to be considered in the choice of a transducer?
64. Which exciters are affected by stray magnetic fields?
65. Why is the output of an electrodynamic exciter sent through a power amplifier?
66. How do you measure the excitation force imparted by an electrodynamic shaker?
67. What is meant by a stinger? Why is it used?
68. How can you increase the output of a piezoelectric actuator?
69. Name some commonly used piezoelectric materials used in pickups and actuators.
70. What is meant by a slip table?
71. What is meant by step relaxation technique?
72. What is meant by ambient testing?
73. Name an exciter which is suitable for exciting rotating structures.
74. Where is operational excitation technique used?

75. What is a modal hammer?
76. Which non-intrusive vibration transducer has the best resolution?
77. What decides the frequency range of a modal hammer?
78. What are the problems with impact excitation?
79. Which excitation technique has the minimum test fixture requirement?
80. What is meant by acoustic excitation?
81. Can a capacitance probe be used on greasy surfaces? Why?
82. Name two exciters which offer good excitation signal control.
83. What advantage does laser holography have over other vibration measurement techniques?
84. What are the difficulties associated with the use of a fibre optic probe?
85. Why is laser holography not very popular in spite of its good resolution?
86. What are the front slope and back slope regions of a fibre optic probe?
87. Which exciters offer a small target area?
88. Name an exciter which is suitable for exciting lightweight structures.

Strike Off What Is Not Applicable and Give Reasons

89. An accelerometer is a transducer with a high/low natural frequency and is meant to be used above/below its natural frequency.
90. An LVDT is a transducer with a high/low natural frequency and is meant to be used above/below its natural frequency.
91. When the frequency of operation of a machine is not known, it is best to use a displacement/velocity/acceleration transducer. Why?
92. For vibration measurements at low frequencies, a displacement pickup/accelerometer is suitable because _____.

Fill in the Blanks

93. A vibration velocity of 15 mm/s at 100 Hz is _____ g and _____ mm.
94. A transducer suitable for vibration measurements on a rotating structure is _____. The main advantages and drawbacks of this transducer are _____.
95. For exciting in the 0 to 800 Hz range, an impact hammer with a _____ tip is desirable.
96. At resonance, the phase lag of the output of a seismic transducer with respect to its input is _____.
97. An electrodynamic transducer is very popular because _____.

Bibliography

1. Boyes, W. (2010). *Instrumentation reference book*. Amsterdam: Elsevier.
2. Buzdugan, Gh., Mihăilescu, E., & Rades, M. (1986). *Vibration measurement (Mechanics: Dynamical systems)*. Netherland: Springer.

3. Christian, L. (2002). *Mechanical vibration and shock*. New York: Taylor & Francis.
4. Collacott, R. A. (1979). *Vibration monitoring and diagnosis*. London: George Godwin Ltd.
5. Cook, R. O., & Hamm, C. W. (1979). Fibre optic lever displacement transducer. *Applied Optics, 18*(19), 3230–3241.
6. de Silva, C. W. (2007). *Vibration monitoring, testing, and instrumentation*. Boca Raton: CRC Press.
7. Doebelin, E. O. (2004). *Measurement systems: Application and design*. New York: McGraw Hill Professional.
8. Ewins, D. J., & Inman, D. J. (2001). *Structural dynamics @ 2000: Current status and future directions*. Baldock: Research Studies Press Ltd.
9. Gatti, P., & Ferrari, V. (1999). *Applied structural and mechanical vibrations-theory, methods and measuring instrumentation*. New York: Taylor & Francis.
10. Hoogenboom, L., Allen, G. H., & Wang, S. (1984). Theoretical and experimental analysis of a fiber optic proximity probe. In *Proceedings of the S.P.I.E. Technical Symposium, East 84, Arlington, Virginia, April 29*, Paper #478-25.
11. Inman, D. J. (2001). *Engineering vibration*. Upper Saddle River: Prentice Hall.
12. McConnell, K. G. (1995). *Vibration testing, theory and practice*. New York: Wiley.
13. Murty, D. V. S. (2004). *Transducers and instrumentation*. New Delhi: Prentice-Hall of India Pvt. Ltd.
14. Northrop, R. B. (1997). *Introduction to instrumentation and measurements*. Boca Raton: CRC Press.
15. Nakra, B. C., & Chaudhary, K. K. (1985, 2004). *Instrumentation, measurement and analysis*. New Delhi: Tata McGraw-Hill.
16. Patton, M. E., & Trethewey, M. W. (1987). A technique for non-intrusive modal analysis of very lightweight structures. In *Proceedings of the 5th International Modal Analysis Conference, 6–9 April*. London: Imperial College of Science.
17. Patton, M. E., & Trethewey, M. W. (1987). A survey and assessment of nonintrusive-modal-testing techniques for ultralightweight structures. *International Journal of Analytical and Experimental Modal Analysis, 2*(4), 163–173.
18. Piersol, A. G. (2002). *Harris' shock and vibration handbook*. New York: McGraw-Hill Professional.
19. Randall, R. B. (2010). *Vibration based condition monitoring*. New York: Wiley.
20. Ramamurti, V. (2000). *Mechanical vibration practice with basic theory*. New Delhi: Narosa Publishing House.
21. Rangan, C. S., Sarma, G. R., & Mani, V. S. V. (1983). *Instrumentation devices and systems*. New Delhi: Tata McGraw-Hill Publishing Company Ltd.
22. Rao, B. K. N. (1996). *The handbook of condition monitoring*. Oxford: Elsevier.
23. Redl, W. A. (1985). *Noise and vibration measurement, prediction and mitigation*. New York: American Society of Civil Engineers.
24. Reeves, C. W. (1999). *The vibration monitoring handbook*. Oxford: Coxmoor.
25. Reza Moheimani, S. O., & Fleming, A. J. (2006). *Piezoelectric transducers for vibration control and damping (Advances in industrial control)*. New York: Springer.
26. Schmitz, T. L., & Smith, K. S. (2012). *Mechanical vibrations, modeling and measurement*. New York: Springer.
27. Smith, J. D. (1989). *Vibration measurement and analysis*. London: Butterworth.
28. Thomson, W. T., & Dahleh, M. D. (2003). *Theory of vibration with applications*. India: Pearson Education India.
29. Wilson, J. S. (2005). *Sensor technology handbook*. Oxford: Elsevier, Newnes.
30. Wowk, V. P. E. (1991). *Machinery vibration: Measurements and analysis*. New York: McGraw-Hill.
31. https://www.bksv.com.
32. https://www.dytran.com.
33. https://www.globalspec.com.
34. https://www.hbm.com.

35. https://www.mbdynamics.com.
36. https://www.micro-epsilon.co.uk.
37. https://www.mide.com.
38. https://www.mtiinstruments.com.
39. https://www.optonor.com.
40. https://www.pcb.com.
41. https://www.sensorland.com.
42. https://www.sensorsportal.com.
43. https://sdyn.in.
44. https://www.sherbornesensors.com.
45. https://www.xcitesystems.com.

Chapter 4
Fundamentals of Acoustics

INTERESTING FACTS: Mankind has been interested in acoustics since the creation of the first musical instruments. Musicians and philosophers have tried to understand the laws of sound production and use them in the development of more and more sophisticated musical instruments. As early as 4000 B.C., music had evolved to a great extent and was appreciated by the Chinese, Hindus, Japanese, and Egyptians. These people observed that there were rules in music, but there were no scientific studies. As early as 3000 B.C., the Egyptians depicted stringed instruments on the walls of their tombs. Pythagoras (582–507 B.C.), the Greek mathematician and philosopher is considered to be the first person to scientifically investigate musical sounds. He conducted experiments on a vibrating string using a simple device called a monochord. The Romans gained knowledge of music from the Greeks. Vitruvius, a famous Roman architect, wrote around 20 B.C. about the acoustic properties of theatres. During the Islamic golden age, Abū Rayhān al-Bīrūnī is said to have proposed that the speed of sound was much lower than that of light. The physics behind acoustical processes began to be understood rapidly during and after the Scientific Revolution. Scientists like Galileo Galilei, Marin Mersenne and Sir Isaac Newton contributed to this field immensely. After the eighteenth century, major advances in acoustics were made due to the works of Hermann von Helmholtz, Baron Rayleigh, Sir Charles Wheatstone, Georg Ohm, Joseph Henry, and Wallace Clement Sabine.

Acoustics deals with the study of sound and mechanical waves in gases, liquids, and solids and has its origin in the study of vibrations and their radiation as acoustic waves. It is concerned with all aspects of sound: production, propagation, control,

transmission, and reception. These pertain to sounds created and received by human beings, machines and measuring instruments. Sound is defined as any dynamic pressure fluctuation superimposed on the mean atmospheric pressure and impinging on the ear drum; it includes all sounds from the weakest which are barely audible, to sounds which cause pain and damage hearing.

This chapter deals with fundamental aspects of acoustics, without a knowledge of which the acoustic measurement techniques which follow in Chaps. 5 and 11 cannot be fully appreciated. Fundamental definitions for describing sound and aspects related to the physics of sound, as well as terminology used in acoustic measurements are described. The theory behind sound propagation, transmission, and the behaviour of sound in enclosed spaces has also been discussed.

4.1 Human Perception of Sound

DID YOU KNOW that the human perception or the psychophysical study of acoustics is called psychoacoustics? The early Greeks tried to seek mathematical explanations for many aspects of music. Though they did not have a name for their studies, they were engaged in psychoacoustics just as much as Gustav Fechner, the present-day Father of psychophysics, and other scientists. The concept of pitch was developed by the time of Pythagoras, but the relation between pitch and frequency was understood only during the time of Galileo in the sixteenth century. Around 350 B.C., Aristotle reported studies on music and sound; he observed that the voice is sweeter than the sound of instruments and that the sound of the flute is sweeter than that of the lyre. Robert Hooke conducted experiments in the seventeenth century to find a relation between the pitch and frequency of a vibrating string. Hermann von Helmholtz's book "On the Sensations of Tone as a Physiological Basis for the Theory of Music" was a major reference for hearing and musical perception for many decades. Helmholtz was influenced by Georg Simon Ohm's "acoustic law" which stated that the ear performs a limited Fourier analysis by determining the harmonic components of a complex sound wave. In 1876, Rayleigh proposed his "binaural ratio" theory for localization of a sound source. Harvey Fletcher, as Director of acoustical research in the Bell Laboratories in the 1990s oversaw a large number of psychoacoustic research projects like measurements of auditory thresholds, intensity and frequency discrimination, tone-on-tone masking, tone-in-noise masking, the phon scale of loudness and the articulation index. There have subsequently been a lot of studies in this area by other researchers.

We necessarily have to deal with sound in our day-to-day life, whether it be spoken communication, music or noise produced by machines. The study of acoustics includes what is called sound, as well as noise. From the point of view of acoustics, both sound and noise are caused by pressure fluctuations about the mean atmospheric

value. There is a big difference though, in the way the two are sensed and this is also highly subjective. In general, higher the sound intensity, higher is the perceived loudness. What one person perceives as sound may very well be perceived as noise by somebody else. In general, noise is defined as any unpleasant or undesirable sound. Typically sounds which have a structured waveform or spectrum are perceived by the ear as being pleasant and would be called music, while those which are unstructured sound unpleasant to the ear and are called noise. Music may be considered to be sound, not noise, since it is pleasant to most people; noise from machinery on the other hand gives pleasure to none other than the owner of the machinery, to whose ears it may sound like music since it is bringing him money! The level of annoyance felt due to noise is highly subjective, varying from person to person, time of the day, health condition, mood, and so on. It also depends on the pitch, loudness and tonal quality of the sound. In general, higher sound levels are tolerated better during the day than at night. Sounds can also cause damage, examples being sonic booms from aircraft or the music of a soprano singer which can shatter windows and glasses.

Noise has become an unavoidable part of everyday life with noise emanating from machines, factories, traffic, etc. For the study of human exposure to noise and for the establishment of noise criteria, it is required to have a good understanding of the way the human ear responds to it. Noise is increasingly being perceived as a pollutant and a serious health hazard these days. Studies have shown that in addition to inducing a threshold shift (hearing loss, possibly restricted to distinct frequency ranges), noise also causes irritability, reduces concentration, increases blood pressure, affects blood circulation, causes cardiac problems and changes resistance of the skin to electric potential. It is therefore important that steps towards noise control are taken. In connection with this drive against noise, legislations have been made to control specific noise-making activities, based on different kinds of zones. In India, allowable noise limits for each zone are set by the Central Pollution Control Board of India.

4.1.1 Sound Pressure and Sound Pressure Level (SPL)

The sensation of sound is caused by pressure fluctuations or oscillations in air. The response of the human ear to sound or noise depends on the sound frequency, the sound pressure level and the pressure waveform; besides, there is a high degree of subjectivity involved. For a young human ear which is in good health and which has not been affected by too much exposure to excessively loud sounds, the range of these pressure amplitudes for audio-acoustics is very large, from 2×10^{-5} N/m^2 or 20 μPa, the threshold of audibility (roughly the sound of a mosquito flying 3 m away) to 20–63 N/m^2, the threshold of pain. The acoustic pressures involved in audio-acoustics are seen to be very small fluctuations in the mean atmospheric pressure

Table 4.1 Typical sound pressure levels L_p (dB)

Sound source	Sound pressure (Pa)	Sound pressure level L_p (dB)
Sound wave at 1 atmosphere (theoretical limit)	101,325	194
Large rocket engine (nearby)	20,000	180
Jet engine at 30 m	630	150
Threshold of pain/ Deepavali atom bomb	63	130
Jet take-off (60 m)	20	120
Jack hammer at 1 m/discotheque	2	100
Major road traffic at 10 m	6×10^{-1}	90
Moving automobile at 10 m	6×10^{-2}	70
Normal conversation	$2 \times 10^{-3} - 2 \times 10^{-2}$	40–60
Quiet residential area	2×10^{-3}	40
Soft whisper (5 m)	6×10^{-4}	30
Normal breathing	6×10^{-5}	10
Threshold of audibility for a healthy young person	2×10^{-5}	0

which has a value of 1.013×10^5 Pa. Due to the remarkably large range of pressures involved (more than a million to 1, or 6 decades), some kind of compressed scale or a logarithmic representation of this quantity (using ratios) is more appropriate than a linear representation which would mean using very large and unwieldy numbers. Besides, the ear responds logarithmically, not linearly, to acoustic stimuli, in terms of both intensities and frequencies. For these reasons, quantities like acoustic pressure, intensity and power (to be discussed later) are represented in terms of the deciBel (dB), which is the logarithm of the ratio of any of the said quantities to a reference level. SPL L_p gives an indication of the loudness of a sound and is the most important acoustic quantity that is measured. It is defined as

$$L_p = 10 \log_{10} \left(\frac{p_{\text{rms}}^2}{p_{\text{ref}}^2} \right) = 20 \log_{10} \left(\frac{p_{\text{rms}}}{p_{\text{ref}}} \right) \tag{4.1}$$

where p_{ref} is the reference sound pressure corresponding to the threshold of audibility of human hearing at the acoustic reference frequency of 1000 Hz and is equal to 2×10^{-5} N/m^2 or 20 μPa; p_{rms} refers to the root mean square (RMS) value of the pressure being measured. A pressure equal to the reference value works out to 0 dB, while 1 Pa is equivalent to 93.98 dB. Table 4.1 gives the SPLs due to some common activities.

Table 4.2 Relationship between power ratio, pressure ratio and change in dB

Power ratio n	Pressure ratio $n^{1/2}$	Change in sound pressure level L_p (dB) $= 10 \log_{10} n = 20 \log_{10} n^{1/2}$
10,000	100	40
100	10	20
4	2	6
2	1.4	3
1	1	0
0.5	0.7	-3
0.25	0.5	-6
0.01	0.1	-20
0.0001	0.01	-40

Table 4.3 Perception of loudness related to L_p (dB)

Change in SPL (dB)	Change in perceived loudness
1	Insignificant
3	Just perceptible
5	Clearly perceptible
10	Twice or half as loud
15	Significant change
20	Four or 1/4 times as loud

Table 4.2 gives the relationship between pressure ratio, power (proportional to pressure squared) ratio and change in SPL in dB. Though the loudness perceived by human beings is correlated to the SPL in dB, which is related to intensity or pressure squared, there is no linear relationship between the two. An increase in SPL of 3 dB, corresponding to a pressure ratio of 1.4 or a doubling of sound energy (pressure squared) is just perceptible to the normal human ear. Thus if a manufacturer makes claims that his/her product produces a sound pressure only half that of his/her rival's product, one should be wary; the change in SPL would be noticeable, but not significantly lower. A change of 10 dB (corresponding to a pressure ratio of 3.16) is perceived as twice or half as loud. Table 4.3 summarizes the subjective perception of changes in SPL in dB.

4.1.2 Frequencies of Interest and Frequency Weighting

Normally, healthy young human beings can detect sounds with frequencies in the range 20–20,000 Hz. With age, the sensitivity to higher frequencies reduces. Pitch is related to frequency; low frequencies are perceived as being low-pitched, while high frequencies are identified as being high-pitched. Pitch is thus the subjective response of the human ear to the spectral content of sound. The frequencies involved in the study of sound encompass the following ranges:

  (i) Sonic: 20–20,000 Hz or the audio-frequency range.
 (ii) Ultrasonic: above 20,000 Hz; this is not considered damaging at levels below 105 dB.
(iii) Infrasonic (subsonic): 20 Hz; these sounds are felt, but not heard and are damaging at levels above 120 dB. Infrasounds as well as ultrasounds affect human senses and cause discomfort.

The sensitivity of the human ear to different frequencies varies, with minimum sensitivity to extremely low and extremely high frequencies. For example, a pure tone of 1000 Hz with an SPL of 40 dB would sound louder than a pure tone of 80 Hz at 50 dB. Besides, a 1000 Hz tone at 70 dB would give the same subjective perception of loudness as a 50 Hz tone at 85 dB. The minimum audible sound level occurs at about 4000 Hz. The frequency range 500 Hz and 2 kHz is most crucial to human beings since speech signals are in this range.

Frequency analysis is required for noise source identification. There are many instruments available for carrying out spectrum analysis of acoustic signals as described in Chap. 5. For the evaluation of human exposure to noise, the sound measuring system must take into account the varying sensitivity of the ear over the audio-frequency range. This is done using internationally standardized frequency weighting networks such as the *A*, *B*, *C* and *D* networks that weight the contributions of the different frequencies to the subjective perception of loudness of sound while arriving at the overall SPL. The two weighting networks most commonly used are the *A* and *C* networks, which correlate with the frequency response of the human ear for different sound levels. Their characteristics are specified in IEC 60651 and have been described in Sect. 5.4.2. While frequency analysis of vibration signals is done using constant bandwidth narrow band filters or fast Fourier transform (FFT) algorithms, constant percentage bandwidth analyzers are used for acoustic signals for the reason that the human ear senses frequencies in a logarithmic fashion. In order to maintain uniformity of measurements using different instruments for frequency analysis, 1/1 and 1/3 octave bands have been standardized by the International Standards Organization; these are described in Sect. 5.4.3.

4.2   Sound Wave Propagation in 1-Dimension (1-D)

INTERESTING FACTS: Do you know that Ernst Chladni and Lord Rayleigh are both credited with the title "Father of Acoustics"? Both were multi-faceted personalities. Chladni's most important work, for which he is called "Father of Acoustics", was his work on vibrating plates and their mode shapes and the calculation of the speed of sound in different gases. Chladni also did pioneering work in meteors and the theory of their extraterrestric origin, which earned him the title "Father of Meteoritics". Rayleigh did not get the Nobel prize in Physics (in 1904) for his works in acoustics, but for his discovery of argon along with William Ramsay. Rayleigh also explained why the sky is blue, a phenomenon now called Rayleigh scattering. He also predicted the existence of surface waves which are now called Rayleigh waves. Rayleigh's textbook "The Theory of Sound", to this day is considered a classical textbook on acoustics. The rayl, the unit of acoustic impedance, as well as a type of surface wave known as a Rayleigh wave are named after him. Also there are craters on Mars and the Moon which are named in his honour.

When a sound source such as a tuning fork vibrates, it causes pressure fluctuations in the surrounding air. These variations spread out from the source and are similar to the ripples in a pond caused by a stone thrown in the water. Though the fluctuations move, the air itself does not move away from the source. Such vibrational disturbances that propagate in a fluid or gaseous medium are called acoustic waves. In such wave motion, pressure disturbances may be generated by a vibrating surface or by turbulent fluid flow and they propagate as longitudinal or compressional waves in the elastic medium (or as stress waves in a solid). The particles oscillate back and forth in the medium in the direction of the waves, resulting in alternate compressions and rarefactions of the particles of the medium as the sound wave passes a given location (Fig. 4.1). The rate at which these disturbances travel through the medium depends on the speed with which the molecules transfer energy from one to another. The study of sound wave propagation with air as the medium constitutes audio-acoustics, while that in water is called under water acoustics. Three basic elements are involved in sound wave propagation: source, medium, and receiver. It is to be noted that sound cannot propagate in vacuum. If the medium is infinite, the waves will propagate in all directions and get attenuated rather quickly. If the sound waves are constrained to travel in one direction only, as happens when they propagate in a fluid-filled narrow tube, the waves travel without alteration in shape or size if the disturbances are small in the first place and if the dissipation in the medium is negligible.
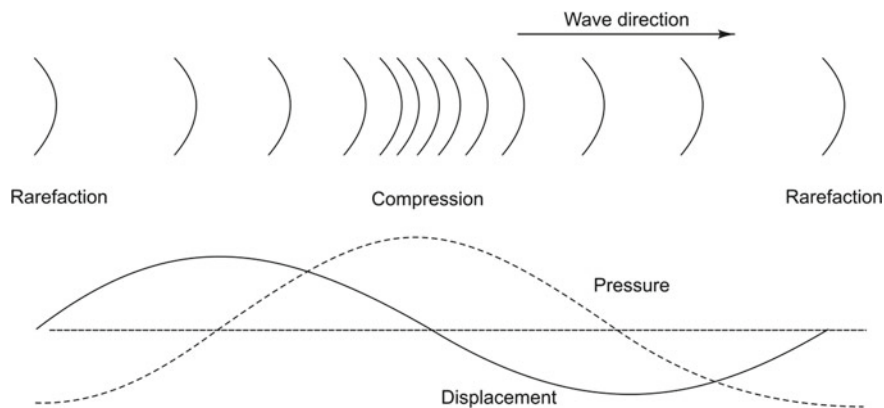
Fig. 4.1 Longitudinal wave

4.2.1 Longitudinal Waves in a Column of Gas

The most popular model for analyzing wave motion is the particle model. A particle is defined as an exceedingly small element consisting of several million molecules with uniform density, velocity, and temperature throughout the element. What typically happens in wave propagation is that a disturbance imparts momentum to some adjacent particles. These particles collide with other particles, transferring momentum to them and setting them into motion. The speed with which the wave propagates is known as the characteristic speed of sound in the medium. Every particle, after colliding with an adjacent particle, oscillates about its equilibrium position and the amplitude of oscillation can be sensed by measuring the sound pressure, which is nothing but the variation over and above the static pressure in the fluid. The equation representing such disturbances is called the wave equation in 1-D. The assumptions made in deriving this equation are as follows:

   (i) The fluid in which the wave propagates is homogeneous and isotropic.
  (ii) The fluid is in thermodynamic equilibrium, except near the propagating wave.
 (iii) The fluid is an ideal gas.
  (iv) It obeys Hooke's law (has linear stress–strain behaviour).
   (v) The body force acting on the fluid is negligible.
  (vi) Damping forces are negligible.
 (vii) The thermodynamic process involved in wave propagation is adiabatic and reversible.
(viii) The motion involved in propagation is so small that products resulting in second and higher order terms are negligible.

   Let us derive the equation of motion for longitudinal waves in a gas. Let us assume that the tube in which it is propagating is so narrow that at any instant, the displacements at all points on any cross-sectional area S are the same. It is

Fig. 4.2 Displacement of an element due to wave propagation
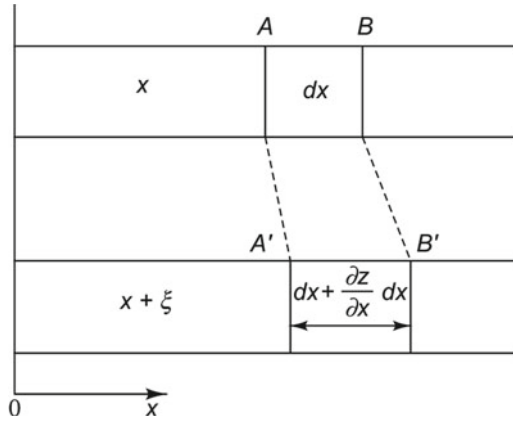
also assumed that there is no friction along the sides of the tube. Figure 4.2 shows the displacement of an element of gas in a tube of cross-sectional area S due to longitudinal wave propagation.

The wave equation may be obtained using Newton's law of motion and the elastic behaviour of the gas. In this figure x and $x + \xi$ indicate the original and displaced positions of the element along the tube. AB indicates the undisturbed element and $A'B'$ that due to the passage of the longitudinal wave. The increase in the volume of the element is $S(\partial \xi / \partial x)dx$.

(i) **Newton's law of motion**

The total force acting on the displaced element $A'B'$ is

$$F = (P_{A'} - P_{B'})S \qquad (4.2a)$$

Here P denotes the total pressure which can be written as a dynamic pressure p superimposed over the static pressure P_0. Therefore,

$$F = -\frac{\partial p}{\partial x} S dx \qquad (4.2b)$$

It is seen that P_0 disappears from the differential equation. From the law of conservation of mass, we have

$$\text{Mass of element } A'B' = \text{mass of element } AB = S\rho_0 dx \qquad (4.3)$$

where ρ_0 is the mean density of gas. The acceleration of the element is $\partial^2 \xi / \partial t^2$. Substituting this in Eq. (4.2b), we can write Newton's law as

$$S\rho_0 dx \frac{\partial^2 \xi}{\partial t^2} = -\frac{\partial p}{\partial x} S dx$$

Or

$$\frac{\partial p}{\partial x} = -\rho_0 \frac{\partial^2 \xi}{\partial t^2} = -\rho_0 \frac{\partial u}{\partial t} \qquad (4.4)$$

Here u is the particle velocity.

(ii) **Elastic behaviour**

The change in pressure of the gas results in a change in its volume due to its

compressibility. Since the frequencies of pressure fluctuations associated with acoustic waves are high, in the range 20 Hz–20 kHz for audio-acoustics, the thermodynamic process is assumed to be adiabatic. Hence,

$$P_0 V_0^\gamma = K \qquad (4.5)$$

where V_0 is the undisturbed volume, γ is the ratio of the specific heats of the gas and K is a constant. Differentiating Eq. (4.5), we get

$$\gamma P_0 V_0^{\gamma-1} dV + V_0^\gamma dP = 0 \qquad (4.6)$$

dP in the above equation is the same as the dynamic pressure p. Hence, the above equation may be written as

$$p = -\gamma P_0 \frac{dV}{V_0} \qquad (4.7)$$

Here the undisturbed volume V_0 is equal to Sdx and the change in volume dV is $S(\partial \xi / \partial x)dx$. Therefore

$$p = -\gamma P_0 \frac{\partial \xi}{\partial x} \qquad (4.8)$$

Differentiating the above equation with respect to x and combining with Eq. (4.4) to eliminate p, we get

$$\gamma P_0 \frac{\partial^2 \xi}{\partial x^2} = \rho_0 \frac{\partial^2 \xi}{\partial t^2}$$

or

$$\frac{\partial^2 \xi}{\partial x^2} = \frac{1}{c_0^2} \frac{\partial^2 \xi}{\partial t^2} \qquad (4.9)$$

where

$$c_0^2 = \gamma P_0 / \rho_0 \qquad (4.10)$$

It has been assumed here that the fluctuating part of wave speed is negligible. The uni-directional wave equation described above is in terms of dynamic displacement ξ. However, the quantity of interest in acoustics is dynamic pressure p which is

related to the displacement by Eq. (4.8). Substituting $\rho_0 c_0^2$ for γP_0 in this equation, we get

$$p = -\rho_0 c_0^2 \frac{\partial \xi}{\partial x} \tag{4.11}$$

The wave equation may also be obtained directly in terms of pressure by differentiating Eq. (4.4) with respect to x as shown below.

$$\frac{\partial^2 p}{\partial x^2} = -\rho_0 \frac{\partial^2}{\partial t^2} \left(\frac{\partial \xi}{\partial x} \right) \tag{4.12}$$

Substituting for $\partial \xi / \partial x$ from Eq. (4.11) into the above equation gives

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{c_0^2} \left(\frac{\partial^2 p}{\partial t^2} \right) \tag{4.13}$$

Thus the 1-D wave equation has the general form

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{1}{c_0^2} \frac{\partial^2 \phi}{\partial t^2} \tag{4.14}$$

The quantity ϕ may represent the pressure amplitude or particle velocity. This equation is the same as that for the transverse vibration of a string where the displacement is perpendicular to the direction in which the wave is travelling. It is the wave equations in terms of particle pressure (Eq. 4.13) and displacement (Eq. 4.9) which have exactly the same form for 1-D propagation; the former is the most commonly used form of wave equation in acoustics.

Equation (4.14) may be solved by separation of variables. It may be written as shown below, to give two separate differential equations.

$$\phi = F(x)G(t) \tag{4.15}$$

The general solution for wave propagation would then take the form

$$\phi = g(c_0 t - x) + h(c_0 t + x) \tag{4.16}$$

Here g and h are arbitrary, independent functions. The first term in the solution represents a wave of constant shape moving in the positive x direction with velocity c_0, while the second term represents a wave of constant shape travelling in the negative x direction with velocity c_0. Let us consider a harmonic wave, which is the simplest solution, travelling in the positive x direction. It may be represented as

$$\phi = a_1 \cos k(c_0 t - x) \tag{4.17}$$

where a_1 may represent the peak pressure and k is the wavenumber and is related to the wavelength λ as shown below.

$$k = \frac{2\pi}{\lambda} \tag{4.18a}$$

The period T of the wave is given by

$$T = \frac{2\pi}{kc_0} = \frac{\lambda}{c_0} \tag{4.18b}$$

and the frequency in Hz is

$$f = \frac{1}{T} = \frac{c_0}{\lambda} \quad \text{or} \quad c_0 = f\lambda \tag{4.18c}$$

It is good to have a rough idea regarding correspondence between wavelength and frequency. Considering that the speed of sound in air is 343 m/s, at 1 kHz the wavelength is close to 34 cm, at 20 Hz it is close to 17 m and at 20 kHz it is only 1.7 cm.

Writing Eq. (4.17) in terms of angular frequency $\omega = 2\pi f$ in rad/s, we have

$$\phi = a_1 \cos(\omega t - kx) \tag{4.19}$$

A more general solution to the forward travelling wave may be written in terms of a combination of sine and cosine waves; this can describe any desired harmonic waveform by an appropriate choice of a_1 and a_2.

$$\phi = a_1 \cos(\omega t - kx) + a_2 \sin(\omega t - kx) \tag{4.20}$$

A general solution consisting of both the forward and backward travelling waves is

$$\phi = a_1 \cos(\omega t - kx) + a_2 \sin(\omega t - kx) + a_3 \cos(\omega t + kx) + a_4 \sin(\omega t + kx) \tag{4.21a}$$

This harmonic solution may be written in terms of complex exponential functions as shown below.

$$\boldsymbol{\phi} = \boldsymbol{A}e^{i(\omega t - kx)} + \boldsymbol{B}e^{i(\omega t + kx)} \tag{4.21b}$$

Here $\boldsymbol{A}$ and $\boldsymbol{B}$ are, in general, complex quantities and are shown in boldface. It is only the real part of the solution which is used, the imaginary part being ignored though it is used in the computation.

Particle velocity is the velocity of a particle in the medium as it transmits a wave and is not the same as the speed of sound. The particle velocity u can be obtained from particle pressure by writing Eq. (4.4) in the form shown below.

$$\frac{\partial u}{\partial t} = -\frac{1}{\rho_0}\left(\frac{\partial p}{\partial x}\right) \tag{4.22}$$

Integrating the above equation over t, we get

$$u = -\frac{1}{\rho_0} \int \frac{\partial p}{\partial x} dt \qquad (4.23)$$

The solution to the wave equation in terms of particle pressure for the forward travelling wave only, may be written as

$$p_+ = a_1 \cos(\omega t - kx) + a_2 \sin(\omega t - kx) \qquad (4.24)$$

Differentiating w.r.t. x, we get

$$\frac{\partial p_+}{\partial x} = ka_1 \sin(\omega t - kx) - ka_2 \cos(\omega t - kx) \qquad (4.25)$$

Integrating the above equation w.r.t. t, we get

$$\int \frac{\partial p_+}{\partial x} dt = -\frac{k}{\omega} a_1 \cos(\omega t - kx) - \frac{k}{\omega} a_2 \sin(\omega t - kx) = -\frac{k}{\omega} p_+ \qquad (4.26)$$

From Eq. (4.23) it follows that

$$u_+ = \frac{k}{\omega \rho_0} p_+ = \frac{p_+}{\rho_0 c_0} = \frac{\partial \xi_+}{\partial t} \qquad (4.27a)$$

$$u_- = -\frac{k}{\omega \rho_0} p_- = -\frac{p_-}{\rho_0 c_0} = \frac{\partial \xi_-}{\partial t} \qquad (4.27b)$$

These expressions relate particle pressure and particle velocity. The total particle pressure and velocity are as shown below.

$$p = p_+ + p_- \quad \text{and} \quad u = u_+ + u_- \qquad (4.28)$$

Figure 4.3 shows the phase relationships between ρ, u, ξ and $\partial \xi / \partial x$ for the waves travelling in the positive and negative x-directions.

Two common boundary conditions which are considered while solving are as follows:

(i) closed tube for which $(u)_x = 0$ at the closed end for all time t and
(ii) open tube for which $(p)_x = 0$ at the open end for all time t. For the open tube, the disturbances extend slightly beyond the open end, and hence the boundary condition may be applied at the imaginary end. It has been shown in the literature that the end correction is $0.6r$ for a tube of radius r with no flange and $8r/3\pi$ for a tube with a large flange.

Let us consider an unflanged tube of length L, closed at one end and open at the other end. The pressure and particle velocity are expressed by the following equations.
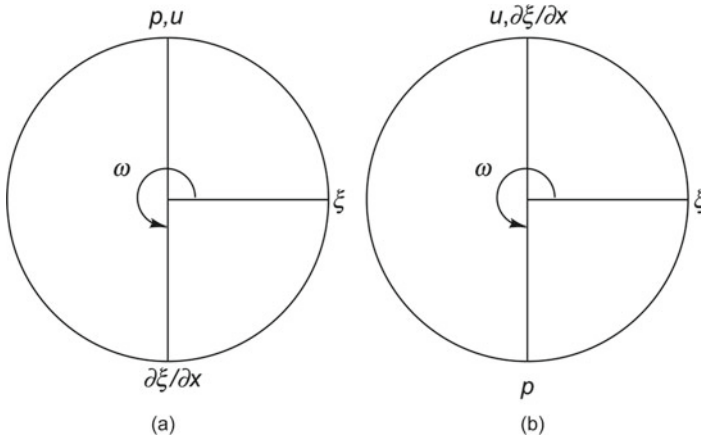
Fig. 4.3 Phase relationships: **a** forward travelling wave, **b** backward travelling wave

$$p = a_1 \cos(\omega t - kx) + a_2 \sin(\omega t - kx) + a_3 \cos(\omega t + kx) + a_4 \sin(\omega t + kx) \tag{4.29}$$

$$u = \frac{a_1}{\rho_0 c_0} \cos(\omega t - kx) + \frac{a_2}{\rho_0 c_0} \sin(\omega t - kx)$$
$$- \frac{a_3}{\rho_0 c_0} \cos(\omega t + kx) - \frac{a_4}{\rho_0 c_0} \sin(\omega t + kx) \tag{4.30}$$

Applying the first boundary condition, i.e., $(u)_x = 0$ for all t, we get

$$0 = a_1 \cos(\omega t) + a_2 \sin(\omega t) - a_3 \cos(\omega t) - a_4 \sin(\omega t) \tag{4.31a}$$

Hence,

$$a_3 = a_1; a_4 = a_2 \tag{4.31b}$$

Substituting these values, the expression for pressure in Eq. (4.29) gets simplified to

$$p = 2 \cos kx (a_1 \cos \omega t + a_2 \sin \omega t) \tag{4.32}$$

This represents a stationary waveform with a pressure antinode (maximum) at the rigid end near the origin. Applying the second boundary condition, namely, that the tube is open at $x = L$, or at $x = L'$ to be more precise with the application of end correction, we have for all values of t

$$(p)_{x=L'} = 0 \tag{4.33a}$$

where

$$L' = L + 0.6r \tag{4.33b}$$

Hence,

$$\cos kL' = 0 \tag{4.33c}$$

i.e.,

$$kL' = (2n - 1)\pi/2, \quad n = 1, 2, 3 \ldots \tag{4.33d}$$

Therefore,

$$\lambda = \frac{4L'}{(2n - 1)} \quad \text{or} \quad f = \frac{(2n - 1)c_0}{4L'} \tag{4.34}$$

Figure 4.4 depicts the pressure variation in an unflanged pipe of length L closed at one end and open at the other for the first four modes.

As in the case of a vibrating system, the pipe sustains natural vibrations or resonances only at certain frequencies. These natural frequencies are found to be func-

Fig. 4.4 Pressure variation in a pipe closed at one end and open at the other end
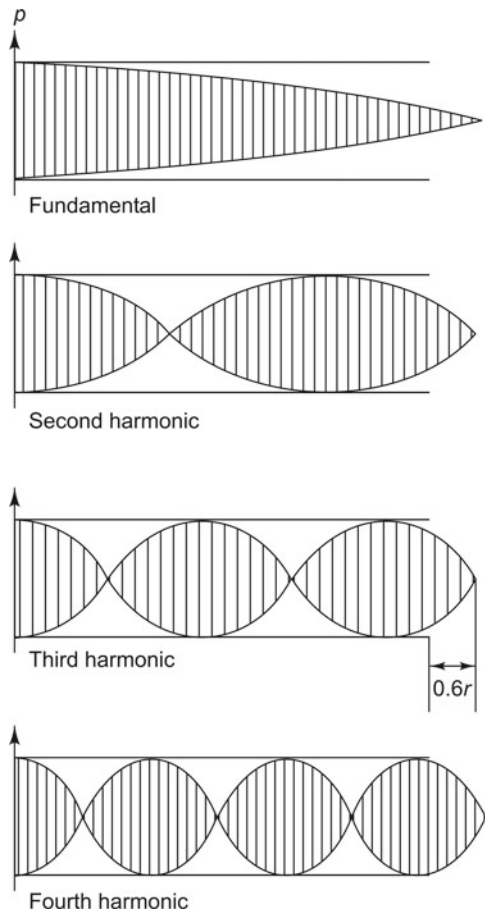
tions of the speed of sound and the length of the tube. The first natural frequency
is called the fundamental and the higher natural frequencies the harmonics. Due to
the nature of the boundary conditions imposed in this case, all harmonics are odd
integral multiples of the fundamental frequency.

4.3   Sound Propagation in 3-Dimensional (3-D) Space: The 3-D Wave Equation

INTERESTING FACTS: We all know that spherical waves emanate from a point source in an ideal isotropic medium such as air. It is also known that wave energy follows the inverse square law followed by conserved quantities in three-dimensional space. This translates to the $1/r$ decay law for the amplitude of the wave. The landmark Huygens–Fresnel principle explains that wave propagation happens as a superposition of spherical waves generated at each point along the wave front. Both Christiaan Huygens and Augustin-Jean Fresnel did a lot of work in the area of wave propagation which is applicable to both optics and acoustics. Huygens was a Dutch mathematician, physicist, and astronomer who made many extraordinary contributions in many diverse fields. He not only formulated the wave theory of light, but also discovered the pendulum clock, derived the centrifugal force for uniform circular motion, and found the true shape of the rings of Saturn and its moon, Titan. He also did research on free fall and experimentally proved the law of conservation of momentum. The French physicist Fresnel was a pioneer in optics who contributed to the wave theory of light, which had been studied extensively by English physicist Thomas Young. Fresnel was initially unaware of Young's experiments, but his experiments on producing interference fringes and diffraction gave him the conviction that the wave theory of light was correct. As a starting point for his mathematical theory of diffraction, Fresnel used Huygen's principle that every point on a wave front can be thought of as a secondary source of spherical wavelets. This revelation made him conclude that light was not a longitudinal wave as previously thought, but a transverse wave, a conclusion Young had reached independently.

Most practical acoustic problems involve propagation in 3-D; hence the plane wave
theory discussed in Sect. 4.2 needs to be extended to the 3-D case. In this section,
the general equation for propagation in 3-D is obtained in Cartesian coordinates as
an extension of the 1-D equation. For applications like spherical radiation studies,
spherical coordinates are more appropriate and for cylindrical radiation, cylindrical
coordinates. With rectangular coordinates, the equilibrium position of the particle
may be defined in terms of coordinates x, y, and z. The particle displacements in the
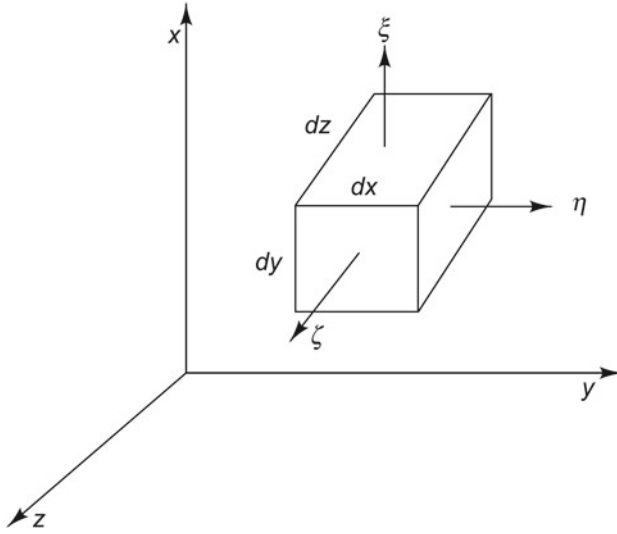
Fig. 4.5 Gas element in 3-D

three directions are ξ, η and ζ and the corresponding particle velocity components are $\partial\xi/\partial t$, $\partial\eta/\partial t$ and $\partial\zeta/\partial t$, respectively. We start off with a gas element in 3-D as shown in Fig. 4.5.

Since the acceleration in each direction is proportional to the force in that direction alone, the motion in each direction may be written as

$$\frac{\partial p}{\partial x} = -\rho_0 \frac{\partial^2 \xi}{\partial t^2}; \quad \frac{\partial p}{\partial y} = -\rho_0 \frac{\partial^2 \eta}{\partial t^2}; \quad \frac{\partial p}{\partial z} = -\rho_0 \frac{\partial^2 \zeta}{\partial t^2} \tag{4.35}$$

The three equations may be differentiated, each with respect to its own coordinate and then added to get

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = -\rho_0 \frac{\partial^2}{\partial t^2} \left(\frac{\partial \xi}{\partial x} + \frac{\partial \eta}{\partial y} + \frac{\partial \zeta}{\partial z} \right) \tag{4.36}$$

Equation (4.7) developed for the 1-D case remains unaltered. If the dimensions of the element are assumed to increase due to the passage of the sound wave, then the increased volume is

$$V + dV = \left[\left(1 + \frac{\partial \xi}{\partial x} \right) \left(1 + \frac{\partial \eta}{\partial y} \right) \left(1 + \frac{\partial \zeta}{\partial z} \right) \right] dx\,dy\,dz \tag{4.37}$$

where $V_0 = dx\,dy\,dz$ is the original volume.

$$\frac{dV}{V_0} = \left(\frac{\partial \xi}{\partial x} + \frac{\partial \eta}{\partial y} + \frac{\partial \zeta}{\partial z} \right) \tag{4.38}$$

Therefore,

$$p = -\gamma P_0 \left(\frac{\partial \xi}{\partial x} + \frac{\partial \eta}{\partial y} + \frac{\partial \zeta}{\partial z} \right) \tag{4.39}$$

Substituting the above equation into Eq. (4.36) gives

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = \frac{1}{c_0^2} \frac{\partial^2 p}{\partial t^2} \tag{4.40}$$

Or

$$\nabla^2 p = \frac{1}{c_0^2} \frac{\partial^2 p}{\partial t^2} \tag{4.41a}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{4.41b}$$

This is very similar to the 1-D wave equation. The particle velocity in terms of the particle pressure may be obtained as

$$u = \left(\frac{\partial \xi}{\partial t} + \frac{\partial \eta}{\partial t} + \frac{\partial \zeta}{\partial t} \right) = -\frac{1}{\rho_0} \int \left(\frac{\partial p}{\partial x} + \frac{\partial p}{\partial y} + \frac{\partial p}{\partial z} \right) dt \tag{4.42}$$

Or

$$u = -\frac{1}{\rho_0} \int \mathrm{grad}(p) dt \tag{4.43a}$$

where

$$\mathrm{grad} = \frac{\partial}{\partial r} \tag{4.43b}$$

In the case of spherical waves, it may be shown that

$$\left(\frac{\partial^2 p}{\partial r^2} + \frac{2}{r} \frac{\partial p}{\partial r} \right) = \frac{1}{c_0^2} \frac{\partial^2 p}{\partial t^2} \tag{4.44}$$

Or

$$\left(\frac{\partial^2 (rp)}{\partial r^2} \right) = \frac{1}{c_0^2} \frac{\partial^2 (rp)}{\partial t^2} \tag{4.45}$$

This equation represents an omni-directional wave radiating out from the source. It is a function of time t and radial distance r from the source and its solution may be written as

$$p(r, t) = \frac{g(c_0 t - r)}{r} + \frac{h(c_0 t + r)}{r} \tag{4.46}$$

Here the first term represents the wave radiating outward from the source; it gets attenuated as a function of $1/r$. The second term represents the backward wave travelling towards the source. For spherical waves which are harmonic, *p(r,t)* in the most general complex form can be written as

$$p(r,t) = \frac{P_+}{r} e^{i(\omega t - kr)} + \frac{P_-}{r} e^{i(\omega t + kr)} \tag{4.47}$$

where P_+ and P_- represent the peak amplitudes of the forward and backward travelling waves, respectively. On the lines of Eq. 4.23 for longitudinal waves in a column of gas, we have the relation between acoustic pressures and particle velocity for the outward travelling spherical wave.

$$u(r,t) = -\frac{1}{\rho_0} \int \frac{\partial p(r,t)}{\partial r} dt \tag{4.48}$$

Substituting for the term corresponding to the forward travelling wave into the above equation and noting that $\rho_0 \omega = \rho_0 c_0 k$, we get

$$u(r,t) = \frac{p(r,t)}{\rho_0 c_0} \left(1 + \frac{1}{ikr} \right) \tag{4.49}$$

There are two regions of interest with regard to the equation relating $u(r, t)$ and $p(r, t)$. They are the regions described by: (i) $kr \ll 1$ and (ii) $kr \gg 1$ as discussed in Sect. 4.4.7.

4.4 Some Important Acoustic Quantities and Relations

Is it not ironical that the French scientist Joseph Sauveur (1653–1716) who coined the term "*acoustique*", derived from an ancient Greek word meaning "able to be heard" was born deaf, did not talk until he was seven years old and had lifelong difficulties talking? He started his life as a teacher of mathematics, got interested in hydraulics, later military fortifications and then got a job with the Engineering Corps. After this he changed jobs to become a teacher at the College Royal in Paris. Sauveur became interested in acoustics. It is astonishing that he had neither a voice nor hearing, yet could think only of music. For all his acoustical experiments he had an assistant who listened to the sounds for him. He conducted detailed studies of the relationship between frequency and pitch, vibration waves in stretched strings, tuning pitch, harmonics, ranges of voices, and musical instruments. He has given the world a number of acoustic terms used to this day. He found that science as understood

in his times could not address the fundamental problems of music. He drew inspiration from Descartes, Mersenne, and Huygens especially, and differentiated between sound and music, recognizing that the former obeys mechanical and mathematical principles and that the latter is highly subjective.

In this section quantities of importance like velocity of sound, acoustic impedance, intensity, power, and radiation fields of a sound source are described. Sound pressure, intensity, and power levels are defined and the relationships between them are obtained. The additive effects of sources and the radiation fields of a source are also described.

4.4.1 Velocity of Sound

For small disturbances, the velocity of sound in a gas is

$$c_0 = \left(\frac{\gamma P_0}{\rho_0} \right)^{1/2} \tag{4.50a}$$

where γ is the ratio of specific heats of the gas, P_0 is the mean pressure and ρ_0 is the mean density. γ appears in the equation because the thermodynamic process involved is adiabatic due to the rapid fluctuations of pressure. P_0/ρ_0 is almost constant for all gases and the velocity of sound is independent of pressure.

Substituting $P_0/\rho_0 = RT/M$ where R is the gas constant, M is the molecular weight and T is the absolute temperature in Kelvin (K), we get

$$c_0 = \left(\frac{\gamma RT}{M} \right)^{1/2} \tag{4.50b}$$

The velocity of sound is thus seen to be proportional to the square root of the absolute temperature. For propagation in liquids

$$c_0 = \left(\frac{B_a}{\rho_0} \right)^{1/2} \tag{4.51}$$

where B_a is the adiabatic bulk modulus of elasticity and ρ_0 is the density. For propagation in solids

$$c_0 = \left(\frac{E}{\rho_0} \right)^{1/2} \tag{4.52}$$

where E is the modulus of elasticity and ρ_0 is the density.

4.4.2 Characteristic Impedance and Specific Acoustic Impedance

When a 1-D wave strikes a flat surface, the impedance it faces is called the specific acoustic impedance z, which is defined as the ratio of acoustic pressure to particle velocity.

$$z = \frac{p}{u} = r_a + i x_a \tag{4.53}$$

where r_a is called the specific acoustic resistance and x_a is called the specific acoustic reactance. In the general case of plane wave propagation in a medium, the forward and backward travelling waves are present; hence p and u in the above equation will have contributions from both terms and will be a function of position. For a wave impinging on a rigid wall, the particle velocity normal to the wall is zero, or the specific acoustic impedance of that component of the wave perpendicular to the wall is infinite. For standing waves or for diverging waves, specific acoustic impedance is a complex quantity. However, in the case of plane wave propagation in an infinite medium (plane progressive waves), we have only the forward travelling wave and the specific acoustic impedance is a real constant of magnitude $\rho_0 c_0$ as shown below.

$$z = \frac{p}{u} = \pm \rho_0 c_0 \tag{4.54}$$

z is a function purely of the medium and has greater significance as a characteristic property of the medium than either ρ_0 or c_0 alone. It is therefore called characteristic impedance and has the unit kg/m^{2}s or MKS Rayl. The velocity c_0 and the wavenumber k are in general, complex quantities, the imaginary parts of which contribute to the resulting absorption in the medium as the wave propagates. If absorption is small, the velocity and wavenumber are wholly real and the characteristic impedance is also wholly real. Table 4.4 shows the velocity of sound in some common materials that we come across and their characteristic impedances.

Table 4.4 Speed of sound in different media

Material	Speed (m/s)	Characteristic impedance $\rho_0 c_0$ (MKS rayls)
Air (at STP)	335	428
Lead	1128	23.2×10^6
Concrete	3109	8.0×10^6
Fresh water	1385	1.48×10^6
Steel	4925	47×10^6
Soft wood	3417	1×10^6
Glass	4771	12.9×10^6

4.4.3 Energy Density and Intensity

Acoustic pressure is the most meaningful quantity since our ears and acoustic transducers sense this scalar quantity. Besides, it is SPL which is used for evaluation of the harmfulness and annoyance of noise sources. But sound energy being a much more fundamental quantity, it becomes imperative to derive a relationship between the two. Energy density D is defined as the energy per unit volume of the fluid and is a function of both space and time and consists of two parts: (i) kinetic energy of the particles and (ii) potential energy of the gas. Knowing the particle velocity, u, the kinetic energy is obtained as

$$D_{KE} = \frac{1}{2}\rho_0 u^2 \tag{4.55}$$

When gas of volume V_0 undergoes an expansion dV, the potential energy it gains can be expressed as

$$D_{PE} = \frac{1}{V_0} \int p \, dV \tag{4.56}$$

It is to be remembered that in the course of sound wave propagation, the gas undergoes consecutive adiabatic compressions and rarefactions. Differentiating the adiabatic law $PV^\gamma = K$, we get $\frac{dV}{V_0} = -\frac{1}{\gamma}\frac{dp}{P_0}$ as described in Eq. (4.7). Substituting this into Eq. (4.56), we get the expression for potential energy as

$$D_{PE} = \frac{1}{\gamma P_0} \int p \, dp = \frac{1}{2}\frac{p^2}{\gamma P_0} = \frac{1}{2}\frac{p^2}{\rho_0 c_0^2} \tag{4.57}$$

Hence, the total energy density is

$$D(x, t) = \frac{1}{2}\left(\rho_0 u^2 + \frac{p^2}{\rho_0 c_0^2}\right) \tag{4.58}$$

The sound intensity I may be defined as the rate at which work is done on a conducting medium by an advancing sound wave and is the rate of flow of sound energy through unit area of the gas normal to the intensity vector; its unit is W/m². This quantity is again a function of space and time and it is a vector since the direction of flow of energy is defined (away from the source). The main use of sound intensity is for location of noise sources. Intensity can be thought of as the rate at which the acoustic pressure does work and is the product of the pressure and the component of particle velocity which is in phase with it.

$$I(x, t) = pu \tag{4.59}$$

The instantaneous pressure of a plane progressive wave for the forward travelling component may be written in the simplest form as

$$p = a_1 \cos(\omega t - kx) \tag{4.60}$$

The instantaneous particle velocity for the forward travelling component is given by

$$u = \frac{p}{\rho_0 c_0} \tag{4.61}$$

It may be noted that particle pressure and velocity are in phase. The energy density at any instant of time is got by substituting Eqs. (4.60) and (4.61) into (4.58).

$$D(x, t) = \frac{1}{2}\left(\frac{p^2}{\rho_0 c_0^2} + \frac{p^2}{\rho_0 c_0^2}\right) = \frac{a_1^2}{\rho_0 c_0^2}\cos^2(\omega t - kx) \tag{4.62}$$

From the instantaneous value of energy density described by the above equation, the average value may be got by averaging over one period, i.e., integrating over one period and dividing by that period. The energy density will be the same at all points in space for a plane progressive wave. Hence

$$D(x, t) = \frac{a_1^2}{\rho_0 c_0^2}\frac{1}{T}\int_0^T \cos^2(\omega t - kx)\mathrm{d}t = \frac{a_1^2}{2\rho_0 c_0^2} \tag{4.63}$$

Writing the above equation in terms of the RMS pressure rather than in terms of the peak value, we get

$$p_{\mathrm{rms}}^2 = \frac{1}{T}\int_0^T a_1^2 \cos^2(\omega t - kx)\mathrm{d}t = \frac{a_1^2}{2} \tag{4.64}$$

The average energy density in terms of the RMS pressure is

$$D = \frac{p_{\mathrm{rms}}^2}{\rho_0 c_0^2} \tag{4.65}$$

The intensity is got by substituting for pressure and particle velocity in Eq. (4.59) as

$$I(x, t) = \frac{a_1^2}{\rho_0 c_0}\cos^2(\omega t - kx) \tag{4.66}$$

On averaging this over one period, we get

$$I = \frac{a_1^2}{2\rho_0 c_0} = \frac{p_{\mathrm{rms}}^2}{\rho_0 c_0} \tag{4.67}$$

Comparing Eqs. (4.65) and (4.67) it is seen that

$$I = Dc_0 \tag{4.68}$$

4.4.4 Sound Power

Acoustic pressure and intensity depend on the distance from the source, especially for 3-D or spherical waves. For a point source, the sound pressure drops to half its value when the distance from the source is doubled and this corresponds to a drop in SPL of 6 dB. Hence, while specifying these quantities, the distance from the source must be mentioned. Besides, these quantities also depend on the reflective properties of the spaces surrounding the source. Hence, it is desirable to have a quantity that describes the acoustical characteristics of a source and which can be used to get intensity or pressure in different acoustical environments and at different distances from the source. Sound power happens to be such a quantity and is used for the noise rating of machines; its value does not depend on the distance or location around a source.

Sound power W is defined as the total sound energy (Joules) emitted by a source per unit time and is measured in Watts. It is obtained by integrating the sound intensity over an imaginary surface area S surrounding a source. If $\bar{I}$ represents the intensity vector, then the sound power radiated by the source is

$$W = \int_S \bar{I} \cdot d\bar{S} \tag{4.69}$$

The dot product is used since the area over which intensity is integrated is perpendicular to the flow of acoustical energy. For a sound source producing uniform spherical waves (or radiating equally in all directions), a spherical surface is most convenient for integration, but sometimes other surfaces are chosen, based on the circumstances for the particular case considered. The acoustic power for a source radiating spherically can be found knowing the intensity I at distance r from the source. Thus

$$W = \int_S \frac{p_{rms}^2(r)dS}{\rho_0 c_0} = \int_S \frac{p_{rms}^2(r)4\pi r^2}{\rho_0 c_0} = 4\pi r^2 I \tag{4.70}$$

For a surface composed of discrete irregular subareas, the acoustic power is obtained is

$$W = \sum_{i=1}^{n} I_i S_i \tag{4.71}$$

where I_i is the intensity of the ith segment and S_i is the surface area perpendicular to the energy flow through the ith segment. Sound power can hence be computed from measured values of sound pressure or sound intensity levels knowing the area over which the measurements were made.

4.4.5 Levels

In view of the fact that sound pressures, intensities and powers encompass a large number of orders of magnitude, these quantities are expressed in terms of levels in dB.

4.4.5.1 Sound Intensity Level

The range of acoustic intensities that the human ear is sensitive to is 10^{-12} W/m^2, corresponding to the threshold of audibility to 10^2 W/m^2, or a range of 14 decades. The intensity level L_I is defined as

$$L_I = 10 \log_{10} \left(\frac{I}{I_{\text{ref}}} \right) \text{dB} \tag{4.72a}$$

Here

$$I_{\text{ref}} = \frac{p_{\text{ref}}^2}{(\rho_0 c_0)_{\text{ref}}} \tag{4.72b}$$

$I_{\text{ref}} = 10^{-12}$ W/m^2 and $(\rho_0 c_0)_{\text{ref}} = 400$ MKS Rayls.

4.4.5.2 Sound Power Level

The sound power level, L_w may be defined as follows:

$$L_W = 10 \log_{10} \left(\frac{W}{W_{\text{ref}}} \right) \tag{4.73a}$$

Here the power, W, is measured in Watts with W_{ref} the internationally agreed reference power being 10^{-12} W.

$$W_{\text{ref}} = I_{\text{ref}} S_{\text{ref}} \tag{4.73b}$$

where $S_{\text{ref}} = 1\,\text{m}^2$. Table 4.5 gives the sound power and power levels of some common sources.

4.4.5.3 Relation Between Sound Pressure, Intensity and Power Levels

Among the three levels mentioned, it is only SPL which can be measured easily; sound intensity and power levels can be calculated from it. The relation between L_p, L_I and L_W can be obtained starting with Eq. (4.72) for sound intensity.

$$\frac{I}{I_{\text{ref}}} = \frac{p_{\text{rms}}^2}{p_{\text{ref}}^2} \cdot \frac{(\rho_0 c_0)_{\text{ref}}}{\rho_0 c_0} \tag{4.74}$$

Table 4.5 Sound power and sound power level of some sound sources

Sound source	Sound power (W)	Sound power level (dB ref 10^{-12} W)
Saturn V rocket	10^8	200
Turbojet engine	10^5	170
Siren	1000	150
Turboprop aircraft at take-off	100	140
Heavy truck engine	100	140
Machine gun	10	130
Heavy thunder	1	120
Trumpet	0.3	115
Chain saw	0.1	110
Accelerating motorcycle	0.1	110
Helicopter	0.01	100
Car at highway speed	0.01	100
Large diesel vehicle	0.001	90
Loud speech	0.001	90
Loud alarm clock	0.0001	80
Usual talking	10^{-5}	70
Quiet vacuum cleaner	10^{-5}	70
Hair dryer	10^{-6}	60
Radio or TV	10^{-7}	50
Low voice	10^{-8}	40
Quiet conversation	10^{-9}	30
Whisper of one person	10^{-10}	20
Human breath of one person	10^{-11}	10
Auditory threshold	10^{-12}	0

Taking logarithm of both sides of the above equation

$$10 \log_{10} \left(\frac{I}{I_{\text{ref}}} \right) = 10 \log_{10} \left(\frac{p_{\text{rms}}^2}{p_{\text{ref}}^2} \right) + 10 \log_{10} \left[\frac{(\rho_0 c_0)_{\text{ref}}}{\rho_0 c_0} \right] \tag{4.75}$$

This can be simplified to

$$L_I = L_p - 10 \log_{10} \left[\frac{\rho_0 c_0}{(\rho_0 c_0)_{\text{ref}}} \right] \tag{4.76}$$

$(\rho_0 c_0)$ for air at 20 °C and 50% Relative Humidity and at a standard atmospheric pressure of 1.013×10^5 Pa is 415 MKS Rayls and $(\rho_0 c_0)_{\text{ref}} = 400$ Rayls. Therefore,

$$L_I = L_p - 0.16 \tag{4.77}$$

For the atmospheric pressure and temperature conditions stated above, L_I and L_p are almost the same, but not for other conditions. The equation for sound power level may be derived as shown below.

$$\frac{W}{W_{ref}} = \frac{p_{rms}^2 4\pi r^2}{\rho_0 c_0} \frac{(\rho_0 c_0)_{ref}}{p_{ref}^2 4\pi r_{ref}^2} = \frac{p_{rms}^2}{p_{ref}^2} \frac{r^2}{r_{ref}^2} \frac{(\rho_0 c_0)_{ref}}{\rho_0 c_0} \tag{4.78}$$

Taking logarithm on both sides,

$$10 \log_{10}\left(\frac{W}{W_{ref}}\right) = 10 \log_{10}\left(\frac{p_{rms}^2}{p_{ref}^2}\right) + 10 \log_{10}\left(\frac{r^2}{r_{ref}^2}\right) - 10 \log_{10}\left[\frac{\rho_0 c_0}{(\rho_0 c_0)_{ref}}\right] \tag{4.79}$$

With $S_{ref} = 1 m^2$, $r_{ref} = 0.280$ m. Hence, for standard atmospheric temperature and pressure,

$$L_W = L_p + 20 \log_{10}\left(\frac{r}{r_{ref}}\right) - 0.16 \tag{4.80}$$

4.4.6 Additive Effects of Sound

Often, the total SPL due to sounds from many sources has to be evaluated. In general, the sounds are of different frequencies with random phases between them and the sources are then said to be incoherent or uncorrelated. An example of uncorrelated sources is the sound from two typewriters operated by different people. However, when the additive effect of sounds of the same frequency is to be found, the phase between them must be considered in the calculation. The sources are said to be fully correlated if they are producing the same waveform as a function of time, i.e., they have the same frequency and phase at a particular point. The cumulative effect can then be found by simply adding the pressures. Sometimes what happens is that all the waves arrive in phase, adding together to produce double the pressure at some points, while at other points they cancel each other out to produce zero pressure. If the sources are coherent, in phase and less than one wavelength apart, the sound pressures may be added to get the total sound pressure from which the total SPL may be got. An example of correlated sources is two stereo speakers which are placed within one wavelength of each other and fed with the same signal. The total sound pressure due to n correlated sources producing SPLs L_{p1}, L_{p2}, L_{p3}, ..., L_{pn} is obtained as follows:

$$L_{p,total} = 20 \log_{10}(10^{L_{p1}/20} + 10^{L_{p2}/20} + 10^{L_{p3}/20} + \cdots + 10^{L_{pn}/20}) \tag{4.81}$$

To find the overall effect due to multiple sound sources which are uncorrelated, the individual sound intensities (or p_{rms}^2 values) have to be added to get the total sound intensity, from which the total sound intensity level may be found. Thus, if two correlated sound sources produce 60 dB each at a listener's location, then $L_{p,total} = 60$ dB + 60 dB = 66 dB. If two uncorrelated sound sources produce 60 dB each, then

$L_{I,\text{total}} = 60$ dB $+ 60$ dB $= 63$ dB. But since L_I is almost equal to L_P for normal atmospheric conditions, we can say that $L_{p,\text{total}} = 63$ dB in the latter case. Hence, the combination of two random noise sources with both sources being identical, results in an increase of 3 dB over the SPL of one source. If the two independent SPLs are different, the combined level will exceed the higher of the two levels by less than 3 dB. When the difference between the two SPLs exceeds 10 dB, the contribution of the less noisy source to the overall noise level is negligible. The total sound pressure due to n uncorrelated sources with sound intensity levels $L_{p1}, L_{p2}, L_{p3}, \ldots, L_{pn}$ is obtained as follows:

$$L_{I,\text{total}} = 10 \log_{10}(10^{L_{I1}/10} + 10^{L_{I2}/10} + 10^{L_{I3}/10} + \cdots + 10^{L_{In}/10}) \qquad (4.82)$$

Sometimes it is necessary to subtract one noise level from another; as when background noise has to be subtracted from the combination of background and machine noise to obtain the SPL produced by a machine alone. The procedure to be followed is similar to the addition of levels.

4.4.7 Radiation Fields of a Sound Source

Considering the simplest form of a sound source, the energy emitted will propagate as spherical waves in all directions. The radiation fields of an acoustic source are the near field, far field, free field, reverberant field, and diffuse field. Figure 4.6a clearly shows the free field and reverberant field, while Fig. 4.6b shows the variation in SPL as a function of distance from the source or in different fields. The salient features of these fields are described below.

4.4.7.1 Near Field

The near field is the region very close to the source where the SPL may vary significantly with a small change in distance from the source. The extent of the near field depends on the frequency and radiation characteristics of the source, as well as its characteristic dimensions. For a spherical sound source, the near field extends to about 2 wavelengths of the lowest frequency (largest wavelength) emitted by the source; for an irregularly shaped and rather large source with a definite radiation pattern (which does not radiate uniformly in all directions), the near field extends from about two to five times the largest characteristic dimension of the source. Sound pressure measurements should be avoided in this region.

The intensity and sound pressure at a given point, at a distance r from the source are related as shown in Eq. (4.49). In this equation, the region in which $kr \ll 1$ is called the acoustic near field. In this region
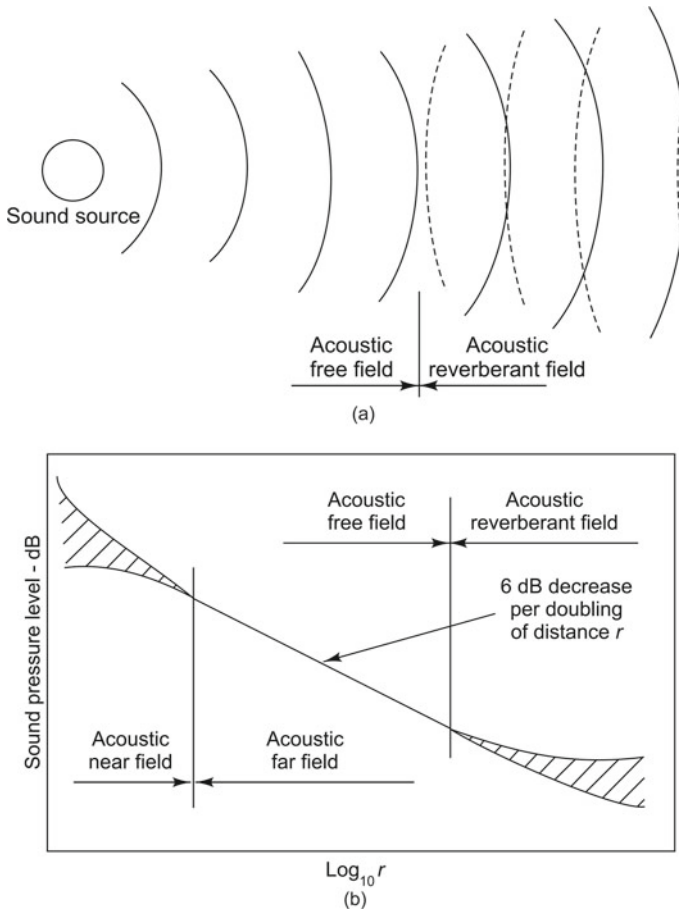
Fig. 4.6 Radiation fields of a source: **a** free field and reverberant field, **b** variation in SPL with field

$$u(r, t) = -i \frac{p(r, t)}{\rho_0 \omega r} \tag{4.83}$$

The region corresponding to $kr \gg 1$ is called the acoustic far field. For 1-D spherical waves, $kr = 10$ is often used as a demarcation between the near field and the far field. Since $\lambda = c_0/f$ where f is the frequency in Hz, we have $k = 2\pi/\lambda$. Hence, the distance to the demarcation between the near and far fields corresponding to $kr = 10$ is $r = 1.6\lambda$. In the near field as can be seen from Eq. (4.83) the particle pressure and velocity are 90° out of phase. Besides, in this region as seen in Fig. 4.6b, the sound field does not decrease by 6 dB for doubling of distance from the source (as it does in the far field). The sound pressure and particle velocity must be in phase for sound power to be radiated.

4.4.7.2 Far Field

The region far away from the source in called the far field. The far field commences where the near field ends and extends to infinity. The transition from near to far field is gradual and the demarcation is assumed to be $kr = 10$. The equation relating particle pressure and velocity in the far field is as shown below.

$$u(r, t) = \frac{p(r, t)}{\rho_0 c_0} = \frac{p(r, t)}{z} \tag{4.84}$$

Here acoustic pressure and particle velocity are in phase as seen from the above equation and this phase relationship is required for wave propagation. The far field comprises the free field and the reverberant field. In the free field, the sound pressure is inversely proportional to the distance r away from the source and therefore the SPL will decay at the rate of 6 dB per octave or 6 dB for every doubling of distance from the source.

4.4.7.3 Free Field/Direct Field

A free field is constituted by a plane wave which propagates in one defined direction. The free field is a region in space where only the directly radiated sound waves from the sound source are present and where there are no reflected waves. Such a field exists when a sound source is located far away from reflecting surfaces or when nearby surfaces are highly absorbent, thus preventing reflected waves. A small sound source (point source) may create a satisfactory plane wave at the measurement location, provided the location is sufficiently far away. The measurement site should be located at a distance at least five to ten times the largest dimension of the source and of the object which is to be placed in the field. It has already been mentioned that in free field conditions, the SPL decreases by 6 dB for doubling of distance from the source. Ideal free fields are difficult, if not impossible, to realize in practice. However, free fields which may be used for instrument verification and for calibration may be created in the laboratory in specially fabricated anechoic chambers, or outdoors away from reflecting surfaces. The anechoic chamber is a room with such highly absorbent walls that all the sound energy striking the walls is absorbed.

Directivity: The SPL at a point due to radiation from a source will depend on the presence of a reflecting surface near a source which will affect not only the radiated sound, but also the apparent directional properties of the source. In the direct field, the equation relating SPL and sound power level must be modified to account for the presence of such reflecting surfaces. This is done by considering a directivity factor, Q which can also characterize the directional sound radiation properties of a source. We have seen that when a spherical sound source of constant power output (not affected by reflecting surfaces) radiates into 3-D space or whole space

$$W = 4\pi r^2 I \tag{4.85}$$

Table 4.6 Directivity factors and indices

Space Q	Pictorial representation	Q	DI (dB)
Whole		1	0
1/2		2	3
1/4		4	6
1/8		8	9

where I is the constant intensity over a spherical surface of radius r. If a sound source is placed on a hard floor, the radiation will be over a hemisphere as the sound source is reflected from the plane. The source is then said to radiate sound into half space and the expression for sound power is

$$W = 2\pi r^2 I \tag{4.86}$$

It is also possible that the source is in a corner where a wall and floor or a wall and ceiling meet. With two such reflecting planes, the radiation will be over 1/4 of a sphere, or into quarter space. The source may be in a corner where two walls and a floor or ceiling meet; with three reflecting planes, sound is radiated into eighth space or 1/8 of a sphere as shown in Table 4.6. A common expression for sound power, independent of the nature of radiation space may be obtained by defining directivity factor Q and a corresponding Directivity Index (DI) in dB and incorporated in the expression for sound power level. Table 4.6 shows the directivity indices for difference spaces. The radiated power in terms of directivity factor is

$$W = \frac{4\pi r^2 I}{Q} \tag{4.87}$$

Directivity index is defined as

$$DI = 10 \log_{10} Q \tag{4.88}$$

The sound power level of a spherical source shown in Eq. (4.80) may be modified to the form shown below to account for directivity index.

$$L_W = L_p + 20 \log_{10} \left(\frac{r}{r_{\text{ref}}} \right) - 0.16 - DI \tag{4.89}$$

4.4.7.4   Reverberant Field

The reverberant field of a source is defined as the region of the sound field from a source which has experienced at least one reflection from the walls, ceiling or floor of the room or enclosure containing the source. The field consists of both directly radiated and reflected sound waves. Such a field exists when there are many reflecting surfaces surrounding the source; the incident and reflected waves then get superimposed to produce a reverberant field. The SPL at a point in the reverberant field will depend on the number of reflections, as well as the absorption coefficient of the reflecting boundaries. In this field there is no 6 dB reduction with doubling of distance away from the source.

4.4.7.5   Diffuse Field

In a reverberant field if there are many reflected waves arriving simultaneously from all directions with equal probability and level, the resulting field is called diffuse field; such conditions are created in the laboratory in a reverberation chamber, a room in which all surfaces are hard and highly reflecting and which essentially contain no sound absorbing materials. Diffuse field spectra created in such rooms may differ from those of an ideal field due to resonances in the room and due to the sound absorption of air. In technical applications where diffuse fields are required, it may be worthwhile applying more than one sound source and mounting reflecting panels which are moved continuously in order to change the dominating room resonances. Sound fields closely resembling a diffuse field may be found in factories where many simultaneous sound sources exist or in buildings with hard walls, such as halls or churches.

4.4.7.6   Pressure Field

In a pressure field, the sound pressure has the same magnitude and phase at any position. Pressure fields may be found in enclosures or cavities which are small in dimensions compared to the wavelength of sound, making the pressure uniform in the enclosure. Microphone "pressure sensitivity" refers to this type of field. Pressure field is used in acoustic calibrators and in piston phones where an exact acoustic pressure field occurs in the couplers which are used for testing earphones or calibrating microphones by the reciprocity technique.

4.5 Sound Transmission from One Medium to Another with Normal Incidence

INTERESTING FACTS: One of the most interesting experiments conducted in Physics laboratories is the "bell-in-vacuum" experiment. We have all watched the air pumped out of a jar containing a ringing bell; as air is being pumped out, the sound of the bell becomes weaker and weaker till it becomes inaudible. In the seventeenth century, scientists believed that sound propagated by means of invisible particles originating at the sound source and moving through space to the receiver's ears. The concept of sound as a wave challenged this view. The German scholar Athanasius Kircher performed the first ever bell-in-vacuum experiment and described it in his book "*Musurgia Universalis*" in 1650. Even after pumping the air out of the jar, he observed that he could still hear the bell, leading him to conclude incorrectly that air was not required for the transmission of sound. The fact was, however, that Kircher's jar was not exactly a vacuum due to the inadequacy in his vacuum pump. By 1660, the Anglo-Irish scientist Robert Boyle had improved vacuum technology; so when he repeated the experiment, he observed the sound intensity reduce almost to zero when air was pumped out. He then concluded correctly that a medium like air is required for sound waves to get transmitted. Though this conclusion is correct, the explanation for the results of the experiment is misleading. Even with the present-day mechanical vacuum pumps, the amount of air remaining in a vacuum jar is more than enough to transmit a sound wave. It is only when the sound pressure is reduced to below 10 millitorrs (1.333 Pa), which is not achievable, that the sound becomes inaudible. Thus, the real reason why the sound becomes inaudible is because the sound transmission gets attenuated due to the impedance mismatch between the solid bell and the rarefied air, between the rarefied air and the solid bell-jar and between the bell-jar and the atmospheric air (the last mismatch remains constant). When the air pressure is low, some of the sound energy is absorbed by the air itself and is converted to heat. Thus the real reason for the sound to become inaudible is not the lack of a medium such as air, as is generally presented in textbooks.

When sound waves from a source strike a surface representing a change from one physical medium to another across an interface, a part of the sound energy is absorbed in the second medium and a part of it is reflected back into the first medium, the two quantities depending on the characteristic impedances of the two media. This is shown in Fig. 4.7. Let medium 1 have a characteristic impedance $\rho_1 c_1$ and medium 2, $\rho_2 c_2$. Let the wavenumbers in the two media be $k_1 = \omega/c_1$ and $k_2 = \omega/c_2$. It should be noted that the frequency of the wave in the two media will be the same.
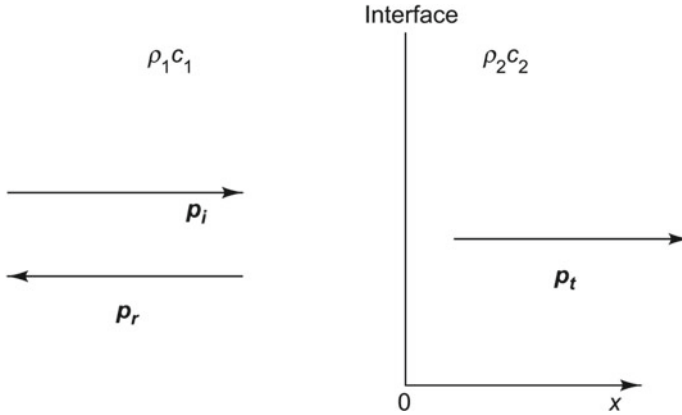
Fig. 4.7 Transmission with normal incidence

Let the incident wave be represented as

$$p_i = A_i e^{i(\omega t - k_1 x)} \qquad (4.90)$$

Here boldface is used to represent complex quantities as before.

The reflected and transmitted waves may be represented as

$$p_r = A_r e^{i(\omega t + k_1 x)} \qquad (4.91)$$

$$p_t = A_t e^{i(\omega t - k_2 x)} \qquad (4.92)$$

Two continuity conditions must be satisfied at the boundary separating the two media.

(i) **Continuity of pressure:** The total pressure in media 1 and 2 should be equal at $x = 0$.

$$(p_i + p_r)_{x=0} = (p_t)_{x=0} \qquad (4.93a)$$

$$A_i + A_r = A_t \qquad (4.93b)$$

Here A_i, A_r and A_t are, in general, complex to allow for phase differences between the three waves.

(ii) **Continuity of particle velocity:** It is assumed that the two media are not physically separated at $x = 0$. This implies that the total normal particle velocity in medium 1 is equal to that of medium 2 at $x = 0$. Thus

$$(u_i + u_r)_{x=0} = (u_t)_{x=0} \qquad (4.94)$$

The particle velocities corresponding to the incident, reflected and transmitted waves may be shown to be

$$u_i = \frac{p_i}{\rho_1 c_1}$$

$$u_r = -\frac{p_r}{\rho_1 c_1} \tag{4.95}$$

$$u_t = \frac{p_t}{\rho_2 c_2}$$

It follows from Eq. (4.94) that

$$\frac{p_i - p_r}{\rho_1 c_1} = \frac{p_t}{\rho_2 c_2} \tag{4.96}$$

Eliminating A_t from Eqs. (4.94) and (4.96), we obtain the relation between the reflected and incident waves as

$$\frac{A_r}{A_i} = \frac{\rho_2 c_2 - \rho_1 c_1}{\rho_2 c_2 + \rho_1 c_1} \tag{4.97}$$

Similarly A_r may be eliminated to give the relation between the transmitted and incident waves as

$$\frac{A_t}{A_i} = \frac{2\rho_2 c_2}{\rho_2 c_2 + \rho_1 c_1} \tag{4.98}$$

Alternately, the two boundary conditions can be combined into a single condition that the specific acoustic impedance is continuous at the boundary and is equal to z_s. Then

$$\left(\frac{p_i + p_r}{u_i + u_r}\right)_{x=0} = z_s = \left(\frac{p_t}{u_t}\right)_{x=0} \tag{4.99a}$$

Or

$$\frac{A_i + A_r}{A_i - A_r}\rho_1 c_1 = z_s = z_2 \tag{4.99b}$$

Equation (4.99b) may be rearranged to the form

$$\frac{A_r}{A_i} = \frac{z_s - \rho_1 c_1}{z_s + \rho_1 c_1} \tag{4.100}$$

The sound power reflection coefficient may be defined as

$$\alpha_r = \frac{I_r}{I_i} = \frac{|A_r|^2/2\rho_1 c_1}{|A_i|^2/2\rho_1 c_1} = \left|\frac{A_r}{A_i}\right|^2 = \left(\frac{\rho_2 c_2 - \rho_1 c_1}{\rho_2 c_2 + \rho_1 c_1}\right)^2 \tag{4.101}$$

The sound power transmission coefficient may be defined as

$$\alpha_t = \frac{I_t}{I_i} = \frac{|A_t|^2/2\rho_2 c_2}{|A_i|^2/2\rho_1 c_1} = \frac{\rho_1 c_1}{\rho_2 c_2}\left|\frac{A_2}{A_1}\right|^2 = \frac{4\rho_1 c_1 \rho_2 c_2}{(\rho_2 c_2 + \rho_1 c_1)^2} \tag{4.102}$$

Fig. 4.8 Standing wave patterns: **a** $\rho_1 c_1 \ll \rho_2 c_2$, **b** $\rho_1 c_1 \gg \rho_2 c_2$
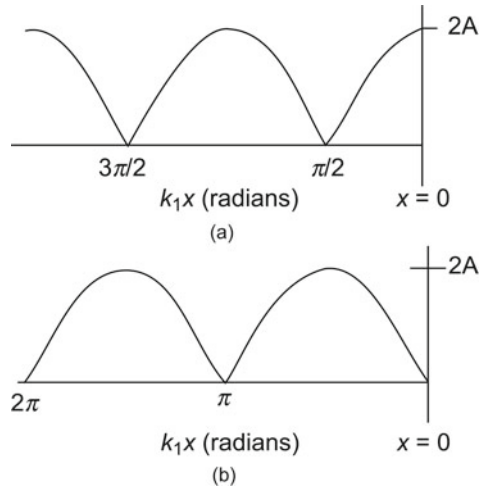
Equations (4.101) and (4.102) show that sound reflection and transmission phenomena are independent of frequency. We can consider three specific cases:

(i) $\rho_1 c_1 \ll \rho_2 c_2$

From Eqs. (4.101) and (4.102), it can be seen that $\alpha_t = 4\rho_1 c_1 / \rho_2 c_2$ or $\alpha_t \approx 0$ and $\alpha_r \approx 1$, implying that almost all the incident sound energy is reflected back into the originating medium, with very little energy getting transmitted to the second medium. When sound is transmitted from air to water which have characteristic impedances of 415 and 1.48×10^6 MKS rayls, respectively, $\alpha_t = 1.12 \times 10^{-3}$ or only 0.112% of the incident energy is transmitted into water. Figure 4.8a shows the standing wave patterns for such a case.

(ii) $\rho_1 c_1 = \rho_2 c_2$

With this condition $\alpha_t = 1$ and $\alpha_r = 0$ implying that all the incident sound energy is transmitted into the second medium.

(iii) $\rho_1 c_1 \gg \rho_2 c_2$

With this condition, we get $\alpha_t = 4\rho_2 c_2 / \rho_1 c_1$ or $\alpha_t \approx 0$ and $\alpha_r \approx 1$ implying that all the incident energy is reflected back into the first medium. Figure 4.8b depicts the standing wave patterns for this case.

4.5.1   Sound Transmission Though a Solid Barrier with Normal Incidence

The equations for this case may be derived on similar lines as for transmission from one medium into another. A quantity of great practical significance stemming from

$\rho_1 c_1$ $\qquad\qquad\qquad$ $\rho_2 c_2$ $\qquad\qquad\qquad$ $\rho_3 c_3$

p_i $\qquad\qquad\qquad$ p_1 $\qquad\qquad\qquad$ p_t

p_r $\qquad\qquad\qquad$ p_2

$x = 0$ $\qquad\qquad\qquad$ $x = L$
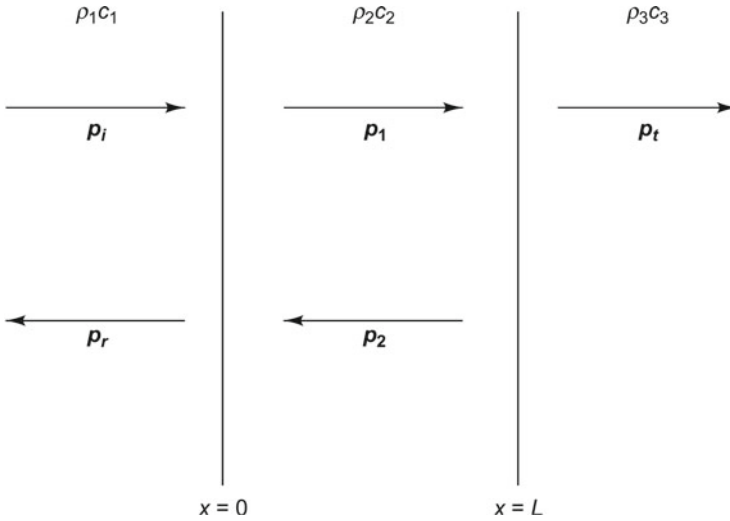
Fig. 4.9 Sound transmission through a barrier

a derivation of this kind is the sound power transmission coefficient of a partition or panel. Such panels find use in noise control applications.

The transmission coefficient is obtained by assuming that we have air as the first and third media in Fig. 4.9 with the barrier as the second medium. The barrier is assumed to be rigid and to have a finite thickness.

The sound power transmission coefficient τ, for such a barrier is defined as the ratio of the transmitted acoustic power to the incident power. Since $W = IS$, S being the surface area common to the two media and since $I_i = \frac{p_{rms(i)}^2}{\rho_0 c_0}$ and $I_t = \frac{p_{rms}^2(t)}{\rho_0 c_0}$, the sound power transmission coefficient is given as

$$\tau = \left| \frac{A_t}{A_i} \right|^2 \tag{4.103}$$

and the transmission loss (TL) of a sound wave through the partition is defined as

$$TL = 10 \log_{10} \left(\frac{1}{\tau} \right) \tag{4.104}$$

4.6   Acoustics of Enclosed Spaces

DID YOU KNOW that Wallace Clement Sabine, an American physicist, is considered the "Father of architectural acoustics"? In fact, he was the architectural acoustician of Boston's Symphony Hall, which is considered as one of the best concert halls in the world for its acoustics. In 1895, when he was a young assistant professor (without any formal training in acoustics) at the Physics Department of Harvard University, he was assigned the task of acoustically improving the Fogg Lecture Hall, a part of the recently constructed Fogg Art Museum. Sabine and his assistants spent many nights moving materials to different lecture halls and measuring sound waves; they had nothing but an organ to produce sound and a stop watch to measure its decay. Sabine came up with a relationship between the surface area of absorption materials placed on walls, seats, ceilings and floors and the volume of the lecture hall. Sabine's formula is used by architects and engineers even today for the design of concert halls to achieve the best reverberation time. The unit Sabine has been named in his honour; one imperial Sabine equals one square foot of 100% absorbing material, and one metric Sabine equals one square metre of 100% absorbing material. The vast field of architectural acoustics has evolved from Sabine's nineteenth-century work.

The behaviour of sound in enclosed spaces needs to be studied since this throws light on the design of auditoria, class rooms, churches, opera halls, and other such spaces. If a sound source is placed in any such enclosure, a part of the energy radiated will get absorbed and a part of the energy will get reflected from the boundary surfaces of the room. A hard, compact, smooth surface will reflect most of the acoustic energy, while a porous, soft surface will absorb most of it. If the surfaces of the room are reflective in nature, the sound produced by a source will undergo multiple reflections at the boundaries. The larger the reflection coefficients, the larger will be the contribution of the reflected sound to the total sound in the enclosed space. The build-up of sound energy will continue even after the source has been turned off. The SPL in such a reverberant sound field will depend on the acoustic power radiated, the size of the room, and the acoustic absorption properties of the boundaries. If the surfaces are highly absorptive in nature, the contribution of the reflected sound becomes less and almost a free field condition will exist where most of the sound is the direct sound. In general, most work spaces may be considered to be semi-reverberant. Before attempting to obtain the expression for the SPL in an enclosure, it is worthwhile looking at the resonant behaviour of rooms.

4.6.1   Acoustic Field in a Small Rectangular Room

The 3-D wave equation expressed in terms of sound pressure is given by

$$c_0^2 \nabla^2 p(x, y, z, t) - \frac{\partial^2 p(x, y, z, t)}{\partial t^2} = 0 \qquad (4.105)$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

If a harmonic solution is assumed, the sound pressure can be written as

$$p(x, y, z, t) = P(x, y, z)e^{i\omega t} \qquad (4.106)$$

Substituting this equation in Eq. (4.105), we get

$$\nabla^2 P(x, y, z) + k^2 P(x, y, z) = 0 \qquad (4.107)$$

where $k = \frac{\omega}{c_0}$ as has been defined earlier.

This is the 3-D Helmholtz equation and may be solved by a separation of variables as shown.

$$P(x, y, z) = P(x)P(y)P(z) \qquad (4.108)$$

Substituting in Eq. (4.107) and dividing by $P(x)P(y)P(z)$, we get

$$\frac{1}{P(x)} \frac{d^2 P(x)}{dx^2} + \frac{1}{P(y)} \frac{d^2 P(y)}{dy^2} + \frac{1}{P(z)} \frac{d^2 P(z)}{dz^2} + k^2 = 0 \qquad (4.109a)$$

On separating the variables, we get

$$\frac{d^2 P(x)}{dx^2} + k_x^2 P(x) = 0 \qquad (4.109b)$$

$$\frac{d^2 P(y)}{dy^2} + k_y^2 P(y) = 0 \qquad (4.109c)$$

$$\frac{d^2 P(z)}{dz^2} + k_z^2 P(z) = 0 \qquad (4.109d)$$

where

$$k_x^2 + k_y^2 + k_z^2 = k^2 \qquad (4.110)$$

The solutions to the above equations are

$$P(x) = A_x \cos k_x x + B_x \sin k_x x \qquad (4.111a)$$
$$P(y) = A_y \cos k_y y + B_y \sin k_y y \qquad (4.111b)$$
$$P(z) = A_z \cos k_z z + B_z \sin k_z z \qquad (4.111c)$$

The constants A_x, A_y, A_z, B_x, B_y and B_z are obtained by applying the boundary conditions associated with the rectangular enclosure with sides a, b, and c. The particle velocity at the rigid walls must be zero. Stating this in terms of pressure from the momentum Eq. (4.4), the boundary conditions are

$$\left.\frac{\mathrm{d}P(x)}{\mathrm{d}x}\right|_{x=0,a} = 0, \quad \left.\frac{\mathrm{d}P(y)}{\mathrm{d}y}\right|_{y=0,b} = 0, \quad \left.\frac{\mathrm{d}P(z)}{\mathrm{d}z}\right|_{z=0,c} = 0 \qquad (4.112)$$

Taking the derivative of Eq. (4.111a) with respect to x gives

$$\frac{\mathrm{d}P(x)}{\mathrm{d}x} = -k_x A_x \sin k_x x + k_x A_x \cos k_x x \qquad (4.113)$$

Applying the boundary conditions at $x = 0$, $B_x = 0$. Similarly applying at $x = a$,

$$k_x A_x \sin k_x a = 0 \quad \text{or} \quad k_x = \frac{l\pi}{a}, l = 1, 2, 3, \dots \qquad (4.114)$$

In a similar fashion $B_y = B_z = 0$ and

$$k_y = \frac{m\pi}{b}, \quad m = 1, 2, 3, \dots \qquad (4.115)$$

$$k_z = \frac{n\pi}{c}, \quad n = 1, 2, 3, \dots \qquad (4.116)$$

Substituting Eq. (4.111) into Eq. (4.108), we get

$$P(x, y, z) = A_{lmn} \cos_{k_{x(l)}} x \cdot \cos k_{y(m)} y . \cos k_{z(n)} z \qquad (4.117)$$

Substituting the values of k_x, k_y, k_z into Eq. (4.110), we get

$$k_{lmn} = \sqrt{\left(\frac{l\pi}{a}\right)^2 + \left(\frac{m\pi}{b}\right)^2 + \left(\frac{n\pi}{c}\right)^2} \qquad (4.118)$$

Equations (4.117) and (4.118) describe the mode shapes and the wavenumbers of the natural modes that can exist in a rectangular room of dimensions a, b and c. The associated natural frequencies are given by

$$\omega_{lmn} = c_0\pi\sqrt{\left(\frac{l}{a}\right)^2 + \left(\frac{m}{b}\right)^2 + \left(\frac{n}{c}\right)^2} \qquad (4.119)$$

The total sound pressure solution in a rectangular enclosure is

$$p(x, y, z, t) = \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} A_{lmn} \cos k_{x(l)} x . \cos k_{y(m)} y . \cos k_{z(n)} z e^{i\omega_{lmn} t} \quad (4.120)$$

It is essentially the largest dimension that determines the lowest natural frequency.

4.6.2 Sound Pressure Level in a Large Enclosure

For obtaining the SPL in a large room, the following aspects are to be considered: the shape of the enclosed area, the sound absorption and reflection characteristics of the different surfaces of the enclosure and in the case of large areas, the effects of sound absorption by air. Figure 4.10 shows a room with a sound source "S" and a receiver "R". There are two types of sound fields associated with the sound source: the direct field consisting of sound waves travelling directly from the source to the receiver and the reverberant field, consisting of sound waves that reach the receiver after at least one reflection. The reverberant sound field results in a considerable increase in the SPL at the receiver. The steady-state sound level is reached when the total sound energy added to the reverberant field is equal to the energy absorbed by the reflecting surfaces.

The energy density in a large enclosure due to a 3-D spherical wave travelling away from a sound source is composed of the direct field and the reverberant field.
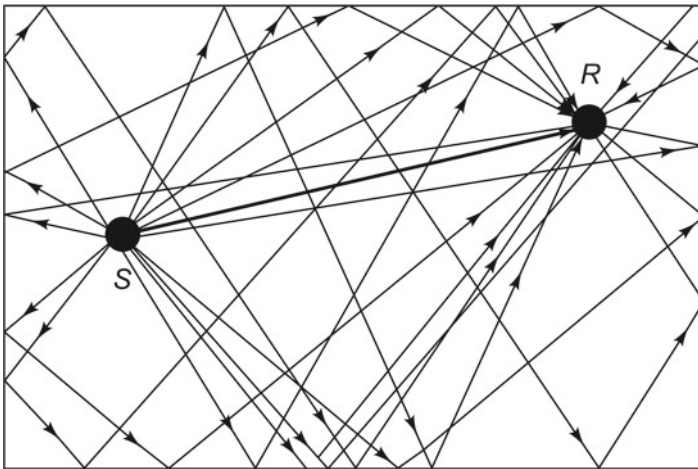
Fig. 4.10 Direct and reverberant sound fields
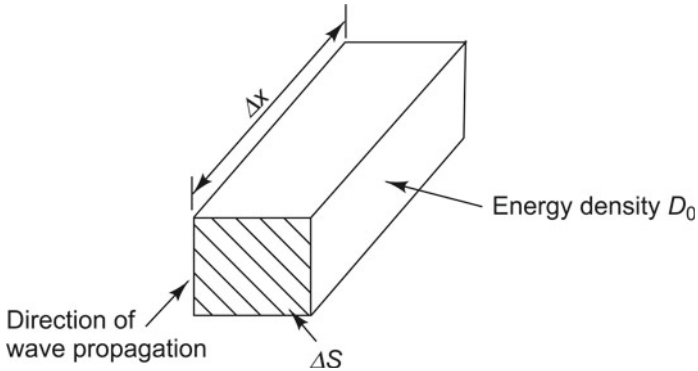
Fig. 4.11 Energy density in the direct field (D_0)

4.6.2.1 Energy Density in the Direct Field (D_0)

The energy density D_0 due to the direct field can be computed from the energy flowing through a small volume ΔV as shown in Fig. 4.11 during a small duration Δt. The total energy E in volume $\Delta V = \Delta x \Delta S$ is

$$E = D_0 \Delta S \Delta x \tag{4.121}$$

The acoustic intensity I is given by

$$I = \frac{W}{\Delta S} \tag{4.122}$$

where W = sound power directly radiated by the sound source and ΔS = unit area.

The sound power may also be computed as the energy contained in volume ΔV divided by the time Δt taken for the sound wave to travel distance Δx. Hence, the power radiated from the source is

$$W = \frac{E}{\Delta t} \tag{4.123}$$

Substituting Eq. (4.121) and (4.123) into (4.122), we get

$$I = \frac{D_0 \Delta S \Delta x}{\Delta S \Delta t} = D_0 c_0 \tag{4.124}$$

where $\frac{\Delta x}{\Delta t} = c_0$, the speed of sound in air.

For a 1-D spherical wave propagation, for any space (full or half or one-fourth or one-eighth) with directivity factor Q

$$I = \frac{W Q}{4 \pi r^2} \tag{4.125}$$

Hence, the expression for energy density in the direct field becomes

$$D_0 = \frac{WQ}{4\pi r^2 c_0} \tag{4.126}$$

4.6.2.2 Energy Associated with Reverberant Field (D_R)

To determine the energy density D_R due to the reverberant field of the enclosure, the sound power associated with the reverberant field that is incident upon a small surface area ΔS is found out. The small energy dE contained in a thin shell-like volume dV at a distance r away from ΔS and centred at ΔS (Fig. 4.12) is

$$dE = D_R dV = D_R dr dS \tag{4.127}$$

where

$$dV = dS dr \text{ and } dS = r^2 \sin\theta d\theta d\psi \tag{4.128}$$

The sound power dW radiated towards the surface ΔS from the volume dV is

$$dW = \frac{dE}{dt} = D_R dS \frac{dr}{dt} = D_R c_0 dS \tag{4.129}$$

Acoustic power is assumed to radiate spherically from dV towards ΔS as depicted in Fig. 4.12. The acoustic intensity dI associated with the sound power dW at ΔS is

$$dI = \frac{D_R c_0 dS}{4\pi r^2} \tag{4.130}$$
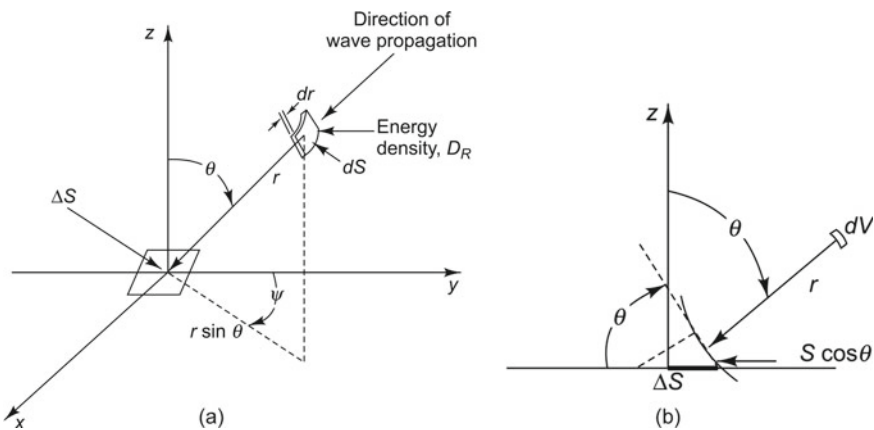
Fig. 4.12 Energy density D_R: **a** sound power due to reverberant field, **b** sound power radiated towards ΔS from dV

The small power dW_i incident upon the surface ΔS is

$$dW_i = dI\,\Delta S \cos\theta = \frac{D_R c_0 \Delta S \cos\theta\,dS}{4\pi r^2} \tag{4.131}$$

where $\Delta S \cos\theta$ is the area element of the sphere of radius r projected on to ΔS normal to r.

Due to multiple reflections it may be assumed that a diffuse field exists. Hence, the total sound power incident on the surface element ΔS is obtained by integrating dW_i over the surface of the hemisphere. Therefore,

$$W_i = \frac{D_R c_0 \Delta S}{4\pi r^2} \int_0^{2\pi} \int_0^{\pi/2} r^2 \sin\theta \cos\theta\,d\theta\,d\psi = \frac{D_R c_0 \Delta S}{4} \tag{4.132}$$

The sound power W_a absorbed by the surface element ΔS is

$$W_a = W_i \alpha = \frac{D_R c_0 \Delta S \alpha}{4} \tag{4.133}$$

The total sound power absorbed by the surface which is equal to the sound power lost from the reverberant field is obtained by adding the power over all surface areas as shown below.

$$W_a = \frac{D_R c_0 S\overline{\alpha}}{4} \tag{4.134}$$

where $\overline{\alpha}$ is the average absorption coefficient of all such absorbing surfaces S_1, S_2, $S_3, \ldots, S_n$ with sound absorption coefficients $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n$.

$$\overline{\alpha} = \frac{S_1\alpha_1 + S_2\alpha_2 + \cdots + S_n\alpha_n}{S_1 + S_2 + \cdots + S_n} = \frac{S_1\alpha_1 + S_2\alpha_2 + \cdots + S_n\alpha_n}{S} \tag{4.135}$$

From the law of conservation of energy, the sound power supplied to the reverberant field should equal the sound power absorbed by the surfaces of the enclosure under steady-state conditions. Besides, the sound power supplied to the reverberant field should equal the remaining power after the sound waves undergo their first reflections. Hence,

$$W(1 - \overline{\alpha}) = \frac{D_R c_0 S\overline{\alpha}}{4} \tag{4.136a}$$

The above equation may be rearranged to give

$$D_R = \frac{4W(1 - \overline{\alpha})}{c_0 S\overline{\alpha}} = \frac{4W}{c_0 R} \tag{4.136b}$$

where R is the room constant defined as

$$R = \frac{S\overline{\alpha}}{1 - \overline{\alpha}} \tag{4.137}$$

The total energy density D in an enclosure is the sum of the energy densities from the direct and reverberant fields, or

$$D = D_0 + D_R = \frac{W}{c_0}\left(\frac{4}{R} + \frac{Q}{4\pi r^2}\right) \tag{4.138}$$

The total energy density in the enclosure can be expressed as

$$D = \frac{p_{\text{rms}}^2}{\rho_0 c_0^2} \tag{4.139}$$

Substituting Eq. (4.139) into Eq. (4.138), we get

$$p_{\text{rms}}^2 = W\rho_0 c_0 \left(\frac{4}{R} + \frac{Q}{4\pi r^2}\right) \tag{4.140}$$

Both sides of the above equation may be divided by p_{ref}^2 and the numerator and denominator of the right side may be multiplied by $(\rho_0 c_0)_{\text{ref}}$ and S_{ref} to give

$$\frac{p_{\text{rms}}^2}{p_{\text{ref}}^2} = \frac{W}{(p_{\text{ref}}^2 S_{\text{ref}})/(\rho_0 c_0)_{\text{ref}}} \frac{\rho_0 c_0}{(\rho_0 c_0)_{\text{ref}}} S_{\text{ref}}\left(\frac{4}{R} + \frac{Q}{4\pi r^2}\right) \tag{4.141}$$

For normal temperature and atmospheric conditions

$$\frac{\rho_0 c_0}{(\rho_0 c_0)_{\text{ref}}} = 1 \quad \text{and} \quad W_{\text{ref}} = \frac{p_{\text{ref}}^2 S_{\text{ref}}}{(\rho_0 c_0)_{\text{ref}}} \tag{4.142}$$

Hence, Eq. (4.141) can be written as

$$\frac{p_{\text{rms}}^2}{p_{\text{ref}}^2} = \frac{W}{W_{\text{ref}}} S_{\text{ref}}\left(\frac{4}{R} + \frac{Q}{4\pi r^2}\right) \tag{4.143}$$

Taking logarithm of both sides of the above equation we get

$$L_p = L_W + 10\log_{10}\left(\frac{4}{R} + \frac{Q}{4\pi r^2}\right) + 10\log_{10} S_{\text{ref}} \tag{4.144}$$

If R is in m^2 and r is in m, $S_{\text{ref}} = 1\,\text{m}^2$ and Eq. (4.144) becomes

$$L_p = L_W + 10\log_{10}\left(\frac{4}{R} + \frac{Q}{4\pi r^2}\right) \tag{4.145}$$

It is to be noted here that in the direct field $\frac{4}{R} \ll \frac{Q}{4\pi r^2}$ and in the reverberant field, $\frac{4}{R} \gg \frac{Q}{4\pi r^2}$. Accordingly, the appropriate terms alone may be used in Eq. (4.145).

4.6.3 Decay of a Sound Field in an Irregularly Shaped Enclosure

The theory explaining reverberation time and the acoustical behaviour of rooms was developed by Sabine. A matter of great concern to both architects and musicians in room acoustics is reverberation time which is defined as the time taken for SPL to fall by 60 dB from the steady-state value, the moment the source is switched off. If an impulsive sound is generated in a room with reflecting boundaries, there are multiple reflections at the boundaries, after which there is more or less a uniform sound field. This field then decays due to absorption of sound energy by the materials of the room surfaces. The rate of decay depends on the absorptive properties of the reflecting surfaces and the distances between them. Figure 4.13 shows the build-up of sound in an enclosure after a sound source has been turned on at time $t = -0.6$ s. Also shown is the decay of sound for three different cases: (a) very little absorption, (b) medium absorption, and (c) large absorption. In this section an expression is derived for the time taken for the sound field in an enclosure to decay by 60 dB after a sound source is turned off.

Let the energy density in the room initially be D. After each reflection, the energy density D decreases by a factor $(1 - \overline{\alpha})$ as shown in Fig. 4.14. Let d be the average distance or the mean free path travelled by a sound wave between reflections and let t' be the average time between reflections or the mean free time calculated as

$$t' = \frac{d}{c_0} \tag{4.146}$$

Hence, the expressions for the energy density become

$$D(t') = D(1 - \overline{\alpha}) \quad \text{after the first reflection}$$

$$D(2t') = D(1 - \overline{\alpha})^2 \quad \text{after the second reflection}$$

$$D(nt') = D(1 - \overline{\alpha})^n \quad \text{after the } n \text{ th reflection} \tag{4.147}$$
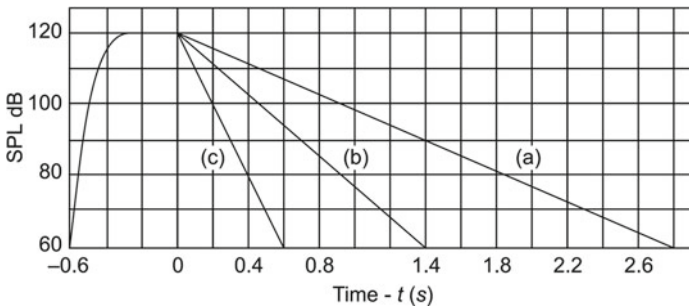
Fig. 4.13 Build up and decay of sound in an enclosure

Fig. 4.14 Effects of several
reflections of a sound wave
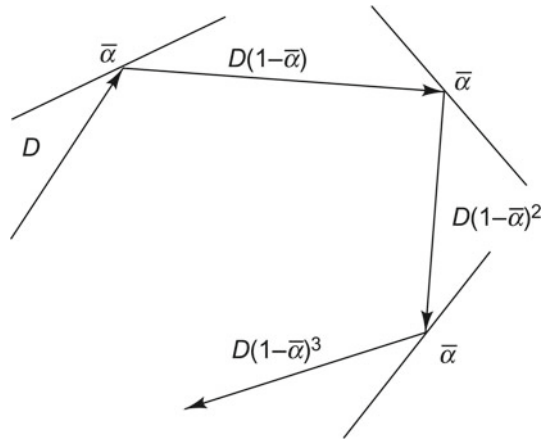
Studies conducted, both analytically and experimentally, on several enclosures of different shapes, have shown that d is given by the expression

$$d = \frac{4V}{S} \tag{4.148}$$

where V is the volume of the room and S is the total surface area of the room. Hence, Eq. (4.146) can be rewritten as

$$t' = \frac{4V}{c_0 S} \tag{4.149}$$

The time t taken by the wave to undergo n reflections is

$$t = nt' \tag{4.150}$$

Substituting the above equation in the expression for t' in Eq. (4.149), we get

$$n = \frac{c_0 St}{4V} \tag{4.151}$$

Substituting for t from Eq. (4.150) and for n from Eq. (4.151) into Eq. (4.147), we get

$$D(t) = D(1 - \bar{\alpha})^{(c_0 St/4V)} \tag{4.152}$$

Plugging in the value of $D = \frac{p_{rms}^2}{\rho_0 c_0^2}$ into the above equation, we get

$$\frac{p_{rms}^2(t)}{p_{rms}^2(t = 0)} = (1 - \bar{\alpha})^{(c_0 St/4V)} \tag{4.153}$$

Taking logarithm on both sides, we get

$$L_{p(t)} - L_{p(t=0)} = 10 \log_{10} \left\{ (1 - \overline{\alpha})^{c_0 St/4V} \right\} \tag{4.154}$$

Reverberation time T is defined as the time it takes for the SPL to decrease 60 dB after the source is switched off and is obtained as

$$-60 = \frac{c_0 ST}{V} 2.5 \log_{10}(1 - \overline{\alpha}) \tag{4.155}$$

The above equation may be written in terms of natural logarithm as

$$60 = 1.086 \frac{c_0 T}{V} \{ -S \ln(1 - \overline{\alpha}) \} \tag{4.156}$$

Room absorption a is defined as

$$a = S \ln(1 - \overline{\alpha})^{-1} \tag{4.157}$$

Substituting for a in Eq. (4.156), the reverberation time can be written as

$$T = \frac{60V}{1.086 c_0 a} \tag{4.158}$$

This is called the Eyring reverberation time formula. If c_0 is in terms of m/s, V is in m^3 and a is in m^2, T in s can be written as

$$T = 0.161 \frac{V}{a} \tag{4.159}$$

If the average sound absorption coefficient $\overline{\alpha}$ is less than 0.1,

$$\ln(1 - \overline{\alpha})^{-1} \approx \overline{\alpha} \tag{4.160}$$

and room absorption a can be written as

$$a \approx S\overline{\alpha} \tag{4.161}$$

The reverberation time can then be written as

$$T = 0.161 \frac{V}{S\overline{\alpha}} \tag{4.162}$$

The transducers and equipment for measuring many of the acoustic quantities described in this chapter are discussed in Chap. 5 and common acoustic experiments in Chap. 11.

List of Symbols

Symbol	Meaning	SI unit
a	room absorption	m^2
a, b, c	room dimensions	m
c_0	velocity of sound	m/s
a_1, a_2, a_3, a_4	amplitude of pressure or particle velocity wave	
$\boldsymbol{A}, \boldsymbol{B}$	complex amplitude of pressure or particle velocity wave	
$\boldsymbol{A_i}$	complex amplitude of incident pressure wave	Pa
$\boldsymbol{A_r}$	complex amplitude of reflected pressure wave	Pa
$\boldsymbol{A_t}$	complex amplitude of transmitted pressure wave	Pa
$\boldsymbol{A_x}, \boldsymbol{A_y}, \boldsymbol{A_z}$	complex constants	
B_a	adiabatic bulk modulus of elasticity	Pa
$\boldsymbol{B_x}, \boldsymbol{B_y}, \boldsymbol{B_z}$	complex constants	
c_0	velocity of sound wave	m/s
d	average distance between reflections or mean free path	m
dE	incremental energy	J
dI	incremental intensity	W/m^2
$dS, \Delta S$	incremental area	m^2
dV	incremental volume	m^3
dW	incremental power	W
$D, D(x, t)$	energy density	J/m^3
DI	directivity index (dB)	
D_0	energy density due to direct field	J/m^3
D_R	energy density due to reverberant field	J/m^3
E	Young's modulus of elasticity	Pa
E	energy	J
f	frequency	Hz
F	force	N
$I, I(x, t)$	sound intensity	W/m^2
I_i	incident sound intensity	W/m^2
I_r	reflected sound intensity	W/m^2
I_{ref}	reference sound intensity	W/m^2
I_t	transmitted sound intensity	W/m^2
k, k_1, k_2	wavenumbers	m^{-1}
k_{lmn}	wavenumbers in room	m^{-1}
K	constant	
L	length of tube	m
L'	length of tube with end correction	m
L_p	sound pressure level (dB)	
L_I	sound intensity level (dB)	
L_W	sound power level (dB)	
M	molecular weight	kg/mol
p	particle pressure	Pa

Symbol	Meaning	SI unit
p_+, p_-	particle pressures of forward and backward travelling waves	Pa
p_i	complex incident pressure	Pa
p_r	complex reflected pressure	Pa
p_t	complex transmitted pressure	Pa
P_+, P_-	complex peak particle pressures of forward and backward travelling waves	Pa
p_{ref}	reference sound pressure	Pa
p_{rms}	root mean square value of pressure	Pa
$p(r, t)$	complex pressure wave	Pa
$p(x, y, z, t)$	3-dimensional pressure wave	Pa
p, dP	particle pressure, dynamic pressure	Pa
P	total pressure (static plus dynamic)	Pa
P_0	static pressure	Pa
Q	directivity factor	
r	radial distance from source	m
r_a	specific acoustic resistance	rayl or Pa s/m
r_{ref}	reference radial distance	m
R	universal gas constant	J mol^{-1} K^{-1}
R	room constant	m^2
S	total surface area of the room	m^2
S	cross-sectional area	m^2
$S_1, S_2, S_3, \ldots, S_n$	areas of absorbing surfaces	m^2
S_{ref}	reference area	m^2
t	time	s
t'	average time between reflections or mean free time	s
T	reverberation time	s
T	period of acoustic wave	s
T	absolute temperature	K
T	time duration	s
TL	transmission loss (dB)	
u	particle velocity	m/s
u_i	complex incident particle velocity	m/s
u_r	complex reflected particle velocity	m/s
u_t	complex transmitted particle velocity	m/s
u_+, u_-	particle velocities of forward and backward travelling waves	m/s
$u(r, t)$	complex particle velocity	Pa
V	volume of the room	m^3
V_0	undisturbed volume or original volume	m^3

Symbol	Meaning	SI unit
W	sound power	W
W_a	absorbed sound power	W
W_i	incident sound power	W
W_{ref}	reference sound power	W
x	original position	m
x_a	specific acoustic reactance	rayl or Pa s/m
z, z_1, z_2	specific acoustic impedance	rayl or Pa s/m
z_s	specific acoustic impedance at boundary	rayl or Pa s/m
$\alpha, \alpha_2, \alpha_3, \ldots, \alpha_n$	sound power absorption coefficients	
$\bar{\alpha}$	average sound power absorption coefficient	
α_r	sound power reflection coefficient	
α_t	sound power transmission coefficient	
γ	ratio of specific heats of gas	
Δx	small distance	m
Δt	small time	s
λ	wavelength	m
ξ, η, ζ	particle displacements in x, y and z directions	m
ρ_0	mean density of gas	kg/m^3
$\rho_0 c_0, \rho_1 c_1, \rho_2 c_2, \rho_3 c_3$	characteristic impedances	rayl or Pa s/m
$(\rho_0 c_0)_{\mathrm{ref}}$	reference characteristic impedance	rayl or Pa s/m
τ	sound power transmission coefficient	
ϕ	particle pressure or particle velocity	
ω	angular frequency	rad/s
ω_{lmn}	natural frequencies in room	rad/s

Abbreviations

1-D	1-dimension
3-D	3-dimensional
FFT	fast Fourier transform
IEC	International Electrotechnical Commission
RMS	root mean square
SPL	sound pressure level
TL	transmission loss

Questions

1. What does the study of acoustics involve?
2. What is the nature of sound waves?
3. What does psychoacoustics deal with?
4. What is the range of frequencies the human ear is sensitive to?
5. From the point of view of waveforms, how do you distinguish between sound and noise?
6. What are the harmful effects of noise?
7. Define the threshold of audibility and that of pain.
8. Differentiate between linear and logarithmic scales for the depiction of physical quantities. Where would you use them?
9. How do you define sound pressure level?
10. Why are sound pressure, intensity and power levels defined in dB?
11. What is the relationship between sound power and pressure ratios?
12. A washing machine manufacturer claims that his machine makes 3 dB less noise than his competitor's. What do you make of his claim?
13. How do you define sonic, ultrasonic, and infrasonic ranges?
14. What is the need to do spectrum analysis of sound?
15. Which analyzers, percentage bandwidth or constant bandwidth, are preferred for machinery diagnostics?
16. What are the assumptions made in the study of the propagation of sound waves in a column of gas?
17. What is the particle model of studying wave motion?
18. What is Hooke's law?
19. Why is the thermodynamic process involved in the longitudinal propagation of sound waves assumed to be adiabatic?
20. How are wavelength and frequency of a sound wave related?
21. What are the boundary conditions to be applied to the pressure in a tube closed at one end and open at the other end?
22. How are sound particle pressure and velocity related in forward and backward travelling waves?
23. What kind of waves emanate from a point source in an isotropic medium such as air?
24. What is the $1/r$ decay law for sound pressure?
25. How is the expression for velocity of propagation of sound in a liquid different from that in a gas?
26. How do you define energy density?
27. On what material properties does the velocity of propagation of sound in a solid depend?
28. What are the boundary conditions applicable to the pressure in a tube closed at both ends?
29. How do you define sound intensity level?
30. Why is sound pressure, in general, denoted by a complex quantity?

31. What are correlated sound sources? Give two examples.
32. Give two examples of uncorrelated sound sources.
33. Are characteristic impedance and specific acoustic impedance the same?
34. What is meant by wavenumber?
35. Is the velocity of sound in air higher or lower than that in steel? Why?
36. How is sound intensity helpful in the localization of a sound source?
37. What is meant by near field with respect to a sound source?
38. What decides the extent of the near field?
39. How do you define far field?
40. What is the reverberant field of a source?
41. How do you calculate the demarcation between the near and far fields?
42. What conditions are conducive for the generation of a free field?
43. What conditions are conducive for the generation of a reverberant field?
44. What is meant by diffuse field?
45. Why is sound power and not sound pressure or intensity used for the noise rating of sources?
46. Under what conditions are sound pressure level and sound intensity level almost the same?
47. When is specific acoustic impedance a real quantity?
48. To find the cumulative effect from multiple sources, when do you add sound pressures and when sound intensities?
49. If two correlated sound sources are producing a sound pressure level of 70 dB each at a listener's location, what is the total sound pressure level?
50. What is the energy density in the direct field?
51. How do you define reverberation time of an acoustic enclosure?
52. What is the room constant R?
53. What is one metric Sabine?
54. What is meant by sound power transmission coefficient?
55. What is directivity index?
56. What is meant by a quarter space?
57. What is the directivity index for one-eighth space?
58. How can you conclude from measurements if a given acoustic field corresponds to the far field?
59. What are the two continuity conditions that will hold good at the boundary between two media when sound propagates from one to the other?
60. If two uncorrelated sound sources produce 80 dB each at a particular location, what is the total SPL?
61. What is average energy density in terms of RMS pressure?
62. What do you mean by acoustic impedance mismatch between two media? What does it result in?
63. What is the general equation for propagation in 3-D space in Cartesian coordinates?
64. What is meant by the elastic behaviour of a gas?
65. What are 1/1 and 1/3 octave bands for sound analysis?

66. What properties should a good acoustic barrier have?
67. For choral music in churches, is it desirable to have a high or low reverberation time?
68. When a sound wave travels from one medium into another, when will standing waves be generated?
69. How do you define sound power reflection coefficient?
70. What are the properties of a good sound absorbing material?
71. What can you say about the particle pressure and particle velocity at a rigid wall?
72. Which dimension of a room determines its lowest natural frequency?
73. How do you compute the average sound absorption coefficient of a room?
74. What is the speed of sound in air?
75. What is the reference acoustic pressure for audio-acoustics?
76. What can you say about the acoustic pressures sensed as being heard in comparison to the mean atmospheric pressure?
77. In the expression for calculation of SPL, why is RMS pressure used?
78. Does sound power decrease with distance from the source?
79. How can you create a diffuse field?
80. Which will have a higher reverberation time, a room with parallel walls or one with non-parallel walls? Why?
81. What do you mean by objective and subjective quantities?
82. What are the objective quantities related to loudness, pitch and tonal quality?
83. What is the difference between the sensitivity of a microphone with a flat wide frequency response and that of the human ear?

Fill in the Blanks

84. The total sound pressure due to uncorrelated sources with sound intensity levels $L_{I1}, L_{I2}, L_{I3}, \ldots, L_{In}$ is _____.
85. When the difference in SPLs between two sources exceeds 10 dB, the contribution of the less noisy source to the overall SPL is _____.
86. It is desirable to have a reverberation time of the order of _____ seconds for a recording studio.
87. If the speech in a room sounds garbled, it implies that the reverberation time of the room is _____.
88. If the sound from a source is to be prevented from reaching a receiver, the possible solutions would be _____.
89. When a sound wave travels from a medium with a high characteristic impedance into one with a lower characteristic impedance, most of the energy gets _____.
90. If two independent SPLs are different at a listener's location, the combined level will exceed the higher of the two levels by less than _____ dB.
91. For small disturbances, the velocity of sound in a gas is a function of _____, _____ and _____.
92. The 3-D wave equation expressed in terms of sound pressure is given as _____.

93. For a point source, the sound pressure drops to half its value when the distance from the source is doubled and this corresponds to a drop in SPL of _____ dB.

94. The total sound pressure level measured from two sources of 85 and 87 dB is _____.

95. A two-fold increase in sound pressure amounts to _____ dB increase in SPL.

96. A decrease of SPL by _____ dB is perceived as half as loud.

Bibliography

1. Ambrose, J. E., & Ollswang, A. J. (1995). *Simplified design for building sound control (Parker Ambrose series of simplified design guides)*. New York: Wiley.
2. Ando, Y. (1985). *Concert hall acoustics*. Springer.
3. Barron, R. F. (2003). *Industrial noise control and acoustics*. Boca Raton, Florida: CRC Press.
4. Benade, A. H. (1990). *Fundamentals of musical acoustics: (Dover books on music)*. Dover Publication.
5. Beranek, L. L. (1988). *Noise and vibration control*. Poughkeepsie, New York: Institute of Noise Control Engineering.
6. Beranek, L. L., & Ver, I. L. (1992). *Noise and vibration control engineering: Principles and applications*. New York: Wiley-Interscience.
7. Beranek, L. L. (1993). *Acoustics*. Woodbridge, New York: Acoustical Society of America.
8. Beranek, L. L. (2003). *Concert halls and opera houses: Music, acoustics, and architecture*. New York: Springer.
9. Beranek, L., & Mellow, T. (2019). *Acoustics: Sound fields, transducers and vibration*. USA: Academic Press.
10. Bies, D. A., & Hansen, C. H. (2003). *Engineering noise control: Theory and practice*. New York, USA: Taylor & Francis.
11. Blackstock, D. T. (2000). *Fundamentals of physical acoustics*. New York: Wiley -Interscience.
12. Chakrabarti, P. K., & Chowdhury, S. (2010). *A textbook on waves and acoustics*. New Central Book Agency.
13. Cremer, L., & Müller, H. A. (translated by T. J. Schultz). (1982). *Principles and applications of room acoustics* (Vols. 1 and 2). Barking, Essex: Applied Science Publishers.
14. Crocker, M. J. (1997). *Encyclopedia of acoustics*. NewYork: Wiley-Interscience.
15. Crocker, M. J. (1998). *Handbook of acoustics*. New York: Wiley.
16. Crocker, M. J. (2007). *Handbook of noise and vibration control*. Hoboken, New Jersey: Wiley.
17. Day, B. F., Ford, R. D., & Lord, P. (1969). *Building acoustics*. Amsterdam, New York: Elsevier Applied Science.
18. Ermann, M. (2015). *Architectural acoustics illustrated*. Wiley.
19. Everest, F. A., & Pohlman, K. C. (2000). *Master handbook of acoustics*. New York: McGraw-Hill.
20. Fahy, F. J., & Walker, J. G. (1998). *Fundamentals of noise and vibration*. London: E & FN Spon.
21. Fahy, F. J. (2001). *Foundations of engineering acoustics*. London: Academic Press.
22. Fahy, F., & Thompson, D. (2015). *Fundamentals of sound and vibration*. CRC Press.
23. Filippi, P., Habault, D., Lefebvre, J. P., & Bergassoli, A. (1998). *Acoustics: Basic physics, theory and methods*. London: Academic Press.
24. Filippi, P. (1984). *Theoretical acoustics and numerical techniques (CISM international centre for mechanical sciences courses and lectures)*. New York: Springer.
25. Fraden, J. (2010). *Handbook of modern sensors*. Springer.

26. Ginsberg, J. H. (2018). *Acoustics-A textbook for engineers and physicists, volume I: Fundamentals*. Springer.
27. Ginsberg, J. H. (2018). *Acoustics-A textbook for engineers and physicists, volume II: Fundamentals*. Springer.
28. Greenspon, J. E. (2011). *Structural acoustics: Deterministic and random phenomena*. CRC Press.
29. Hall, D. E. (1993). *Basic acoustics*. Malabar, FL: Krieger Publishing.
30. Harris, C. (1991). *Handbook of acoustical measurements and noise control*. New York: The McGraw-Hill.
31. Hartmann, W. M. (2013). *Principles of musical acoustics*. New York: Springer.
32. Hemond, C. J. (1983). *Engineering acoustics and noise control*. Englewood Cliffs, New Jersey: Prentice Hall.
33. Kinsler, L. E., & Frey, A. R. (1982). *Fundamentals of acoustics*. New York: Wiley.
34. Kinsler, L. E., Frey, A. R., Coppens, A. B., & Sanders, J. V. (1999). *Fundamentals of acoustics*. New York: Wiley.
35. Kleiner, M. (2014). *Acoustics of small rooms*. CRC Press.
36. Kleppe, J. A. (1989). *Engineering applications of acoustics*. Boston, Massachusetts: Artech House Publishers.
37. Kuttruff, H. (2016). *Room acoustics*. CRC Press.
38. Lindsay, R. B. (1974). *Physical acoustics*. New York: Van Nostrand Reinhold Co.
39. Maekawa, Z., & Lord, P. (2007). *Environmental and architectural acoustics*. Taylor & Francis eBookstore.
40. Maher, R. C. (2020). *Principles of forensic audio analysis*. Spain: Springer.
41. Malecki, I. (1969). *Physical foundations of technical acoustics*. New York: Pergamon.
42. Mechel, F. P., & Munjal, M. L. (2004). *Formulas of acoustics*. Berlin, New York: Springer.
43. Migeot, J. L. (2016). Acoustics, Lulu.com.
44. Miles, R. N. (2020). *Physical approach to engineering acoustics*. Spain: Springer.
45. Morse, P. M., & Ingard, K. U. (1986). *Theoretical acoustics*. Princeton, New Jersey: Princeton University Press.
46. Möser, M., & Zimmermann, S. (2004). *Engineering acoustics: An introduction to noise control*. Berlin, New York: Springer.
47. Munjal, M. L. (2013). *Noise and vibration control (IISc lecture notes series book 3)* Kindle Edition, WSPC.
48. Newhouse, V. (2012). *Site and sound: The architecture and acoustics of new opera houses and concert halls*. New York: Monacelli Press.
49. Nilsson, A., & Bilong, L. (2015). *Vibro-Acoustics* (Vol. 1). Berlin, Heidelberg: Springer.
50. Peters, R. J., & Reeves, C. W. (1998). *The noise and acoustics handbook*. UK: Coxmoor Publishing Co.
51. Peters, R. J. (2011). *Acoustics and noise control*. Routledge.
52. Pierce, A. D. (1989). *Acoustics: An introduction to its physical principles and applications*. Woodbury, New York: Acoustical Society of America.
53. Rayleigh, J. W. S., & Lindsay, R. B. (1976). *The theory of sound* (Vols. 1 and 2). New York: Dover Publications.
54. Reynolds, D. D. (1981). *Engineering principles of acoustics: Noise and vibration*. Boston: Allyn & Bacon Inc.
55. Rossing, T. D., & Fletcher, N. H. (2004). *Principles of vibration and sound*. New York: Springer.
56. Seto, W. W. (1971). *Acoustics, Schaum's outline series*. New York: McGraw-Hill.
57. Skudrzyk, E. J. (1971). *Foundations of acoustics, basic mathematics & basic acoustics*. New York: Springer.
58. Smith, B. J., Peters, R. J., & Owen, S. (1996). *Acoustics and noise control*. London, New York: Longman.
59. Schmitz, T. L., Smith, K. S. (2012). *Mechanical vibrations*. Springer.
60. Sujatha, C. (2010). *Vibration and acoustics: Measurement and signal analysis*. Tata McGraw-Hill Education.

61. Subrahmanyam, N., & Lal, B. V. (1986). *A textbook of sound*. New Delhi: Vikas Publishing House Pvt. Ltd.
62. Temkin, S. (1981). *Elements of acoustics*. New York: Wiley.
63. Tohyama, M. (2018). *Sound in the time domain*. Springer.
64. Vasques, C. M. A., & Rodrigues, J. D. (2014). *Vibration and structural acoustics analysis: Current research and related technologies*. Springer.
65. Wilson, C. E. (1994). *Noise control: Measurement, analysis, and control of sound and vibration*. Krieger Publishing Company.
66. Xiang, N. (April 1, 2017). *Architectural acoustics handbook* (A Title in J. Ross Publishing's Acoustic). J. Ross Publishing.

Chapter 5
Equipment for Measurements in Acoustics

INTERESTING FACTS: For centuries, scientists thought that sound could not be captured. But today we can record it with the finest of details. How has the science of audio measurement evolved? In 1807, Thomas Young used a flexible diaphragm with a pin inserted into it to capture sound. By shouting into the diaphragm, he had the sound waves scratched by the pin onto a glass cylinder coated with lamp black. The French scientist Édouard-Léon Scott de Martinville took this idea forward to develop the 'phonautograph' in 1854. He concentrated the sound waves using the ears of decapitated dogs as receiving horns. He put a feather across the small end of the ear and when he shouted, the sharpened tip of the feather 'wrote' the sound waves on the lamp black on a cylinder. In 1882, Lord Rayleigh devised a setup consisting of a glass tube with a small reflective disk placed such that it could pivot along a diameter and rotate proportionally to the particle velocity of sound waves in the tube. One end of the tube was open but had a tissue across it such that random drafts would not upset the sound waves. When a beam of light was shone onto the vibrating disk, it was reflected back onto a graduated target and thus he could measure its rotation and the amplitude of the sound wave. In 1986, Thomas Edison and Emile Berliner invented the carbon button microphone within a week of each other. It was George Washington

Pierce, who first connected a carbon button microphone to the moving coil of a galvanometer and measured sound electrically. The electrical era in acoustic measurements really began in 1917, when Western Electric engineers combined four separate and unrelated inventions to create a reliable machine for practical sound measurements: they were (i) d'Arsonval's galvanometer, (ii) the 'condenser' microphone, invented by Edward C. Wente, (iii) the thermophone devised around the same time by Harold D. Arnold and Irving B. Crandall, and (iv) a vacuum tube amplifier. In 1917, AT&T produced a sound level meter (SLM), which was huge and cumbersome. Later came computerized measurements, paving the way for sound evolving from being merely a scientific theory to a quantity that can be measured accurately.

This chapter gives a general introduction to the various kinds of microphones and acoustic exciters, which are available in the market, as well as their characteristics and the selection criteria commonly used to identify the device most suitable for a particular application. It describes the principle of operation of the sound level meter, which is used extensively in acoustic measurements, along with 1/1 and 1/3 octave filters. It also gives the background information regarding the characteristics, specifications, and applications of various acoustic calibrators and calibration techniques. Finally, it describes anechoic and reverberation chambers, the two chambers used for many standard measurements.

5.1   Parameters to Be Considered in the Choice of Microphones

Microphones have a wide range of applications such as acoustic measurements, telecommunications, broadcasting, recording of music, consumer electronics, etc. For any specific microphone application, the acoustic engineer can choose from a number of different transduction principles. A microphone may sense the pressure, the pressure gradient or the particle velocity, which is then converted to an electrical signal. Most precision measurement microphones are pressure sensing condenser microphones, the reason being that they detect what the human ear detects, i.e., pressure. A constant electrical charge is used for conversion of the diaphragm displacement into an analogue electrical signal. Pressure sensing microphones can be used if particle velocity or sound intensity is required to be determined. The measuring microphones affect the sound field, especially at higher frequencies, because the wavelength and microphone dimensions tend towards the same order of magnitude. However, the influence depends on the type of sound field (pressure-field, free-field or diffuse-field). When selecting a microphone, it is important to first consider the measurement requirements and understand how they influence the performance of the microphone.

5.1.1 Technical Considerations

Technical considerations for good acoustic and electric performance are discussed below.

5.1.1.1 Microphone Type Depending on Sound Field

When making acoustic measurements, it is to be ensured that the microphone itself does not influence the sound field being measured and change the sound pressure so that it will not lead to a measurement error. The influence varies depending on the type of field, i.e., free-field or pressure-field or diffuse-field and on the type of microphone, its dimensions, as well as the diaphragm impedance (to a small extent). This influence may be small enough to be negligible in some cases, whereas in other cases, it may be unacceptable, amounting to several dB, and is often a function of frequency of sound. The pressure-sensing microphone can be designed for use in different types of sound fields, i.e., pressure-field, free-field and diffuse-field (equivalent to random-incidence response). Microphone types that have been designed for specific applications and that, therefore, correct for the influence of the field are commercially available. These microphones are optimized to have a flat frequency response called the optimized response in one of these sound fields. The differences between the responses are seen only at higher frequencies, differing by less than 0.1 dB below 1000 Hz. If such specifically designed microphones are not used, the user must evaluate, and if possible, correct for the influence.

Pressure-field microphone: The sound pressure in a pressure field has the same magnitude and phase at any location within the field. Pressure fields are found in enclosures or cavities, which are small in dimensions compared to the wavelength. Typical examples are couplers applied for testing of earphones or calibration of microphones. The frequency response of a pressure-field microphone is generally optimized to be as flat as possible, for uniform pressure on the outer surface of its diaphragm. For this, the diaphragm should be as stiff as the walls of the coupler within which it is applied. Since this is not practically possible, appropriate corrections may be made.

Free-field microphone: Free-field microphone types have a flat free-field frequency response characteristic to sound waves that are perpendicular to the diaphragm (at 0° incidence). When making outdoor measurements away from reflecting surfaces or indoors in an office with a lot of natural acoustic damping, a free-field microphone is the best choice. The basic difference between a free-field microphone and a normal pressure microphone is essentially in the design compensation of the frequency response of the former such that it can be directed at the noise source with minimal errors resulting from the introduction of the microphone in the sound field. The influence of the microphone body on the sound field may be corrected by post-

processing of measured sound spectra, but the compensation might also be built into the microphone itself. The compensation is made by introducing heavy damping for the diaphragm resonance, leading to a frequency response characteristic, which is essentially flat in an undisturbed free-field.

Diffuse-field microphone: Dedicated microphones for diffuse-field measurements are not very common, the reason being that the diffuse-field correction or the pressure change created by the microphone itself is so small that many pressure-field microphones also have good diffuse-field characteristics. For measurements in enclosed areas with hard reflective surfaces where reverberations are likely, pressure-field microphones modified for random-incidence measurements to be present are the best because the random-incidence response of a pressure-field microphone is much 'flatter' or more constant across the frequency range than that of a free-field microphone.

5.1.1.2 Technical Specifications of a Microphone

Frequency response: The microphone should have a wide frequency range and flat frequency response. Since microphones generally have a wide operational frequency range, frequency response should be considered in relation to other requirements such as the type of sound field.

Dimensions: The size (diameter) of the microphone is to be chosen according to the highest frequency range of interest and the dynamic range of sound pressure levels (SPLs) and usually reduces with higher frequency and SPL to be measured. It is also true that the larger microphones have higher sensitivity than the smaller ones.

Equivalent noise dB$_N$ (A): This is the level of sound pressure required to produce an output voltage corresponding to the inherent A-weighted noise voltage, measurable with 3% non-linear distortion and is dictated by the inherent noise of the microphone and pre-amplifier combination.

Dynamic range: This is the range from the lowest to the maximum SPL, the microphone can handle. For a wide linear dynamic range, low inherent noise and low distortion are required. The lower limit of the dynamic range is dictated by the inherent noise of the microphone and pre-amplifier combination. This is of concern for use in office environments where levels as low as 10–12 dBA may have to be measured. The upper limit of the dynamic range is dictated by the maximum SPL; e.g., in measurements of exhaust systems where high SPLs (140–150 dB) may be encountered. It is normally either the lower or the upper limit of dynamic range that is of interest depending on the application.

Sensitivity: This is described in terms of output mV per unit input Pa pressure fluctuation. Pressure sensitivity implies sensitivity to the actual pressure acting on the

microphone, whereas free-field sensitivity implies sensitivity to the pressure that existed in the sound field before insertion of the microphone, these two being equal at low frequencies.

Open-circuit sensitivity (S_o): This is defined as the pressure-field sensitivity, valid with an idealized pre-amplifier which does not load the microphone. It is the ratio of open-circuit voltage to pressure and is determined at a frequency of 250 or 1000 Hz and is used for calibration and monitoring of the microphone.

Loaded sensitivity (S_c): When the microphone cartridge is connected to the pre-amplifier, its input voltage is attenuated by the pre-amplifier input capacitance. This effect is valid over a wide frequency range. The loaded sensitivity is described in terms of the open-circuit sensitivity as

$$S_c = S_o + G \text{ (dB)}$$

where G is the gain of the microphone and pre-amplifier combination.

Phase response: This is of concern when choosing microphones for sound intensity measurements. It is not the absolute phase response that is of consequence, but the relative phase response between a pair of microphones, necessitating the phase response characteristics to be closely matched. Specially manufactured pairs of microphones, with matched phase responses, are available for sound intensity measurements.

Type of polarization: There are two different types of microphone constructions, one that employs an external voltage supply called the polarization voltage to establish the charge through a large resistor to polarize the air gap between the back plate and diaphragm (externally polarized) and one where the polarization charge is stored in an electret layer on the back plate of the microphone (pre-polarized). Externally polarized microphones are ideal for general field and laboratory use and for high temperature measurements. Pre-polarized microphones are suitable for portable sound level meters where their lowweight and non-requirement of a polarization voltage are desirable. Such microphones also offer better performance in very humid environments.

Directivity: The directional characteristic is the variation in relative sensitivity of the microphone as a function of the angle of incidence. This can be represented graphically by polar plots given at fixed frequencies. The directional characteristic is especially important if the microphone is used for free-field measurements.

Stability of the transducer: The stability of a measurement microphone is a highly desirable feature. It implies high stability of sensitivity and frequency response during short-term and long-term fluctuations of temperature, humidity, etc.

Calibration chart: The calibration chart is an important piece of documentation providing essential information about the performance characteristics of the microphone. It should have comprehensive specifications and performance documentation

in the form of individual calibration charts. The open-circuit sensitivity, frequency response, dynamic range, etc. are the specifications given on the calibration chart. It allows the user to verify the performance of the microphone and pre-amplifier system and to calibrate the measurement system. The microphone should have high suitability for measurement and for calibration using practical and accurate methods. Emphasis is also to be given to periodic calibration with standard noise sources so that an allowable accuracy of ± 1 dB is available throughout the guaranteed frequency range of the microphone and the sound level meter.

5.1.1.3 Environmental Considerations

When selecting a microphone, it is important to consider the environment where measurements are to be made. The microphone should have negligible influence from the environment and should operate satisfactorily over a range of environmental conditions, with very little influence from ambient pressure, temperature, humidity, wind effects, vibration, magnetic and electromagnetic fields, dust, pollution, etc. It should have good mechanical robustness, bump and shock resistance, chemical resistance, corrosion resistance, etc. Some features that should be considered are as follows:

- *Robustness:* In a typical laboratory environment, all microphones can be used, but for field use, more robust general purpose microphones are required.
- *Temperature:* At normal temperatures (-30 to $+125\,°C$), all microphones can be used. At high temperatures (up to $300\,°C$), however, special microphones (e.g., Falcon Range^{TM}) should be used and at very high temperatures, above $300\,°C$, a probe microphone should be used. The tip of the probe on the probe microphone can withstand temperatures up to $700\,°C$.
- *Atmosphere:* All microphones may be used in normal air, but for measurements where corrosive industrial gases exist, for example, when making measurements in industrial chimneys, corrosion resistant (Falcon Range^{TM}) microphones should be used.
- *Wind:* The effect of wind is to produce a turbulent air stream around a microphone and this is wrongly sensed as sound. Wind noise in outdoor measurements can greatly degrade measured acoustic data and therefore a windscreen should be used. It is generally made of a specially prepared moisture and corrosion resistant porous polyurethane foam. Where wind speeds are greater than 24 kilometres per hour (kmph), wind noise may make measurements invalid, even with a wind screen (which typically reduces noise by approximately 15 dB for wind speeds up to 120 kmph). This is because the spectra of measured sound and wind often overlap, especially for low frequency acoustic sources, making it impossible to separate them by band selective filtering.
- *Humidity:* In general, humidity does not influence the sensitivity and frequency response of a microphone. However, some microphones have a layer of quartz on the diaphragm, which absorbs moisture, leading to a decrease in tension of

the diaphragm and a corresponding decrease in microphone sensitivity. Such a situation can arise when sudden changes in temperature and humidity occur when going from a warm, humid environment to an air-conditioned building. The moisture will attenuate the sensitivity of the microphone and as a side effect, increase the inherent noise level. For high humidity, pre-polarized microphones are more reliable because they are more resistant to the attenuation of the polarization voltage which occurs at high humidity levels.

- *Effect of magnetic field:* Microphones are typically designed with materials that provide a very low sensitivity to magnetic fields. Special microphones (Falcon Range™ from Brüel & Kjær, Denmark) are available, which have significantly lower sensitivity to magnetic fields than earlier microphones.
- *Effect of vibration:* Vibration sensitivity is defined as the ratio of output voltage of the microphone to casing acceleration. The vibration sensitivity of the microphone normal to the diaphragm is well defined as it is determined by the mass of the diaphragm; the sensitivity is much smaller in all other directions. Pre-amplifiers and electrical adaptors also contribute to the vibration sensitivity of the transducer.

5.2 Various Types of Microphones

DID YOU KNOW that microphones date back to the 1600s when scientists looked for ways of amplifying sound? The English physicist Robert Hooke developed an acoustic cup and string style phone which is the forerunner of sound transmitters. In 1827, Sir Charles Wheatstone worked on a device that could amplify weak sounds and christened it the 'microphone'. It was Emile Berliner (best known for his invention of the Gramophone), who in 1876, invented what is considered the first modern microphone, while working with Thomas Edison. In 1878, David Edward Hughes, a British-American inventor/music professor, developed the first carbon microphone, variants of which are still being used today. In 1916, the condenser microphone was invented by E.C. Wente while working at Bell Laboratories. In 1920s, the RCA Company developed the first ribbon microphone when radio broadcast for news and entertainment became popular. In 1928, Georg Neumann and Co. designed the first commercial condenser microphone. In 1931 Western Electric developed the first electrodynamic microphone. In 1957, Raymond A. Litke, an electrical engineer with Educational Media Resources developed the first wireless microphone for multimedia applications. In 1964, Bell Laboratories came up with the first electret microphone, which offered greater reliability and higher precision at a lower cost and with a smaller size than the microphones which existed

then. Later came improvements in condenser and dynamic microphones, as well as the development of the clip-on microphones. The twenty-first century has seen the emergence of MEMS microphones that serve a variety of applications, as well as arrays of microphones arranged on the surface of a solid sphere, allowing the sound to be captured from a variety of directions.

In this section, various types of microphones such as carbon granule, condenser, electret capacitor, electrodynamic/moving coil, piezoelectric, and ribbon microphone are discussed. Their principles of operation and typical specifications are also described.

5.2.1 The Carbon Granule Microphone

The carbon button microphone (Fig. 5.1) is one of the earliest types of microphones, having been used extensively in all early telephones and in recording and broadcasting systems. In this device, the sound field acts upon an electroconductive break diaphragm that develops pressure on a capsule of compressed carbon powder. This capsule or carbon button, as it is called, contains carbon granules between two carbon discs. The sound waves incident on the diaphragm cause it to vibrate, transmitting the displacement to the carbon particles and changing their pressure-dependent contact resistance. The front and rear contacts are insulated and brought out to terminals. When a direct current (DC) voltage is applied across the capsule, the alternating resistance produces an alternating voltage drop, which is proportional to the displacement of the diaphragm. The mouthpiece acts as a horn and increases the acoustic pressure on the diaphragm.
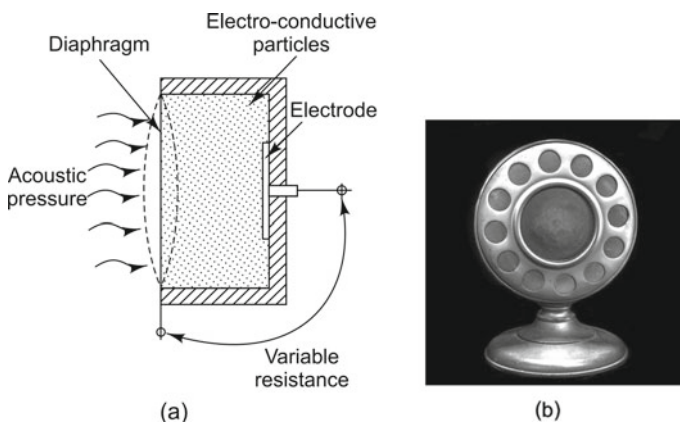
Fig. 5.1 Carbon button microphone: **a** principle of operation and **b** photograph (https://digilab. libs.uga.edu)

Typical specifications

Sensitivity: 0.1–0.2 V/μ bar
Resistance: 30–100 Ω
Frequency range: 250–3200 Hz

Advantages Carbon microphones are cheap and durable.

Disadvantages

(i)   The main disadvantage of the carbon microphone is the so-called 'carbon hiss'.
(ii) If the hermetic seal of the button is damaged, moisture can cause the granules to pack, decreasing the resistance of the microphone and making it less sensitive.
(iii) It does not have a very good frequency response.

5.2.2   Condenser Microphone

The condenser microphone was perfected as a much-needed substitute for the carbon microphone, which was very noisy. It was subsequently used in broadcasting, until it was replaced in the recent past by the dynamic microphone. Today it is almost exclusively used for all technical acoustical measurements. This microphone is so called because it makes use of the principle of a condenser (capacitor) to convert acoustic signals to an output voltage. The classic design of a pressure-sensing microphone is shown in Fig. 5.2a. It consists of a metal housing with a thin metallic membrane or diaphragm, which is adequately protected by a grid and which constitutes one element of a parallel plate condenser. A back-plate is mounted in parallel, behind the delicate and highly tensioned diaphragm (which is designed to have a high resonant frequency), and forms the other plate of the condenser. This fixed plate is called the
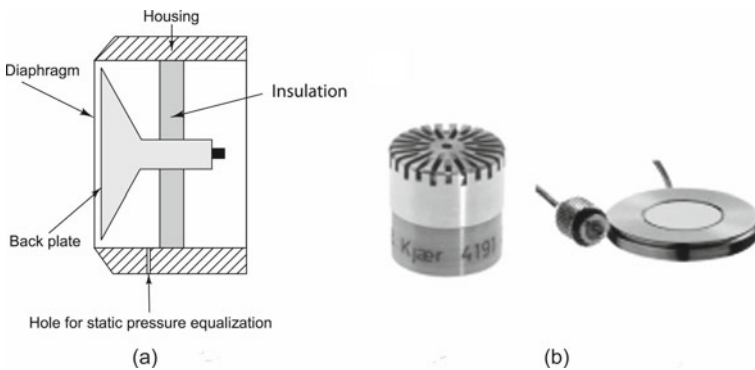
Fig. 5.2 Condenser microphone: **a** schematic drawing, **b** photograph of condenser microphone, **c** photograph of surface microphone (Courtesy of Brüel & Kjær, Denmark)

polarization electrode and is maintained at a potential of about 200 V. A suitable electrical insulator supports this polarization electrode and insulates it from the housing at ground potential. The diaphragm and the front of the back plate form the plates of the active capacitor, which generates the output of the condenser microphone.

When sound is incident upon the diaphragm, the alternating pressure of the sound wave causes small displacements of the diaphragm, resulting in a small change in condenser capacitance. There is a small static pressure equalization vent (which is nothing but a narrow air channel) in the housing of the microphone, allowing air leak to the interior, so that the static pressure inside the microphone assembly is always equal to the ambient atmospheric pressure. The distance between the microphone diaphragm and back plate must be constructed to very stringent mechanical tolerances and is typically set to 20 μm. Similarly, the tension of the diaphragm, which is stretched like the skin of a drum, must be carefully controlled, since the sensitivity of the microphone is inversely proportional to the stiffness of the diaphragm system. The higher the tension, the higher is the stiffness and the lower the microphone sensitivity. The thickness of the diaphragm typically varies from about 1.5 to 8 μm. The stray capacitance between the back plate and the housing is kept as small as possible.

The stiffness and mass of the diaphragm system determine the diaphragm resonance frequency, which in turn determines the upper frequency limit of the transducer. Thus, the microphone size, the sensitivity, and the frequency range are interdependent and cannot be selected in isolation, requiring the user to make a compromise when selecting a microphone. Generally, small microphones work well at higher frequencies and create smaller disturbances in the sound field, but they also have lower sensitivity. The microphone has to be located very close to the pre-amplifier to minimize stray capacitances. Besides, the bias supply must be extremely well regulated and ripple-free, since every variation will result in distortion of the acoustic signal. Figure 5.2b shows condenser microphones of various sizes. Table 5.1 shows their typical specifications. The instantaneous output voltage may be derived from the formula below:

$$V \times C = Q$$

$$(V_0 + e) \times \frac{\varepsilon_0 \kappa \times A}{D_0 + d} = \frac{V_0 \times \varepsilon_0 \kappa \times A}{D_0} \qquad (5.1)$$

$$\therefore\ e = V_0 \frac{d}{D_0}$$

Table 5.1 Typical specifications of condenser microphones

Diaphragm size (in.)	Lower frequency (Hz)	Upper frequency (Hz)	Dynamic range (dB)	Sensitivity (mV/Pa)
1	2–6	8–18	−80 to145	50
1/2	3–7	9–40 k	−10 to 150	50
1/4	4	70–100 k	30–170	3.5
1/8	7	140	30–180	1.0

where A is the area of capacitor plates, C is the instantaneous capacitance between plates, D_0 is the distance between plates at rest position, d is the displacement of the moveable plate (diaphragm) from rest position, V is the instantaneous voltage between plates, V_0 is the polarization voltage, e is the voltage change caused by plate displacement, Q is the constant charge on the capacitor plate, κ is the dielectric constant of air and ε_0 is the permittivity of free space.

It is seen that the output voltage of the system and hence the sensitivity is proportional to the bias voltage and the diaphragm area, and inversely proportional to the gap and the diaphragm stiffness. There is a linear relationship between the output voltage and displacement, even if the corresponding capacitance changes are non-linear.

Surface microphones (Fig. 5.2c) currently available in the market are designed for measurement of the true surface pressure and are highly useful for automotive and aerospace applications. In these microphones, the diaphragm is kept flush with the microphone housing in order to minimize the wind-generated noise of the microphone and the vent for static pressure equalization is placed just next to the diaphragm on the front of the microphone. This is highly appropriate in applications where the microphone is exposed to turbulent pressure fluctuations and where the static pressure can vary rapidly with microphone position. These transducers are meant to be mounted directly on a vehicle during wind-tunnel tests; they may be mounted even on curved surfaces with the help of mounting pads. They are generally rugged, stable and corrosion-resistant. They also have built-in pre-amplifiers. The sensitivities of these microphones are optimized in order to achieve a good noise floor and still allow measurements at higher levels without clipping. The devices also support IEEE 1451.4 (TEDS) for remote identification and reading of their calibrated sensitivity.

Advantages

   (i) Flat frequency response over the entire audio range.
  (ii) Low self-noise.
 (iii) High stability.
 (iv) Reasonably high sensitivity.
  (v) Fairly good stability with respect to time, temperature, and humidity.
 (vi) Small dimensions of the microphone making diffraction of sound negligible.
(vii) Very low sensitivity to magnetic fields.

Disadvantages

 (i) Water droplets form on the diaphragm and back plate (electrode) when the microphone is taken from humid environment to a cool room. Therefore, the capacitance of the microphone changes.
(ii) Polarization voltage of 200 V is required for the back plate.

5.2.3   Electret Capacitor Microphone (ECM)

An electret capacitor microphone also known as pre-polarized microphone is a type
of condenser microphone, which has overcome all the inconveniences of the tradi-
tional capacitor microphone, like the need for a polarization power supply. It is now
used almost universally, in applications such as telephones and consumer electronics
and has replaced the carbon microphone for general purpose applications. It is an
excellent, easy to use and inexpensive device, which can be operated on a 5 V supply
voltage, but it normally requires an integrated pre-amplifier. The development of the
polymer electret film and the field-effect transistor has spurred its growth. The exter-
nally applied charge described in Sect. 5.2.2 is replaced by a fixed charge in an electret
material in this microphone. An electret is a stable high temperature polymer material
that has been permanently electrically charged or polarized. The name comes from
electrostatic and magnet. The electret contains trapped or 'frozen' electrical charges,
which produce the required electrical field in the air gap. This charge remains inside
the electret for thousands of years. The static charges in the material get aligned
and permanently embedded in an electret in the same way that a magnet is made by
aligning the magnetic domains in a piece of iron. Electrets are made by first melting
a suitable dielectric material like plastic or wax and then allowing it to re-solidify
in a powerful electrostatic field, causing the static charges to align themselves in the
direction of the electrostatic field to produce a permanent electrostatic 'bias'. The
sensitivity of the microphone depends on the field strength in the air gap. Typically,
the field strength and equivalent polarization voltage are similar to those used for
microphones with external polarization, namely 200 V. These pre-polarized micro-
phones, which are generally more complex than ordinary condenser microphones
are mainly intended for use in portable battery operated and hand-held instruments.

There are three major types of electret microphones:

- *Foil-type or diaphragm-type:*   This is the cheapest type, where a film of electret
  material is used as the membrane or diaphragm itself. Obviously, the quality of
  this device is poor, since the electret material does not make a particularly good
  diaphragm.
- *Back electret:*   In this type, the diaphragm is made of an uncharged material
  and an electret film is applied to the back plate of the microphone capsule. This
  design yields a transducer of better quality.
- *Front electret:*   This is the newest type, where the condenser is formed by the
  diaphragm and the inside surface of the capsule, the back plate being totally
  eliminated. Any conductive film may be used for the diaphragm in this design
  and the electret film is applied to the inside front cover.

Figure 5.3a shows the principle of operation of the foil type electret microphone, and
Fig. 5.3b shows photographs of electret microphones.

Electret microphones were traditionally considered low-cost and low-quality
transducers. However, the best ones today match the capacitor microphones in almost
all characteristics except low noise performance and can even offer the long-term sta-
bility and ultra-flat response of a condenser microphone. The capacitor in the electret
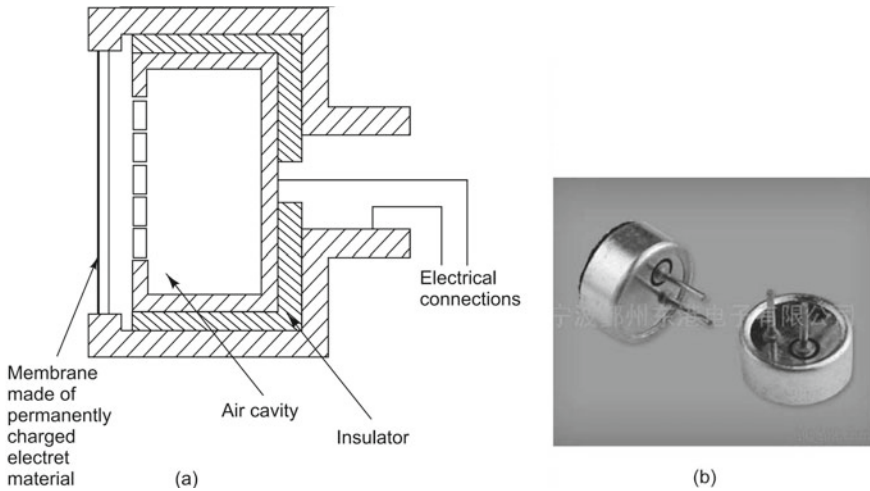
Fig. 5.3 Electret microphone: **a** principle of operation, **b** photographs (Courtesy of www.ecvv. com)

is carefully sealed away from floating charges so that the electret does not get neutralized. These microphones have many applications such as high-quality recording and built-in microphones in small sound recording devices and telephones.

Advantages

  (i) This microphone is remarkably small.
 (ii) Polarization voltage is not required.
(iii) It has omni-directional characteristics.
 (ii) It is very inexpensive.

Typical specifications

Size: 9 mm diameter and 6 mm height.
Frequency bandwidth: 15 kHz.
Sensitivity: 0.6 mV/μbar.
Power supply: 2–10 V at <1 mA.

5.2.4 Electrodynamic/Moving Coil Microphone

This microphone is often found in recording studios, broadcast and motion picture production, home high fidelity and video recording systems and on stages for live sound reinforcement. The principle of the dynamic microphone was known in 1877 when Bell developed the telephone, but lack of electronic amplification made

it impossible to use. Dynamic microphones are very similar to conventional loud-speakers in that both have a diaphragm (or cone) with a voice coil attached near the apex. Both have a magnet with the coil in the annular gap. The difference is in how they are used. A dynamic microphone operates like a speaker 'in reverse'. In fact, many communication systems use small speakers with lightweight cones as both speaker and microphone. They simply move the same transducer from one end of the amplifier to the other! The diaphragm, carrying a coil of fine wire placed in a strong annular magnetic field, is exposed to sound waves and is moved by changing sound waves. This moves the coil, which causes current to flow as lines of flux from the magnet are cut. The voltage e induced in the coil is proportional to its amplitude of vibration, which in turn, depends on the sound pressure and is expressed as

$$e = Blv \tag{5.2}$$

where B is the flux density in Tesla (webers/m^2), $l = \pi d N$ is the total length of wire in the coil, d is the diameter of the coil and N the number of turns; v is the velocity of the diaphragm in m/s.

Figure 5.4a shows the construction of a moving coil microphone and Fig. 5.4b shows a photograph. The dome and diaphragm act like a rigid piston and carry the coil of wire, which moves in the annular gap of the magnetic field. The pole pieces of the microphone are of soft iron and permanent magnets provide the field. The diaphragm is acted upon by the acoustic pressure, so that the microphone is a pressure microphone, with omni-directional characteristics.
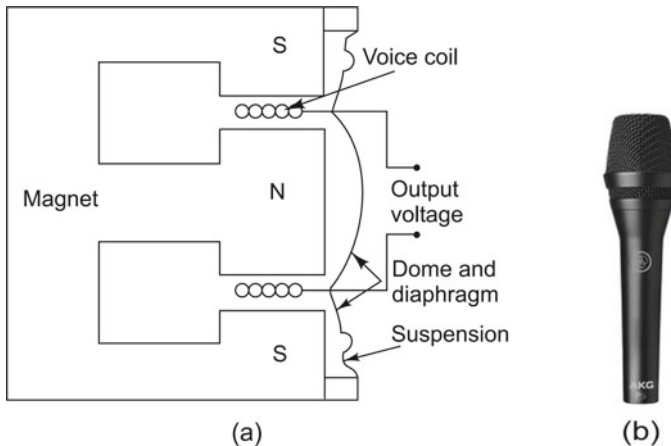
Fig. 5.4 Electrodynamic microphone: **a** schematic drawing, **b** photograph (Courtesy of https://www.harmanaudio.in)

Advantages

(i) Dynamic microphones are rugged and reliable.
(ii) They do not need bias voltages or batteries or external power supplies.
(iii) They are capable of smooth response and are also available with tailored response for special applications.
(iv) They have a fairly good sensitivity with an excellent signal-to-noise ratio (SNR).
(v) Dynamic microphones are of the low-noise type and are relatively rugged.
(vi) Most microphones of superior quality are still dynamic microphones.
(vii) They are not susceptible to moisture.
(viii) They operate at fairly high temperatures.
(ix) They need very little maintenance and with reasonable care will maintain their performance for many years.

Disadvantages

(i) Their sensitivity is quite low, no more than -90 or -100 dB, so amplification is essential. At least 40 dB can be gained with transformers, bringing the output up to -50 dB when applied to the amplifier, which is not too bad.
(ii) The other drawback of these microphones is that they are susceptible to external magnetic fields.
(iii) The high-quality microphones are relatively expensive.

Typical specifications

Internal impedance: $<10\,\Omega$.
Size of dynamic element: 1 in. diameter.
Frequency response: 40–18,000 Hz.
Output level: -53 dBV/Pa @ 1000 Hz.

5.2.5 Piezoelectric Microphone

Piezoelectric microphones employ crystals that can become electrically polarized and produce voltages proportional to the strain. Since the piezoelectric effect is reversible, a piezoelectric microphone will function as a source when an alternating voltage is applied to its terminals. The description of the piezoelectric effect is made complicated by the many directional quantities and the crystal symmetries that enter. Rochelle salt is one of the first materials to be used in acoustic transducers. Single crystals of Rochelle salt have been employed widely in microphones. Unfortunately, the crystals deteriorate in the presence of moisture and get permanently damaged if the internal temperature exceeds 46 °C. Historically, quartz crystals have been of great significance as transducers. They are impervious to water as well as most corrosive materials, can be subjected to extreme temperatures and are easily manufactured.
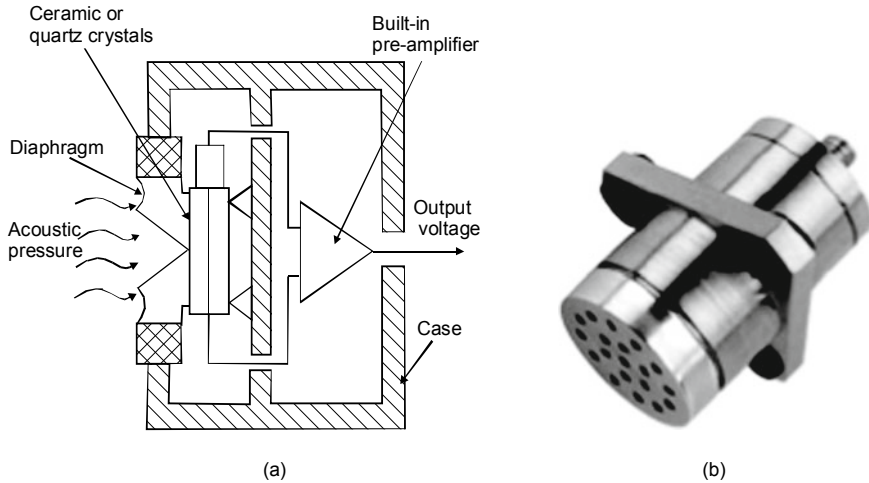
Fig. 5.5 Piezoelectric microphone: **a** schematic drawing, **b** photograph (Courtesy of Brüel & Kjær, Denmark)

They can generate longitudinal waves, shear waves or combinations. Other useful materials are sintered ceramics including barium titanate, lead zirconate, lead titanate and mixtures of these and associated compounds. If a microphone is described as 'crystal', it usually contains Ammonium dihydrogen phosphate (ADP); if it is called 'ceramic', barium titanate is the active element. Piezoelectric microphones are often constructed in a manner similar to that shown in Fig. 5.5. Here, the sound waves act on a light diaphragm whose centre is linked to an end or corner of the piezoelectric element by means of a driving pin. Although a single element could be used, two elements are usually sandwiched together to form a bimorph. Stresses in the crystals, produced by a sound field, generate an output proportional to the acoustic pressure. The voltage output of a bimorph element is proportional to its strain. Many designs incorporate a built-in pre-amplifier next to the crystal; this arrangement reduces the electrical noise and output impedance.

Advantages They have satisfactory frequency response and are relatively high in sensitivity, low in cost and small in size. Piezoelectric microphones are therefore widely used in public-address systems, sound level meters, and hearing aids.

Disadvantages Piezoelectric microphones give low output at moderate internal impedance, and must always be used with amplifiers.

Specifications

Charge sensitivity: 0.16 pC/Pa.
Frequency response (± 1 dB): 2–4 kHz and (± 3 dB): 1–10 kHz.
Dynamic range: 100–180 dB.
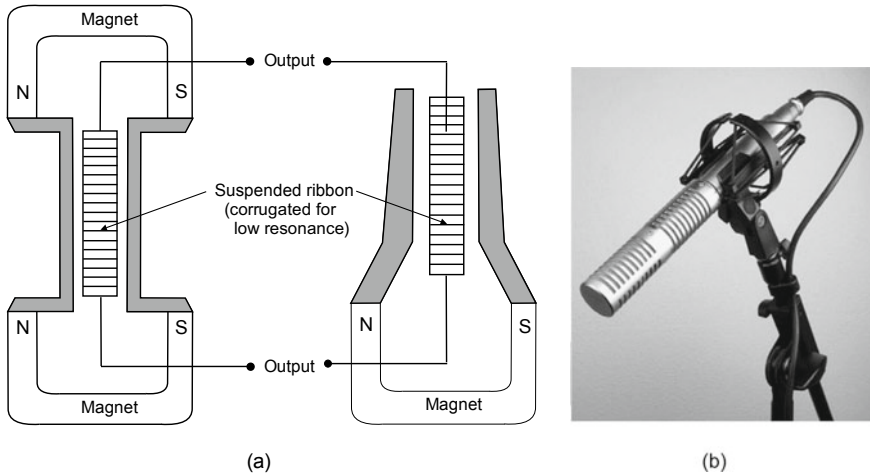
5.2.6   The Ribbon Microphone

The microphones discussed in Sects. 5.2.1–5.2.5 are called pressure microphones because they are activated by the acoustic pressure acting on one side of a diaphragm. These microphones are essentially omni-directional because pressure is a scalar quantity. The ribbon microphone is a pressure gradient sensitive microphone and therefore has directional sensitivity. The pressure gradient is a vector quantity parallel to the direction of propagation and in phase with the displacement.

Let us consider a harmonic pressure wave $p(x, t)$ impinging on a 'surface' not 'diaphragm'. 'Surface' implies exposure to pressure equally on front and back, while 'diaphragm' implies exposure on the front side only. The net driving force will be equal to the product of the difference in pressures between the two sides and the common surface area A. This pressure difference can be computed as the space rate of change of acoustic pressure (pressure gradient) $\partial p / \partial x$ multiplied by the effective acoustical distance L between the two sides of the surface. Let the pressure wave have an angle of incidence θ. When θ is zero, the sound wave is incident normally on the surface. The expression for the desired angle-dependent force F is as below:

$$F = -(\partial p / \partial x)AL \cos \theta = ikpAL \cos \theta = i\omega\rho_0 vAL \cos \theta \qquad (5.3)$$

This force is at right angles to the pressure and is proportional to the propagation constant $k = \omega/c_0 = 2\pi/\lambda$ where c_0 is the speed of sound. This equation holds for $kL \ll 1$, i.e., when the size of the microphone is small compared to the wavelength of sound. Equation (5.3) states that the velocity of the surface v is proportional to p and independent of frequency; this is true irrespective of whether the pressure wave is plane or spherical as long as the distance from the source is large and wavelength is small. If the latter conditions are not satisfied, the velocity will be inversely proportional to frequency, giving rise to what is called the proximity effect.

The most common pressure gradient microphone is the ribbon microphone. Figure 5.6 shows a ribbon microphone. The pressure gradient sensitive surface or the ribbon is a corrugated aluminium foil and is suspended in a strong magnetic field. Moving the ribbon will change the electromotive force (emf) generated which is proportional to v. The distance L depends on the size of the baffle in which the ribbon is suspended. The generated emf is typically small; the velocity can be made higher by reducing the mass of the surface. The sensitivity is proportional to $\cos \theta$. The ribbon microphone is hence equally sensitive to sound coming from the front and back and is completely insensitive to sounds coming from 90° or 270°. By combining the cosine sensitivity of the ribbon microphone with the isotropic sensitivity of a pressure microphone, we get a cardioid microphone where the response is proportional to $(1 + \cos \theta)$. Such a microphone favours sounds coming from 0°, and discriminates against sounds coming from 180°.

Fig. 5.6 Ribbon microphone: **a** schematic drawings for two different designs, **b** photograph (Courtesy of www.cascademicrophones.com)

Advantages

 (i) Ribbon microphones are very sensitive and accurate.
(ii) They can also be oriented so that noise sources can be put in a zone of low
     sensitivity. These characteristics have made the ribbon microphone the standard
     for broadcasting, and the lozenge-shaped shiny microphone very popular.

Disadvantages They are very delicate and expensive to make and repair.

Specifications

Frequency response: 20 Hz–20 kHz.
Maximum SPL: 132 dB.
Output sensitivity: 30 mV/Pa.
Equivalent noise level: 6 dB SPL, A-weighted as per IEC 651.
Dynamic range: 126 dB.
Directional pattern: Bi-directional.

5.2.7 MEMS Microphone

One of the latest developments in the area of microphones is that using Micro Electro
Mechanical Systems (MEMS). Owing to their extremely small size, these microphones find applications in wearable products such as earphones for hands-free calls
and headphones with active noise cancelling (ANC) functions, action cameras, digital cameras, as well as smartphones and speech recognition interfaces. The MEMS
microphone has a very compact package, which houses a MEMS chip, an acoustic

transducer and an application-specific integrated circuit (ASIC) for signal processing. The MEMS chip is manufactured by forming a microstructure on a silicon wafer using photolithography, etching, etc., the same techniques that are used in semiconductor manufacturing technology. In this process, many layers of different materials are deposited on a silicon wafer and then the unwanted material is etched away. A moveable diaphragm or membrane and a fixed back plate are created over a cavity in the base wafer. The sensor back plate is made in the form of a stiff perforated structure that allows air to move through it easily, while the diaphragm is a thin solid membrane that bends due to changes in air pressure caused by sound waves. The microphone is connected to the external environment through a sound hole. The diaphragm and the back plate are arranged such that there is a small gap between the sound hole and the back chamber of the MEMS chip.

When a sound wave falls on the diaphragm, the changes in air pressure cause the diaphragm to flex. The thicker back plate, on the other hand, remains stationary as the air moves through its perforations. The movement of the membrane changes the capacitance between the diaphragm (vibrating plate) and the back plate (fixed plate), which act as two plates of a parallel-plate capacitor. The MEMS microphone is therefore said to be of the 'capacitance type'. When the gap changes due to vibration, the capacitance changes and the change in capacitance is converted to an electrical signal by the ASIC, which uses a charge pump to place a fixed charge on the diaphragm. The ASIC then measures the voltage variations due to the changes in capacitance caused by the motion of the diaphragm due to sound waves. Analogue MEMS microphones produce an output voltage proportional to the instantaneous pressure fluctuation. They usually have three pins: the output, the power supply voltage, and ground. In most applications, a low noise audio ADC is also required to convert the output of the microphones into digital format for processing and/or transmission. The chip is only a few millimetres in size and typically around three MEMS microphones are mounted in one smartphone.
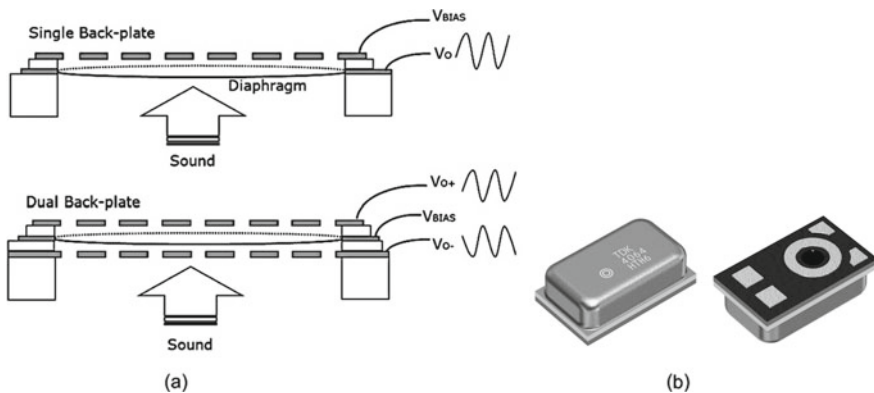
Fig. 5.7 MEMS microphone: **a** Schematic diagram, **b** Photograph (Courtesy of https://product.tdk.com)

Figure 5.7 shows schematic drawings of general single-back plate and double-back plate MEMS microphones. The single-back plate microphone is an asymmetrical structure in which one back plate faces the diaphragm. The electrostatic force due to the bias voltage causes the diaphragm to get displaced to one side (back plate side) even when there is no sound wave acting on it. When a sound wave impinges on the diaphragm, the latter gets displaced asymmetrically and the total harmonic distortion (THD) increases with increasing sound pressure level, leading to clipping. On the other hand, the double-back plate type is a symmetrical structure in which two pieces of back plates are arranged face-to-face on both sides of the diaphragm. With this arrangement, the diaphragm does not get displaced in the absence of a sound wave because the electrostatic force due to the bias voltage acts uniformly on both sides of it. So even if a large sound pressure acts on the diaphragm, hardly any clipping occurs; thus it has superior THD-SPL characteristics. This feature along with a high SNR allows it record with good quality even for remote sound sources.

The sensitivity of these microphones depends on the magnitude of the output voltage for a given sound pressure. For a MEMS microphone, the higher the displacement of the diaphragm is, the more sensitive the microphone is, for the same sound pressure. Therefore, when a MEMS microphone is downsized, the area of the diaphragm becomes smaller, leading to reduced sensitivity. However, the sensitivity of this microphone also depends on its material, thickness, internal stress, and the distance from the back plate. Many of the present-day MEMS microphones have an extremely flat frequency response and a wide range of frequencies from low to ultrasonic range.

Typical Specifications

Sensitivity: -40 dB (ref. 1V/Pa).
Output impedance: $300\,\Omega$.
SNR: 55 dB.
Operating voltage: 2 V.
Current consumption: $200\,\mu A$.
THD: 1–10%.

Advantages of MEMS Microphone

- Low power consumption.
- Very small size.
- Very good SNR.
- Improved noise cancellation.
- Wide frequency response.
- Good high-temperature and vibration withstand capability.
- Very good sensitivity.
- High reliability and robustness.

Disadvantages of MEMS Microphone

- Very complex design process and procedure.
- Low cost achievable only with mass production.

- Need for large amplification.
- Difficulty in integration of MEMS microphone.
- Better quality of materials which can withstand higher power desirable.

5.3   Acoustic Exciters

DID YOU KNOW that horns were the earliest devices used to magnify sounds? They amplify sounds mechanically. From 1880 to the 1920s, Thomas Edison, Magnavox and Victrola developed effective horns, but they could amplify the sound to only a certain extent. In 1861, Johann Philipp Reis developed a simple type of electronic loudspeaker, which was barely able to reproduce noise. In 1876, Alexander Graham Bell tried to develop a speaker based on Reis' work. In 1877, Werner Von Siemens proposed the idea of the electromagnetic coil driven speaker; unfortunately, there was no technology available to amplify the sound. Between 1877 and 1921, many inventors toyed with the idea of the electrodynamic loudspeaker, but the outputs were highly distorted sounds. Developments in vacuum tubes in the 1910s helped achieve good amplifiers. When voice radio evolved in the 1920s, the horn and electrodynamic loudspeaker became very popular. In 1921, Radio Corporation of America sold their first speaker. Major advances in loudspeaker design had to wait till our understanding of electricity, radio, sound waves, mechanics, chemistry, and physics was mature. One of the challenges in loudspeaker design was developing a speaker, which could faithfully reproduce both low and very high frequencies. This became possible over the years since the speaker's abilities to reproduce frequencies got even better with the progress in entertainment systems. Later developments in loudspeakers include the electrostatic flat panel speaker, ribbon-driven speakers and piezoelectric speakers, to name a few. With more and more research going on, speakers are becoming smaller, more efficient and more durable.

Loudspeakers convert electrical energy to acoustic energy. The commonly used speakers are: (a) electrodynamic loudspeaker and (b) electrostatic loudspeaker (ESL). There is also a very special type of acoustic exciter called the electropneumatic transducer (EPT).

## 5.3.1   Technical Specifications of a Loudspeaker

- *Type of speaker:* Depending on the frequency range of operation, speakers may be classified as woofers (low frequency), squawkers (mid-frequency), tweeters (high frequency) or full-range speakers.

- *Rated power:* This is the maximum input power that can be given to a speaker before unacceptable distortion occurs or alternately before the loudspeaker gets thermally destroyed. Specifications are in terms of the nominal (or continuous) power and peak (or maximum short-term) power that a loudspeaker can handle. A woofer may be damaged at much less than its rated power if it is driven beyond its mechanical limits at lower frequencies (e.g., by heavy bass music). Tweeters can also be damaged by music or by a sine wave input at high frequencies due to amplifier clipping.
- *Impedance:* This value is required for appropriate impedance matching with the power amplifier and is typically 4 or 8 Ω.
- *Frequency response:* This is the range of frequencies over which the speaker produces a constant output and is the measured output or SPL when the speaker is powered over a specified range of frequencies for a constant input level. It is often quoted as the maximum deviation in SPL such as within ±3 dB. Generally, a frequency response plot is provided by the manufacturer.
- *Sensitivity:* This is the SPL produced by a loudspeaker and is measured on-axis in anechoic conditions (in a non-reverberant environment) typically at one or more specified frequencies when driven with an input of 2.83 V corresponding to the voltage across a standard 8 Ω speaker driven at 1 W. It is usually specified in dB SPL, and is measured at 1 m.
- *Maximum SPL (dB):* This is the highest output the loudspeaker can produce, without getting damaged or without exceeding a particular distortion level. The rating given by manufacturers is often inflated and is misleading since it is given without reference to frequency range or distortion level.
- *Directivity:* This can be represented by polar plots given at fixed frequencies as for the case of microphones. Curves of equal SPL are plotted representing the propagating wave front. At low frequencies, sound propagation is omni-directional, meaning sound spreads out in all directions; at high frequencies it becomes much more directional, separating into lobes.
- *Baffle or enclosure type:* It has also been found that the production of a high-fidelity loudspeaker requires that the speaker be put in a closed box. Hence, the type of enclosure is mentioned for enclosed systems only: typically sealed, bass reflex, etc.

5.3.2 Electrodynamic/Moving Coil Loudspeaker

Figure 5.8a is a simplified drawing of a moving-coil loudspeaker which is nothing but a dynamic microphone in reverse. Of course, the designs are quite different, because in practice a loudspeaker with a large power rating cannot be used as a microphone. In the loudspeaker, a light voice coil is mounted such that it can move freely in the magnetic field of a strong permanent magnet. The speaker cone is attached to the voice coil and is connected by a flexible mounting to the outer ring of the speaker support.
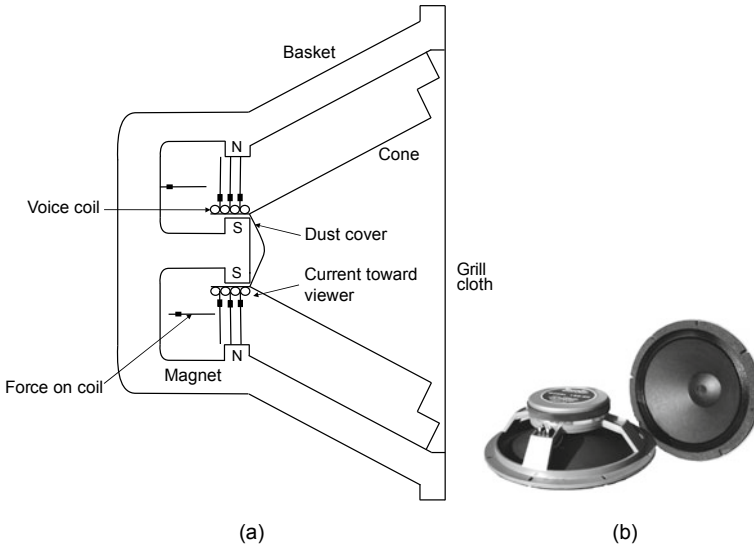
Fig. 5.8   Moving coil loudspeaker: **a** schematic drawing, **b** photograph (www.iconicspkrs.com)

A magnetic field is created by current flowing through the moveable voice coil; this interacts with the magnetic field of the speaker's magnet, forcing the coil and the attached cone to move back and forth. This, in turn, moves the air to produce a sound which is a replica of the original electrical wave. The vibrating diaphragm (cone) is usually appreciably larger than the voice coil, to enhance the efficiency of radiation at the lower frequencies. Figure 5.8b shows the photograph of a dynamic speaker.

Loudspeakers are almost always the limiting element in the fidelity of a reproduced sound at home or in a theatre. The free cone speaker is very inefficient at reproducing sound wavelengths larger than the diameter of the speaker and becomes directive at high frequencies. For this reason, a number of speakers with different diaphragm diameters are often used in high-fidelity applications to cater to various frequency ranges. There are relatively massive speakers, i.e., speakers of 25 to 30 cm diameter (woofers) radiating the lower frequencies (subwoofers and woofers) up 500 Hz, smaller speakers for the mid-range frequencies (squawkers), i.e., speakers of 10 to 12.5 cm diameter and still smaller ones of around 7.5 cm diameter to radiate the highest frequencies (tweeters and super tweeters) for the range 300–20,000 Hz. Higher power is required in the bass range making multiple drivers a practical necessity. These various speakers can be driven through electrical filtering networks that deliver to each its appropriate range of frequencies or can be band-limited naturally through the characteristics of their own electrical, mechanical, and radiation impedances.

Loudspeakers are generally enclosed for the reason that an unenclosed speaker behaves as an acoustic dipole, giving rise to a poor low frequency response (because sound from the back of the diaphragm cancels that from the front) and highly directional radiation. By mounting the loudspeaker in an infinite baffle, it is forced to

behave as a monopole and radiate into the half space in front of the baffle. To create such an infinite baffle, the baffle is folded around the back of the loudspeaker, or in other words, the loudspeaker is put in an enclosure.

Advantages

  (i) Wide dynamic range.
 (ii) High power handling.
(iii) High sensitivity.
(iv) Relatively simple design.
 (v) Very rugged.
(vi) Small size and low cost.

Disadvantages

  (i) The electrodynamic loudspeaker is a low impedance device with high current requirements.
 (ii) The total weight of the moving parts of a dynamic speaker: its voice coil, suspension system, and cone is much more than the air that the speaker drives, thus preventing response to high frequency signals.
(iii) A magnetic speaker resonates at many frequencies, behaving somewhat like a bell. The transient response is poor since the 'ringing' continues long after the original note has stopped.
(iv) It is not possible to achieve high efficiency (especially at low frequencies) along with compact enclosure size and adequate low frequency response. Therefore, one has to make a trade-off when designing a speaker system.
 (v) It has a low efficiency of around 5%, resulting in say 10 W of sound power for 200 W of electrical power from an amplifier.

Typical specifications

  Impedance: 4–8 Ω.
  Music power rating: 200 W.
  Power rating for sine wave: 100 W.
  Frequency response ($+/-3$ dB): 40 Hz–20 kHz

5.3.3 Electrostatic Loudspeaker (ESL)

The ESL uses a thin flat diaphragm sandwiched between two electrically conductive grids forming a pair of capacitor plates, with a small air gap between the diaphragm and grids. The diaphragm is usually made from a polyethylene terephthalate (PET) film of thickness 2–20 μm, with exceptional mechanical properties and with a conductive coating material such as graphite. The membrane is imparted an electrostatic charge with a high DC polarizing voltage, while the electrodes are fed with ground
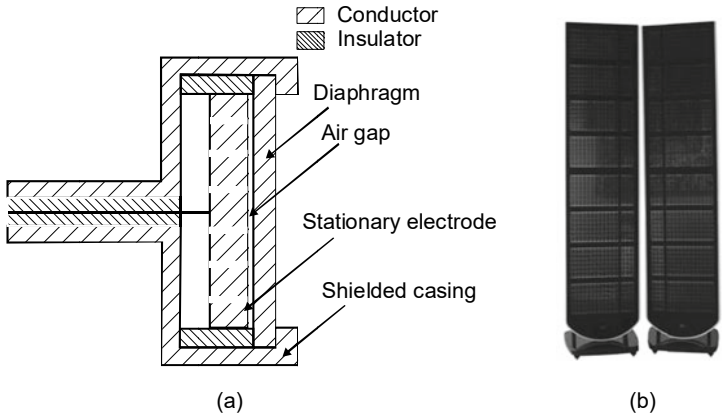
Fig. 5.9 Electrostatic loudspeaker: **a** schematic drawing, **b** photograph (Courtesy of http://www.kingsaudio.com.hk)

potential. Typical polarizing voltages are of the order of 2–3 kV. One plate is held stationary, while the other, the diaphragm, moves in response to mechanical or electrical excitation. If a transient voltage is applied across the plates, the diaphragm moves in response to the changing charge. For low-distortion operation, the diaphragm must operate with a constant charge on its surface, rather than with a constant voltage. The diaphragm's conductive coating is chosen and applied in such a manner as to give it a very high surface resistivity. The grids are driven by the audio signal, front and rear grids being driven in phase opposition to each other. Using grids on both sides cancels out non-linearity. Hence, a uniform electrostatic field proportional to the audio signal is produced between the grids. This causes a force to be exerted on the charged diaphragm, and its resulting movement drives the air on either side of it. The grids have to be able to generate a very uniform electric field, while still allowing sound to pass through. Suitable grid constructions are therefore perforated metal sheets, a frame with tensioned wire or wire rods. Figure 5.9 shows an electrostatic speaker.

The principle of operation of an ESL is as follows: A charge, Q, placed in a field of strength $V_{sig}/2d$ expressed in N/C or V/m experiences the following force:

$$F = QV_{sig}/2d \tag{5.4}$$

where F is the force (N), V_{sig} is the signal voltage, Q is the strength of the charge (C), and $2d$ is the distance between the electrodes (m), d being the distance between the diaphragm and one electrode. Since $Q = CV_0 = $ constant, the total capacitance between the plates will vary as the diaphragm moves and the voltage falls. Here C is the total capacitance (F) of the system, and V_0 is the polarizing voltage (V).

Advantages

 (i) The main advantage of the speaker is the extremely lowweight of the diaphragm, due to which it offers a faithful reproduction (both in amplitude and phase) of high-frequency content, making it superior to the heavy moving coil speaker. This is highly desirable for music, since overtones and higher harmonics contribute significantly to the overall richness of the sound.
 (ii) Most electrostatic speakers are tall, flat and thin in design and generally do not have enclosures, acting as vertical dipole line sources with highly directional characteristics. This brings about different acoustic behaviour in comparison with conventional electrodynamic loudspeakers. Curved panels are also used, making the positioning of the speakers in an enclosure a little less stringent.
(iii) Since electrostatic speakers do not need an enclosure, problems associated with cabinet design can be got rid of.
(iv) With rapid progress in the research and development of materials and cheaper polymers, these speakers are bound to become more popular.

Disadvantages

 (i) The chief disadvantage is poor bass response due to phase cancellation arising from the lack of an enclosure, and the difficulty in reproducing low frequencies with a diaphragm with small allowable displacement. This is often remedied through a hybrid design using a dynamic loudspeaker to handle lower frequencies, with the electrostatic diaphragm to handle middle and high frequencies.
 (ii) The directionality of these speakers can also be a disadvantage since the area where the effect of proper stereo imaging can be heard in an auditorium becomes small.
(iii) These speakers are sensitive to ambient humidity levels.
(iv) They have a tendency to attract dust, insects, conductive particles and moisture, leading to deterioration in diaphragms and necessitating periodic replacement.
 (v) They are large in size and expensive, putting them out of the reach of most consumers.

Typical specifications

Frequency response: 50 Hz–25 kHz.
Sensitivity: Equivalent to 84 dB/W @1 m.
Impedance: 8 Ω.
Maximum power: 260 W (music).

5.3.4 Ribbon Driven Speakers

The principle of operation of a ribbon speaker is very similar to that of an electrodynamic loudspeaker, the difference being that a ribbon driver uses a strip of very thin

Ribbon diaphragm

Transformer

+

Audio input

-

N   S

Magnet

(a)   (b)
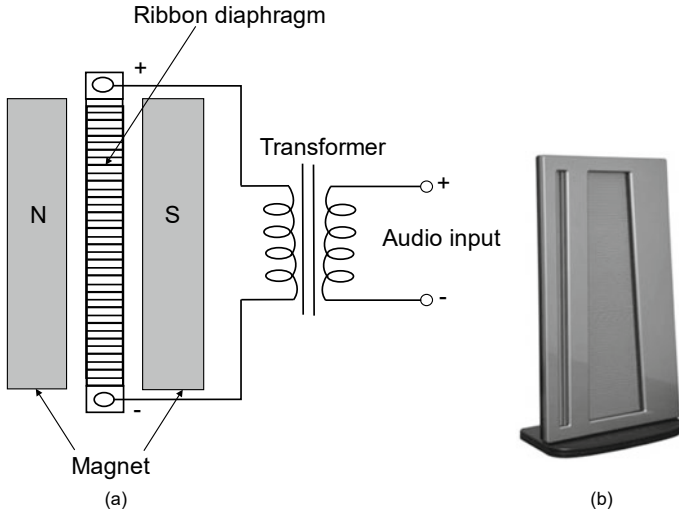
Fig. 5.10   Ribbon speaker: **a** Schematic drawing, **b** photograph (Courtesy of apogeeacoustics.com)

(about the width of human hair) and light conductive material (typically aluminium) as a diaphragm suspended between the north and south poles of a magnet as seen in Fig. 5.10. In the dynamic speaker, on the other hand, there is a cone attached to a voice coil suspended in a magnetic field and this drives the air. The ribbon can be thought of as a voice coil stretched over its entire length. It may be straight, but is generally pleated for additional strength. It is held tight at its top and bottom and is free to move in the middle as a diaphragm; besides, it is tensioned in the factory for optimum performance. The ribbon speaker is usually mounted in a flat, open-air panel that radiates sound both to the rear and front.

When an electric current is passed through the electrically conductive ribbon, it creates a magnetic field around the latter; the generated field interacts with the field of the permanent magnet, causing the ribbon to move back and forth, displacing the air and creating sound. Thus, the ribbon works as both the voice coil and the diaphragm. Since the impedance of the ribbon is very low, often a fraction of an ohm, it has to be matched to the relatively higher impedance of the power amplifier and a transformer is used for this.

True ribbon drivers are called 'line-source transducers' since they produce sound over a line rather than from a point. A ribbon's radiation pattern varies drastically with frequency. At low frequencies, when the ribbon's length is much smaller than the wavelength of sound, it acts as a point source and radiates sound spherically around it. When the frequency increases and the wavelength of sound becomes comparable to the ribbon's dimensions, the radiation pattern is more like a cylinder around the ribbon than a sphere. At very high frequencies, the ribbon radiates only in the horizontal direction. This is advantageous in that the listener hears more direct sound

from the speaker and less reflection from the walls and ceiling.

Advantages

- Very low mass of the ribbon (one quarter) as compared to the mass of the coil in a dynamic speaker.
- Good response to transients and high frequencies because of lower inertia.
- Lack of a box or cabinet which results in improved performance.
- Stunning sound quality, unmatched by dynamic drivers.
- 10 times the radiating area of a dome tweeter's diaphragm.

Disadvantages

- Low sensitivity.
- High amplifier power required to drive the speaker.
- Very low impedance of the ribbon requiring an impedance matching transformer.
- Difficulty in positioning in a room with small variations changing the sound.
- Excitation of resonant frequency produces the irritating sound of crinkling aluminium foil.
- Exact tension is required for the ribbon.

Typical specifications of ribbon tweeter

Frequency response: 1–40 kHz.
Nominal Impedance: 8 Ω.
Sensitivity: 102 dB @ 1 W/m.
Diaphragm: Laminated Aluminum.
Magnet: Nd-Fe-B.
Mass: 1000 gm.

5.3.5   Piezoelectric Speaker

Piezoelectricity was discovered by the French brothers Pierre and Jacques Curie in 1880. It was used in the early 1900s in vibration transducers, but it was not until the 1950s with the invention of high-quality piezoelectric ceramics that the first speakers using this technology became popular. The piezoelectric speaker works on the reverse piezoelectric effect in which application of an electric charge to a piezoelectric crystal, produces a mechanical strain on the crystal, causing it to deform and thereby generate a small physical force. This principle is used in the construction of thin-form-factor speakers, which are also called piezo speakers, piezo buzzers or ceramic speakers.
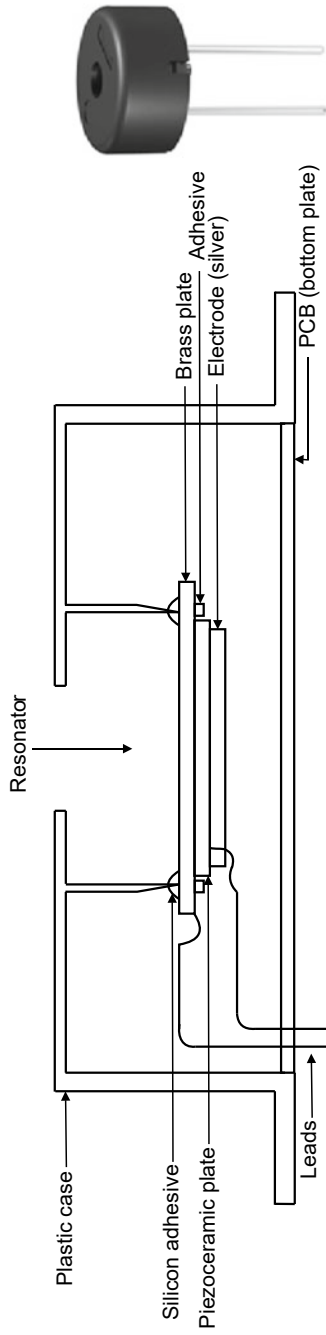
Fig. 5.11 Piezoelectric speaker: **a** schematic drawing, **b** photograph (Courtesy of https://product.tdk.com)

Figure 5.11 shows a cross-sectional view of a typical piezoelectric speaker. It makes use of a piezoceramic plate coated with silver on both sides. One side is bonded to a thin brass plate around the plate's circumference with an adhesive and becomes one electrode, the other side of the plate becoming the second electrode. The brass plate is fixed to a supporting ring, which is part of a moulded plastic casing, using a silicone adhesive. When a voltage is applied to the electrodes from the output of an amplifier, the piezoceramic plate develops a strain proportional to the voltage; this causes a diaphragm, which is fixed onto the plate to bend. As a result, the air in front of the plate moves and produces sound. The electrical signal thus gets transformed into a sound wave. There are two ways in which an audible lead zirconate titanate (PZT) crystal may be excited to produce sound: by driving it with a single-ended signal (0 to $+5$ V) or with a bipolar signal (-2.5 to $+2.5$ V). Piezoelectric speakers are often integrated into a horn loudspeaker.

Advantages

- They are environmentally robust.
- They are useful in underwater applications.
- They are suitable for ultrasonic applications.
- They are of low cost and small size.
- In spite of being extremely thin, they have acceptable sound characteristics.

Disadvantages

- Piezoelectric speakers require high drive voltages and drive currents.
- They have poor performance at low frequencies.
- The SPLs produced are small.

Typical specifications

- With 10.6 V RMS (30 V_{p-p} sine) input, output $= 0.46$ W ($= 43$ mA, 10.6 V).
- With 21.21 V RMS (60 V_{p-p} sine) input, output $= 1.85$ W ($= 87$ mA, 21.21 V).

5.3.6   *Electropneumatic Transducer*

This is a highly specialized transducer, which is capable of producing an enormous amount of acoustic power, of the order of a few 1000s of Watts resulting in SPLs of the order of 150 dB or more and is used for conducting high level acoustic tests. It is an electrodynamically controlled air modulator, which converts the pressure energy of air into acoustic energy quite efficiently and can reproduce almost any type of random signal fed into its coils. The principle of operation of this transducer is similar to that of a siren. The transducer consists of a magnetic structure which houses and supports a pneumatic valve. The magnetic structure is designed to receive and
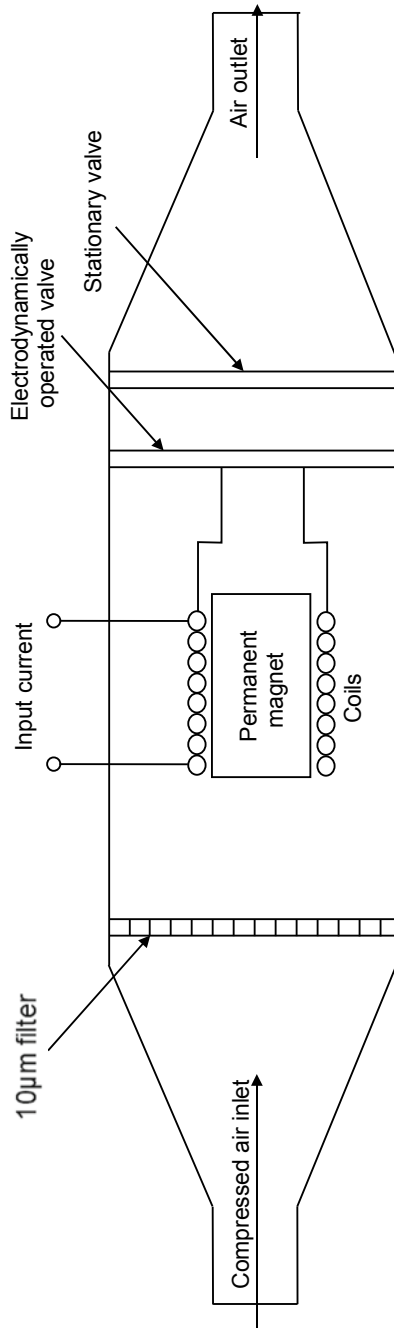
Fig. 5.12 Electropneumatic transducer
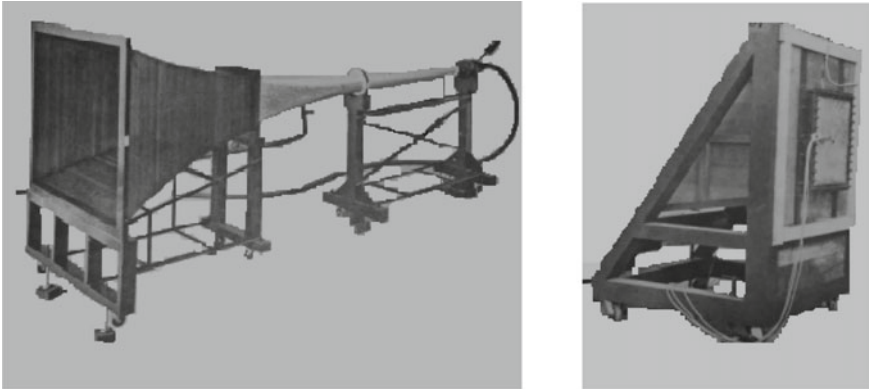
Fig. 5.13 Acoustic excitation using EPT: **a** EPT with horn, **b** test frame with test plate (Courtesy of IIT Madras)

direct air flow through an air filter to the pneumatic valve. This electrodynamically operated pneumatic valve responds to any applied electrical waveform and controls or modulates the velocity of the air stream supplied from a compressor. The resulting velocity fluctuations superimposed upon the direct flow velocity generate acoustic energy. This sound output is generally sent out through a horn which concentrates the sound into a directed beam so that most of the energy is radiated in one direction. The basic elements of the EPT are shown in Fig. 5.12. Air is delivered at constant pressure from a pneumatic compressor to the plenum chamber from where it escapes through one or more orifices. The orifice areas, which are controlled by the action of the electrodynamic driver, modulate the air stream velocity and produce acoustic energy. Figure 5.13a shows the EPT and horn assembly. Figure 5.13b shows a test plate (mounted on a test frame) to be subjected to acoustic excitation.

Advantages A very large amount of acoustic power is obtained from one source; no dynamic speaker can deliver this amount of power.

Disadvantages The transducer needs a steady supply of compressed air.

Specifications

Rated acoustic power: 4000 W.
Maximum air flow rate: 12.7 m^3/min.
Maximum required air pressure: 2.76×10^5 Pa.
Maximum current: 6 A (RMS).
Nominal impedance: 2 Ω 100 Hz.
Nominal diameter: 21.5 cm.

5.4 Sound Level Measurement

INTERESTING FACTS: Did you know that the first sound level meter produced in 1917 by AT&T and its many successors were physically huge? In fact, these devices were so cumbersome that in 1988, Leo Beranek in his classic book, 'Acoustical Measurements' described their essential accessories as 'a strong back or a rolling table'. By today's standards, their accuracy was questionable. To understand how the need for sound level meters arose, it is essential to know how a technician from the New York Noise Abatement Commission made noise level measurements in Times Square, New York, in 1929. He made measurements in different parts of the city to understand traffic noise. He tried to quantify 'deafness produced by noise', using a receiver, a microphone, a phonograph and a wobble oscillator. In 1932, the objective type of SLM was developed; it was an assembly of a microphone, an amplifier and an indicating instrument. It was very bulky since it used vacuum tubes and automobile type storage batteries. One of the earlier SLMs had a mass of almost 50 kg. Besides, there was no kind of standardization in those days. All types of microphones were used, right from the crudest carbon microphone to the most sophisticated condenser microphone of the time; meters were of the peak reading type, vacuum tube voltmeters, average type meters and RMS meters. Lightweight acoustic measuring instruments evolved rapidly after the transistor was developed in 1947 by John Bardeen, Walter Brattain and William Shockley. Today we have very sleek portable SLMs which are a few 100s of grams, which have internal memory, excellent post-processing software and alternative ways of displaying results.

5.4.1 Sound Level Meter

Sound pressure has always been the acoustic parameter of measurement interest since it is this quantity that the human ear detects. Microphones have and are being widely used for the measurement of sound pressure in spite of the fact that modern technology has made it possible for sound intensity and particle velocity measurements to be made quite easily. The sound level meter is the basic instrument designed for the measurement of sound, both objectively as well as subjectively, in approximately the same way as the human ear responds. It has already been mentioned in Chap. 4 that the decibel (dB) is chosen as the scale for sound pressure level measurement since this logarithmic scale compresses the entire range of sound pressures of practical interest in audio-acoustics from 20 μPa, which is the threshold of audibility, to 20–

63 Pa, which is the threshold of pain (when the source is radiating at a frequency of 1 kHz). Sound pressure level depends on the distance and direction with respect to the source, as well as the nature of acoustical environment, such as free-field or reverberant- or diffuse-field.

The chief features of the SLM are:

- Adjustable preamplifier.
- Frequency weighting networks: Linear, A, B, C and D.
- Internal 1/1 and 1/3 octave filters or an output to external filters.
- Main amplifier.
- Detector circuit.
- Indicating meter having 'fast', 'slow', 'peak' and 'impulse' responses.

It is required that the acoustical pressure fluctuations be measured with a suitable electro-acoustic transducer which most faithfully converts the sound pressure into voltage. Hence, the microphone constitutes the front end of any SLM. The most suitable microphone for the SLM is the condenser microphone. The electrical output of the microphone being quite small, it is amplified by a preamplifier before being processed. In most SLMs, the microphone is attached directly to the instrument. In others, the microphone and its preamplifier constitute a separate unit with a shielded cable connecting it to the instrument proper. The SLM itself is usually shaped to avoid disturbances due to its own reflection of sound on the microphone. The effect of the operator holding the meter should also be carefully considered.

The indicating meter (with a logarithmic scale in dB) usually meets the need for subjective and objective measurements of noise. The former measurement simulates human binaural perception of sound quantitatively from a source, for which use of internationally standardized A, B, C, and D frequency response weighting networks is made; these impart the characteristics of the human ear to the measuring instrument. The objective measures are indicative of the true sound from sources, correlating well with their vibrations; for this the Linear weighting network is used. The SLM is designed to cater to both these needs in the field with portability and reliability. Some SLMs come with built-in 1/1 and 1/3 octave analysers for a simple form of spectrum analysis.

Other optional features include a calibrated logarithmic attenuator, as well as peak and impulse reading capabilities using a detector circuit to capture the properties of the fluctuating sounds. The details regarding the detector circuit for time averaging are given in Sect. 5.4.4. The last stage of an SLM is the read-out unit which displays the SPL in dB (LIN) or dB (A). Most modern SLMs have digital displays that indicate the maximum RMS value measured using a rectifier. The signal may also be available at output sockets, in either AC or DC form, for connection to external instruments such as tape recorders or analysers, to allow recording and/or further processing. The piston phone or acoustic calibrator is also a standard accessory for an SLM. The essential components of the SLM are indicated in the block diagram in Fig. 5.14. Figure 5.15 shows a photograph of the meter.
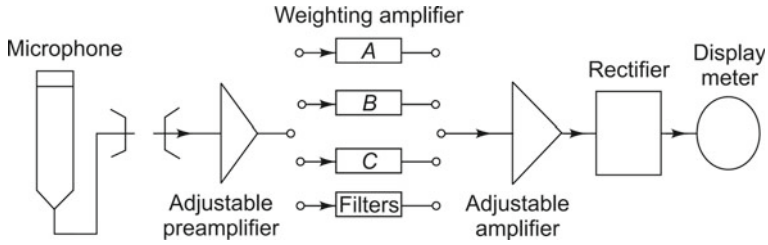
Fig. 5.14 Sound level meter

Fig. 5.15 Photograph of sound level meter (Courtesy of Brüel & Kjær, Denmark)

5.4.2 A, B, C, D and LIN Frequency Weighting Networks

There are generally four types of weighting networks, which adjust the response of the instrument to simulate the response of the human ear, which is logarithmic in its response to both loudness and frequency. These networks fall within the limits specified by American National Standards Institute (ANSI) or International Electrotechnical Commission (IEC). The function of the A-weighting network is to mimic the response of the human ear. Its characteristic is based on the historical equal-loudness contours and is matched reasonably well to human hearing sensitivity for a wide range of sounds. This weighting is stipulated for virtually all governmental and industrial regulations, as well as noise exposure studies and is also frequently

used for environmental noise measurements. It has a good correlation with hearing damage. Originally, the A-weighting was meant only for quiet sounds with SPLs in the vicinity of 40 dB but has now been made mandatory for all levels.

The linear weighting or LIN scale gives an overall value as would be sensed with a microphone having a flat frequency response in the audio-acoustic frequency range. This weighting network that provides a nominally uniform response up to 20 kHz is used for the objective technical evaluation of noise from sources. The B and C scales are meant to be used when the noise is louder than that which would be measured with an A scale. C frequency weighting is meant for describing the loudness of industrial noise and is used for the measurement of the peak value of a noise in some legislations; it is fairly 'flat', with only a small attenuation at both low and high frequencies. Some of the earlier instruments also had the D network for the evaluation of air traffic noise. This network had a gain about 10 dB higher than that of the A scale in the 3–4 kHz region. It may be said that the B and D scales are almost obsolete now. The frequency responses of the A, B, C, D and LIN weighting networks are shown in Fig. 5.16.

The weighted SLM gives only a single number reading for sound pressure level. It does not give any information regarding the frequency content of the signal. However, it can be used to find out whether a noise signal contains frequencies primarily above 1 kHz. To do this, measurements should be made using both the A- and C-weighted networks. The A-weighted sound pressure level in dB (A) should be subtracted from the C-weighted sound pressure level in dB (C); if the difference is large, the signal is primarily composed of frequencies less than 1000 Hz, if it is small, it consists of frequencies above 1000 Hz.
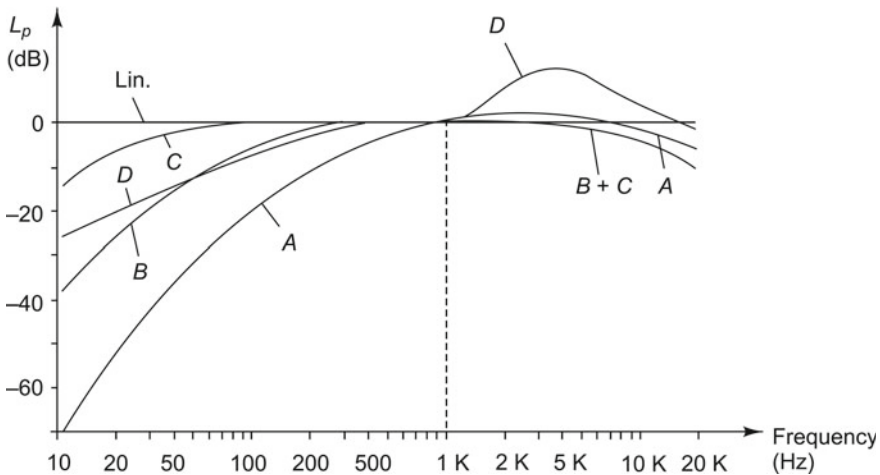

Fig. 5.16 Response of weighting networks

5.4.3   1/1 and 1/3 Octave Filters

For vibration measurements, constant bandwidth narrow band analysers may be used. For acoustic measurements, on the other hand, the entire frequency range from 20 to 20,000 Hz, i.e., the audio frequency range is divided into octave or one-third octave bands. The SLM is coupled to 1/1 or 1/3 octave analysers to obtain the spectral distribution of the sound energy. Such analysers are constant percentage bandwidth analysers and are used for analysis of acoustic signals, the reason being that the ear responds to frequencies in a logarithmic fashion. The advantage of constant percentage bandwidth filters is that two neighbouring filters combine to form one filter with a flat top, but with double the width (on a logarithmic scale). The octave bands cover the range of frequencies 20 Hz to 20 kHz with a number of 1/1 octave band pass filters. The term 'octaving' means doubling and each filter has a centre frequency and bandwidth twice that of the previous one. The 1/1 filter is the widest amongst octave filters used, with a bandwidth of 1 octave. But generally a higher frequency resolution is desired for spectrum analysis of a noise signal than that afforded by this weighting network. Such a refinement can be achieved by using subdivisions of smaller bandwidths. For example, in a 1/3 octave band analyser, each 1/1 octave filter is divided into three filters with the centre frequency of the central filter being the same as that of the original 1/1 octave band pass filter. The 1/1 and 1/3 octave filters are called constant percentage bandwidth filters, with the 1/1 octave filter having a bandwidth close to 70% of its centre frequency and the 1/3 octave filter, 23% of its centre frequency. The 1/3 octave filters are the more popular of the two, their advantage being that their bandwidth at frequencies 500 Hz corresponds well to the frequency selectivity of the human auditory system.

When a set of 1/1 or 1/3 octave band pass filters are used to analyse a noise signal, only the sound energy with frequency components contained in the filter pass band of each filter is allowed to pass through the respective filters. Such filters are sufficient if the energy in the sound signal is fairly evenly distributed and relatively flat over a rather broad range of frequencies. However, often a signal may contain significant amounts of energy at one or more discrete frequencies. Clues to this can be obtained by noting whether the SPL in a particular 1/1 or 1/3 octave band is much higher than that in the others, or whether a signal sounds as if it contains some pure tones. When this is the case, it may be necessary to conduct a more refined analysis of the signal to determine the frequencies at which these high energy levels exist. This can be accomplished by obtaining the spectrum levels of the sound signal using constant bandwidth filters with narrow bandwidths.

If f_c is the centre frequency of a bandpass filter, f_u the upper frequency limit and f_l the lower frequency limit, then the relationships as given in Table 5.2 exist for 1/1 and 1/3 octave band filters. Table 5.3 shows the lower, upper and centre frequencies of 1/1 and 1/3 octave band filters. It is seen from these tables that the bandwidths steadily widen as the centre frequencies increase. Figure 5.17 shows examples of 1/1 and 1/3 octave pass bands.

Table 5.2 1/1 and 1/3 octave band-filter frequencies

Diaphragm	1/1 octave band pass filter	1/3 octave band pass filter
Upper cut off frequency f_u	$f_u = f_c \times 2^{1/2} = 2 f_l$	$f_u = f_c \times 2^{1/6} = 2^{1/3} f_l$
Lower cut off frequency f_l	$f_l = f_c \times 2^{-1/2}$	$f_l = f_c \times 2^{-1/6}$
Centre frequency of next band	Present $f_c \times 2$	Present $f_c \times 2^{1/3}$

Table 5.3 Preferred band limits and centre frequencies (Hz) for 1/1 and 1/3 octave bands

Band	1/1 Octave bands			1/3 Octave bands		
	f_l	f_c	f_u	f_l	f_c	f_u
12	11	16	22	14.1	16	17.8
13				17.8	20	22.4
14				22.4	25	28.2
15	22	31.75	44	28.2	31.5	35.5
16				35.5	40	44.7
17				44.7	50	56.2
18	44	63	88	56.2	63	70.8
19				70.8	80	89.1
20				89.1	100	112
21	88	125	177	112	125	141
22				141	160	178
23				178	200	224
24	177	250	355	224	250	282
25				282	315	355
26				355	400	447
27	355	500	710	447	500	562
28				562	630	708
29				708	800	891
30	710	1000	1420	891	1000	1122
31				1122	1250	1413
32				1413	1600	1778
33	1420	2000	2840	1778	2000	2239
34				2239	2500	2818
35				2818	3150	3548
36	2840	4000	5680	3548	4000	4467
37				4467	5000	5623
38				5623	6300	7079
39	5680	8000	11,360	7079	8000	8913
40				8913	10,000	11,220
41				11,220	12,500	14,130
42	11,360	16,000	22,720	14,130	16,000	17,780
43				17,780	20,000	22,390
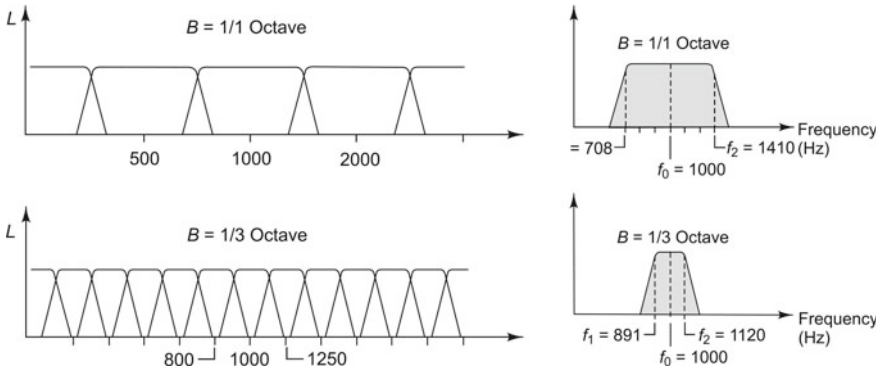
Fig. 5.17 Octave and third-octave filters (Courtesy of http://www.bksv.com)

Table 5.4 Frequency ratios for narrow band octave filters

Centre frequency	Frequency ratio
Octave	2
One third octave (1/3)	1.25
One quarter octave (1/4)	1.1892
One fifth octave (1/5)	1.1486
One sixth octave (1/6)	1.1225
One eighth octave (1/8)	1.0905
One tenth octave (1/10)	1.0717

For more detailed analysis, very narrow band pass filters are available today. Their centre frequencies are obtained from the lowest centre frequency by multiplying them successively by $2^{1/n}$ where n may be equal to 3, 4, 5, etc. Filter bandwidths down to 1/96 octave have been realized. Table 5.4 shows the frequency ratios for typical narrow bands commencing from 31.5 Hz.

5.4.3.1 Spectrum Level

The spectrum level of a sound signal is defined as the sound level contained in a filter with a frequency band pass of 1 Hz centred at a frequency f. The spectrum level is usually measured with constant band pass, narrow band frequency analysers. With respect to the sound pressure level, if the spectrum level $L_{P(SL)}$ of the signal is known, the band sound pressure level $L_{P(BL)}$ over a specified band (a measure of the total energy contained in the band) is obtained as

$$L_{P(BL)} = 10 \log_{10} \left[\int_{fl}^{fu} \frac{p_{rms}^2(f)}{p_{ref}^2} df \right] \tag{5.5}$$

If the sound energy is fairly flat and evenly distributed over the frequency band of interest, the above equation can be simplified as

$$L_{P(BL)} = 10 \log_{10} \left[\frac{p^2_{\text{rms(avg)}} \Delta f}{p^2_{\text{ref}}} \right] \tag{5.6}$$

where $p^2_{\text{rms(avg)}}$ is the average mean square value of sound pressure in the bandwidth Δf. The band sound pressure level can be further simplified as

$$L_{P(BL)} = L_{P(SL)} + 10 \log_{10} \Delta f \tag{5.7}$$

5.4.4   Time Averaging

The sound level value is displayed as a reading on an indicating meter or as a digital value, after exponential time averaging of the signal. If the sound level fluctuates too fast, it is not possible to get a meaningful reading. For this reason, two detector response characteristics were standardized: the F (for Fast) and S (for Slow) characteristics. F typically has a time constant of 125 ms and provides a fast reacting display response. S has a time constant of 1 s, giving a slower response which helps average out the display fluctuations, which would otherwise be impossible to read using F mode. Besides, there are two more forms of averaging: I for impulse and P for Peak hold. Figure 5.18 shows the detector characteristics in these various modes of operation.

Impulsive noise consists of one or more bursts of sound energy, each of duration less than about 1 s. Impulses are characterized by the peak pressure, the rise time
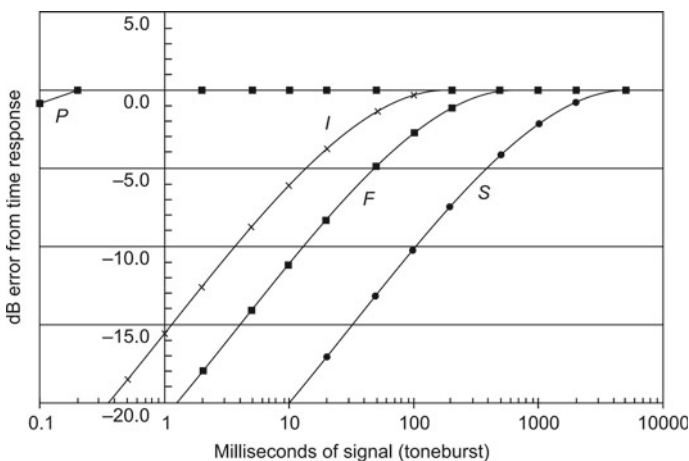

Fig. 5.18   Characteristics of detector circuit

and the duration of the peak. The *I* or impulse mode is to be used if the sound to be measured consists of isolated impulses or a large amount of impact noise; under such circumstances, the normal *F* and *S* detector responses are not short enough to detect and display transient noise in a manner similar to the human perception of impulsive sounds. The *I* characteristic has a time constant of 35 ms. The perceived loudness of short duration sounds is typically lower than that of steady continuous sounds, but the risk of damage to hearing is not necessarily lower. For this reason, some SLMs include a circuit for capturing either the peak value of a sound or its maximum RMS value, irrespective of its duration during the monitoring period. This is done by using a hold circuit. Time constant for the peak detector is $50\,\mu s$.

5.4.5   Integrating Sound Level Meters

The hearing damage potential of a given sound environment depends not only on its level, but also on its duration. Therefore, both these quantities must be measured to indicate the energy received. For constant sound levels, this is easy, but if the sound level keeps varying, a single integrated value computed over a well- defined sampling period should be used. For the measurement of fluctuating noise, such as traffic noise or for community noise-annoyance assessments, integrating SLMs are available to measure energy equivalent level L_{eq}, which is nothing but the equivalent continuous sound level over a prescribed time. In the absence of any uniform method for determining community reaction to intruding noise, equivalent continuous noise level L_{eq} has become the basic measure. For an A-weighted L_{eq}, the symbol L_{eq}, is used and may be computed according to the equation shown below.

$$L_{Aeq} = 10 \log_{10} \frac{1}{T} \int_0^T \left[\frac{p_A(t)}{p_{\text{ref}}(t)} \right]^2 dt \qquad (5.8a)$$

where T is the total measurement time, $p_A(t)$ is the *A*-weighted continuous acoustic pressure and p_{ref} is the reference acoustic pressure ($20\,\mu$Pa). This L_{eq} has the same energy content and consequently the same hearing damage potential as the fluctuating sound. It is customary to calculate a noise exposure level normalized to a nominal 8 h working day. When the *A*-weighted SPLs over a period of time are presented in the form of a histogram, L_{eq} is derived from the following equation

$$L_{eq} = 10 \log_{10} \left[\frac{1}{100} \left\{ f_1 10^{L_1/10} + f_2 10^{L_2/10} + f_i 10^{L_i/10} + \cdots + f_n 10^{L_n/10} \right\} \right]$$
$$(5.8b)$$

Here f_i denotes the percentage of the total time for which the *A*-weighted noise levels lie within the band interval with the corresponding centre frequency and L_i is the *A*-weighted level.

L_{eq} is the basis for calculating L_{DN}, the day-night average sound level and L_{NP}, the noise pollution level. Day-night average (DNL) or L_{DN} is defined as the *A*-weighted

equivalent sound pressure level expressed in dB averaged over a period of 1 day
or 24 h, with an additional level of 10 dB imposed on the equivalent sound levels
at night, i.e., between 10 p.m. and 7 a.m, to account for increased annoyance due
to noise during the night hours. This average is used to define the level of noise a
community is exposed to and for aviation noise analysis.

The noise pollution level (NPL) or L_{NP} is related to L_{eq} and was developed by
the U.S. Department of Housing and Urban development (HUD) as a guideline for
all types of community and environmental noise sources in U.S.A. L_{NP} is based
on the evidence that human annoyance for a specific noise exposure is a function
of the average level, as well as the variability of the sound source (assuming that
the less steady it is, the more distracting and annoying it becomes). L_{NP} combines
the ambient noise level with the fluctuations of the noise over time leading to the
definition:

$$L_{NP} = L_{eq} + k_0 \sigma \tag{5.9}$$

where L_{eq} is measured in dB (A), σ is the standard deviation of the instantaneous A-
weighted SPLs with time, k_0 is a constant which is assigned a value 2.56, derived using
the best fit to data from a number of studies of subjective response to variable noise
levels. This measurement system applies to any environment, unlike those specifi-
cally concerned with aircraft/road traffic. Table 5.5 gives some guidelines regarding
acceptable values of LNP. The variance may be calculated as

$$\sigma^2 = \sum t_j L_j^2 - \left(\sum t_j L_j \right)^2 \tag{5.10}$$

where t_j are the fractions of time L_j occur during 24 h.

5.4.6 Acoustic Calibrators

This section explains techniques for the calibration of microphones. The calibra-
tion can either be performed in the field or in a calibration laboratory. Calibration
essentially means determination of the sensitivity of the device which is generally
expressed in terms of mV/Pa. The method used to calibrate a microphone must have
a known amount of uncertainty. Calibration is important for the measurement to be
recognized by legal authorities or, if compliance with international standards is to be

Table 5.5 LNP criteria proposed by HUD

Category	L_{NP}, dB (A)
Clearly acceptable	<62
Normally acceptable	62–74
Normally unacceptable	74–88
Clearly unacceptable	>88

claimed. For absolute confidence in the results, it is advisable to perform a field calibration before and after a measurement, irrespective of whether the measurements are absolute or relative.

Most sound level calibrators are portable, easy-to-use and characterized by the production of a well-defined sound pressure at a single frequency, usually in the range 200 Hz–1 kHz, at which the calibration is performed (IEC 942 'sound calibrators'). To calibrate across the entire frequency range, a multitone calibrator may be used. Such a calibrator provides a number of pure tones at single frequencies in steps of one octave. When using calibrators that produce a single frequency, the calibration is strictly valid only at that reference frequency. However, considering the fact that microphones generally have a flat frequency response over a very wide range of frequencies, it can be expected that they will give the same electrical output at all frequencies in the specified frequency range for sound pressures of equal magnitude, making calibration at a single frequency sufficient.

5.4.6.1 Piston Phone

A piston phone is a calibrated reference sound source, which provides a quick and simple means of calibration in the field or laboratory conditions for calibration of precision condenser microphones and high sensitivity piezoelectric pressure sensors. It is a self-contained, easy to operate, battery-powered instrument which generates a very stable and accurate reference SPL for calibration by the to and fro movement of a piston in a coupler cavity. The reference SPL is generated by varying the volume of the cavity at a controlled frequency by a known displacement, and is proportional to the variation of the coupler volume, i.e., to the stroke and area of the piston. The cavity volume is that which is enclosed at the head of the piston phone by the transducer and its dedicated coupling adaptor which ensures that an equal cavity volume is maintained irrespective of the sensor used. Two opposed reciprocating pistons that are driven by a precision-machined, rotating cam disk alter the cavity volume by a known displacement at a tachometer-controlled frequency. A pressure change accompanies the change in volume of the cavity, resulting in the accurate reference SPL. The ambient barometric pressure, temperature, and humidity influence this SPL. A barometer is normally included to provide a correction for the ambient pressure. The error in SPL is a function of the error in the measurement of the stroke. The piston phone allows correct calibration to be made even in very noisy surroundings.

When such a calibrator is used for microphone calibration, the well-defined sound pressure output from the calibrator is applied to the front of the microphone and the sensitivity may be determined by dividing the output voltage by the sound pressure. When the calibrator is used to calibrate the entire measurement channel or sound level meter, the sound pressure is applied to the front of the microphone as before and a gain adjustment is made in the measurement system or sound level meter to give the correct reading. The sound calibrator always establishes a pressure field and hence a suitable correction has to be applied for a free-field calibration. It is to be noted that when calibrating a sound level meter, the weighting filter should not be

Fig. 5.19 Photographs of calibrators (Courtesy of Brüel & Kjær, Denmark)

switched on. When using a piston phone with a 250 Hz calibration frequency, a signal reading which is 8.6 dB less than the actual value is got if the A weighting filter is switched in. Please see Sect. 5.4.2 for information on weighting filters. Figure 5.19 shows photographs of some calibrators.

Typical specifications

Generated SPL: 134 dB SPL ($\pm$0.09 dB) (100 Pa)
Reference conditions: ambient pressure: 101.3 kPa, ambient temperature: 20 $^\circ$ C, ambient humidity: 65% RH.
Frequency: 250 Hz ($\pm$0.5%).
Distortion: $<$1.5%.
Coupler volume: 20 cm^3.

5.4.6.2 Electrostatic Actuator Calibration

The electrostatic actuator is a reliable device for determination of a microphone's frequency response under laboratory conditions. It is applicable to microphones with a metallic diaphragm. The actuator comprises a stiff and electrically conducting metal plate which forms an electrical capacitor along with the microphone diaphragm. The metallic grid is positioned close to (at approximately 0.5 mm of) the diaphragm of the microphone. By applying high voltages to the actuator, electrostatic forces equivalent to a pressure-field or sound pressure of approximately 100 dB are produced on the microphone diaphragm. The actuator method has an edge over sound based methods in that it provides a simpler means of producing a well-defined calibration pressure over a wide frequency range, but its accuracy is not high enough for the determination

of the microphone sensitivity, which has to be done with an acoustic calibrator/piston phone at a reference frequency. This method cannot simulate a free-field or a diffuse-field condition and appropriate corrections must be made if such calibrations are required. The electrostatic force or pressure produced by the actuator is practically independent of environmental factors.

The actuator is not suited for absolute calibration since the sound pressure produced is heavily dependent on the distance between the actuator and the diaphragm. Besides, this method requires the system to use high voltages and a removeable grid.

5.5 Acoustic Chambers

The most commonly used acoustic chambers are anechoic and reverberation chambers.

DID YOU KNOW that the world's oldest wedge-based anechoic chamber is the one built by Bell Laboratories in 1940 and is called The Murray Hill anechoic chamber? The interior room measured approximately 9.144 m high, 8.534 m wide and 9.754 m deep, with cement and brick walls about 0.9144 m thick to attenuate outside noise. Large fibre glass wedge-shaped absorbers were fixed on the interior surfaces of the chamber to create an impedance match with the surrounding air. Wedges serve as waveguides with all incident sound energy getting internally reflected into them. This chamber was once considered the world's quietest room and was used to measure loudspeaker and microphone directivity and frequency response functions (FRFs). The development of reverberation chambers was not driven by the necessity to conduct acoustic experiments, but by the need to record rich classical music. Buildings such as churches, church halls, and ballrooms were often chosen for this purpose because of their natural reverberation characteristics. Artificial echo/reverberation chambers had to be designed in the early 1900s for sound recording due to limitations of early recording systems. Most commercial popular recordings were made in specially constructed artificial echo chambers. The echo chamber at EMI's Abbey Road in London was one of the first studios in the world to be built specifically for recording purposes in 1931; to this day, it remains an important example of the early twentieth-century electro-acoustic echo chambers.
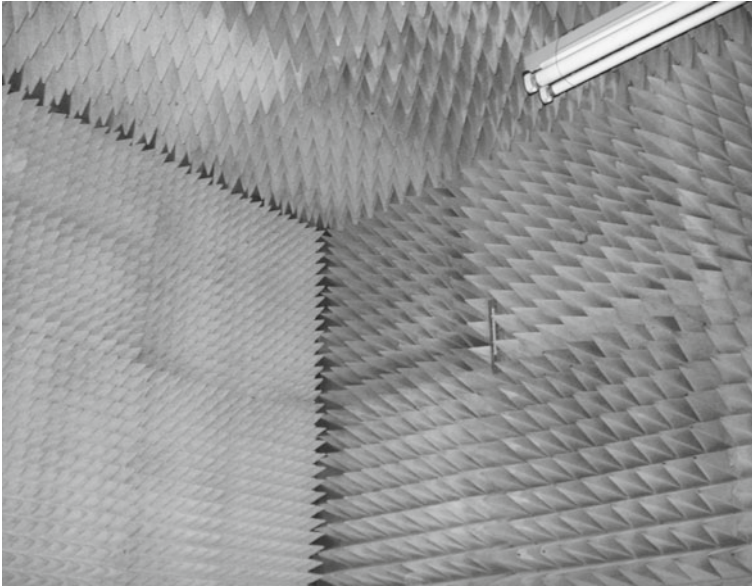
5.5.1 Anechoic Chamber

The word *anechoic* literally means echoless. Anechoic chambers are echo-free enclosures with sound absorption of 99% or more, or a reflection of 10% or less. In practice, anechoic chambers are designed, not just as echoless rooms, but also as rooms which are free from external noise. These rooms are used in the laboratory to create a free-field condition, which implies that the sound level from a spherically radiating sound source in the chamber should decrease by 6 dB for every doubling of distance from the source. Also, there should be no external noise interference and no excessive variations in temperature, pressure and humidity.

To achieve these conditions, the rooms have to be large with massive walls and ceiling to isolate airborne and structure-borne sounds from outside. Anechoic chambers must have very low background noise and this consideration calls for superior sound insulation and isolation. The boundary surfaces of the chamber (walls, floor and ceiling) are covered with highly absorbing materials with normal incidence absorption co-efficient not less then 99% throughout the frequency range of interest. These rooms are often constructed as a vibration isolated room within another room. Since no sound can penetrate into such rooms, any sound produced inside the chamber can be studied for its intensity, frequency content, audibility, etc. Anechoic chambers are essentially used for research in acoustics and development work in fundamental research of speech and hearing. They are also used for studying sound patterns of loudspeakers, microphones and of consumer goods like washing machines, refrigerators, air-conditioners, automobile engines, etc., from the point of view of product comparison or noise control. Though the noise power of a machine can be calculated from sound intensity measurements made at site over a geometrical surface enclosing the machine, many quality standards are still based on the use of anechoic chambers.

The sound-absorbing materials typically used are porous materials like special acoustic foam, shaped in the form of wedges or perforated sheet metal wedges, filled with fibreglass insulation or acoustic tiles. The shape and dimensions of the acoustic foam wedges are dependent on the performance level required and cut-off frequency desired. The larger the lengths of the wedges, the lower are the frequencies they are capable of absorbing. The lowest frequency at which the sound energy absorption falls below 99% or the reflected sound pressure level exceeds 10% is the cut-off frequency and the corresponding wavelength, the cut-off wavelength. Fig. 5.20 is a photograph of an anechoic room, clearly showing the wedges.

Anechoic chambers should provide an ambient noise level of at least 20 dB, preferably 25 dB, below the lowest level of the signal to be measured. Their behaviour should adhere to the inverse square law variation of sound intensity under free-field conditions. Consequently, the sound pressure level should decrease by 6 dB with every doubling of distance from the source. Under practical conditions, deviations of up to 1.5 dB are not uncommon in a fully anechoic room as per ISO 3745. Anechoic chambers are classified as: (a) fully anechoic: all six surfaces are sound absorbent and (b) hemi-anechoic: the floor is a normal reflecting plane.

Fig. 5.20 Walls of anechoic chamber (Courtesy of IIT Madras)

5.5.2 Reverberation Chambers

A reverberation chamber has hard surfaces that reflect sound and is used to simulate a diffuse field in the laboratory. It is designed such that no two walls are parallel, ensuring that multiple reflections take place at various angles of incidence to produce a non-directional or 'diffuse' sound field within the chamber. The sound power emitted by a machine can then be found from the SPL in the chamber knowing the characteristics of the room such as its volume, surface area and reverberation time (RT). There are several design and construction factors to be considered in the construction of a reverberation chamber. The important ones are the cut-off frequency, inside and outside dimensions, exterior ambient and target interior noise levels, vibration isolation target and its means of achievement, shape, including non-parallel inside walls to avoid standing waves and non-porous walls to have high reflection. A very simple design could be in the form of twin parallelepiped pentagons (Fig. 5.21) such that no two walls are parallel, thus achieving multiple reflections, so as to achieve a near diffuse-field. The floor of the chambers is kept horizontal, while the ceiling of the chambers is slanting so that they are not parallel. Generally, the chambers are designed in two different sizes, one being a large reverberation chamber and the other being a smaller one. The difference between the volumes of the chambers should be more than 10% as per the ISO 140-1, ISO 140-3 standards.

Fig. 5.21 Isometric view of
twin reverberation chambers
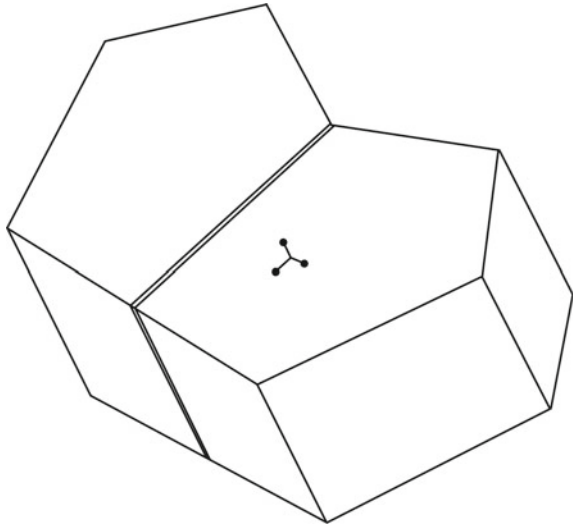
5.5.2.1   Methods to Achieve Diffusivity

Several methods have been used in order to achieve a relatively high degree of
diffusivity in reverberation chambers. Some of the more important ones are discussed
further.

- **Rotating reflector:** A motorized rotating reflector placed in the room effec-
  tively creates a variable boundary of the room, resulting in oscillation and the
  appearance and disappearance of modal patterns for a constant sound source.
  The circulation of the air in the room also breaks the modal patterns to achieve
  a good diffuse sound field.
- **Reflecting surfaces:** Highly reflective surfaces are added to the rooms by hang-
  ing objects from the ceiling. This increases the sound reflecting area for a given
  room volume, without appreciably increasing the room absorption. The sound
  diffusion is thus enhanced through multiple reflections.
- **Non-parallel walls:** Such a construction enables the sound waves in the room
  to impinge on the walls randomly at various angles of incidence. For a constant
  volume, the total number of normal modes for an irregularly shaped room is
  almost that for a room with parallel walls. However, the degeneracy of modes
  due to the overlap of modal patterns in an irregularly shaped room is considerably
  less and, consequently a more diffuse sound field is obtained.
- **Multiple frequency sound source:** A signal such as a warble (frequency mod-
  ulated) tone or a third octave band noise has energy over a wider frequency
  range than a pure tone. Such a source is capable of generating a series of modal
  patterns, each catering to a narrowly grouped frequency band. The application
  of such a sound source is usually required but is unrelated to the design and
  construction of the chambers.

List of Symbols

Symbol	Meaning	SI unit
A	area of capacitor plates, surface area	m^2
B	flux density	T
c_0	speed of sound	m/s
C	capacitance	F
d	displacement of diaphragm from rest/electrode	m
D_0	distance between plates at rest	m
dBA	A-weighted SPL (dB)	
$dB_N(A)$	A-weighted noise level due to inherent noise of microphone (dB)	
d	diameter of the coil	m
D_0	distance between plates at rest position	m
e	change in voltage/ voltage induced in coil	V
E	instantaneous voltage between plates	V
f	frequency	Hz
f_c	centre frequency of band pass filter	Hz
f_i	percentage of total time for which A-weighted SPL is in ith band	
f_1	lower frequency of band pass filter	Hz
f_u	upper frequency of band pass filter	Hz
F	Force	N
G	microphone gain	
k	wavenumber	m^{-1}
k_0	constant	
l	length of conductor	m
L	distance between two sides of surface	m
L_{Aeq}	equivalent continuous A-weighted SPL over a prescribed time (dB)	
L_{DN}	day-night average SPL (dB)	
L_{eq}	equivalent continuous SPL over a prescribed time (dB)	
L_i	A-weighted SPL in ith band (dB)	
L_j	jth measured SPL (dB)	
L_{NP}	noise pollution level (dB)	
$L_{P(BL)}$	band sound pressure level (dB)	
$L_{P(SL)}$	spectrum level (dB)	
N	number of turns of coil	
$p_A(t)$	A-weighted continuous acoustic pressure	Pa
p_{ref}	reference acoustic pressure	Pa
$p(x, t)$	pressure	Pa
Q	charge	C
S_o	open-circuit sensitivity of microphone	V/Pa
S_c	loaded sensitivity of microphone	V/Pa

Symbol	Meaning	SI unit
t_j	fraction of time over which SPL L_j occurs	s
T	total measurement time	s
v	velocity	m/s
V	instantaneous voltage	V
V_0	polarization voltage	V
V_{sig}	signal voltage	V
Δf	frequency bandwidth	Hz
ε_0	permittivity of air	F/m
κ	dielectric constant	
λ	wavelength	m
θ	angle of incidence	rad
v	velocity	m/s
σ	deviation of instantaneous A-weighted SPLs (dB)	

Abbreviations

AC	alternating current
ADC	analogue-to-digital converter
ADP	ammonium dihydrogen phosphate
ANC	active noise cancellation
ANSI	American National Standards Institute
ASIC	application-specific integrated circuit
DC	direct current
ECM	electret capacitor microphone
EPT	electropneumatic transducer
ESL	electrostatic loudspeaker
IEC	International Electrotechnical Commision
ISO	International Organization for Standardization
kmph	kilometres per hour
LIN	linear
MEMS	micro electro mechanical systems
NPL	noise pollution level
RMS	root mean square
PET	polyethylene terephthalate
PZT	lead zirconate titanate
SLM	sound level meter
SNR	signal-to-noise ratio
TEDS	transducer electronic data sheet
THD	total harmonic distortion

Questions

1. What is a condenser microphone used for?
2. What are the applications of a microphone?
3. Microphones influence the sound field mainly at high/low frequencies. Strike out what is not applicable.
4. What is meant by a pressure field? Give an example of one.
5. What is the size of the condenser microphone preferred for frequencies above 20 kHz?
6. What are the parameters to be considered in the choice of a microphone?
7. How does one select a microphone based on the field being measured?
8. What is meant by frequency response of a microphone?
9. How does microphone sensitivity influence its size?
10. What decides the frequency response of a microphone?
11. What is the typical frequency response characteristic of a pressure field microphone?
12. When is a free-field microphone used? In what environments?
13. What is a diffuse-field microphone?
14. What do you mean by directivity of a microphone?
15. How does humidity affect a condenser microphone?
16. What is a carbon granule microphone? What are its drawbacks?
17. What information does a calibration chart of a microphone give?
18. What is the need for a windscreen in a microphone?
19. What is the nominal frequency range of a 1 in. condenser microphone?
20. What environmental factors should be considered in the choice of a microphone?
21. How does one select a microphone for measurement based on its diameter?
22. What do you mean by dynamic range of a microphone?
23. What is the use of a polarization voltage for a microphone?
24. Which is the most popular microphone for technical measurements?
25. What is meant by stability of a transducer?
26. How do you define sensitivity of a microphone?
27. On what factors does the sensitivity of a condenser microphone depend?
28. What is a pre-polarized microphone?
29. When does phase response of a microphone become important?
30. What is the need for static pressure equalization in a condenser microphone?
31. When is a probe microphone typically used?
32. What are phase-matched microphones? Where are they used?
33. How does the tension of a condenser microphone affect its sensitivity?
34. What microphones are used in environments with high humidity?
35. What is the working principle of a condenser microphone?
36. Where are surface microphones used? What are their disadvantages?
37. How does an electrodynamic microphone work?
38. What are the advantages of a piezoelectric microphone?
39. What is a ribbon microphone?

40. What kind of microphones are used in wearable devices?
41. What are the parameters that constitute the technical specifications of loudspeakers?
42. How does a moving coil loudspeaker work?
43. What is the typical construction of an electrostatic loudspeaker?
44. What are the advantages of ribbon driven speakers?
45. Where is an electropneumatic transducer typically used?
46. What is meant by a crystal microphone?
47. Name a microphone which has very little sensitivity to a magnetic field.
48. What are the main parts of a sound level meter?
49. What does a sound level meter measure?
50. What are LIN, A and C frequency weighting networks?
51. What is the need for moving coil loudspeakers with diaphragms of different sizes?
52. What is an electret microphone?
53. What is the significance of the A-weighting network?
54. What is the principle of operation of an electropneumatic transducer?
55. Why are many loudspeakers enclosed?
56. Which readings are higher, those measured on the LIN or A scale?
57. What are constant percentage bandwidth analysers?
58. What are the two common acoustic chambers?
59. What does 'octave' mean?
60. What are the limitations of a moving coil microphone?
61. What is an anechoic chamber? What are the two common constructions of anechoic chambers?
62. What are the advantages of an electrostatic speaker?
63. What is the advantage of a double-back plate type MEMS microphone?
64. Where are piezoelectric speakers used?
65. What is meant by band sound pressure level?
66. What is an integrating sound level meter?
67. What is meant by L_{eq}?
68. Why is time averaging used in sound level measurements?
69. What are the various detector settings available in a sound level meter?
70. What are the methods used to achieve diffusivity in a reverberation chamber?
71. What are the advantages of a MEMS microphone?
72. What are some of the factors to be considered in the construction of a reverberation chamber?
73. What is meant by piezoelectric effect? Name 2 mechanical quantities that can be measured using it.
74. What do you mean by free-field condition? How can you simulate it in the laboratory?
75. Name a device that works on the reverse principle of an electrodynamic microphone?
76. Differentiate between band pass filters used for objective and subjective measurements.

77. Why is a 1/1 octave filter called a % bandwidth filter? Name the centre frequencies of 1/1 octave band filters.
78. Why is the condenser microphone preferred to other microphones for engineering acoustics? What are its drawbacks?
79. What do you mean by subjective and objective measurements?
80. Differentiate between linear and logarithmic scales. When is each used?
81. What are the chief features of a reverberation chamber?
82. What are the upper and lower cutoff frequencies of a 1/1 octave filter of centre frequency f_c?
83. What is a hemi-anechoic chamber?
84. What are the common frequencies of acoustic calibrators?
85. What is a piston phone?
86. What are the sound-absorbing materials used in an anechoic chamber?
87. What is the centre frequency of the 1/3 octave filter next to that with centre frequency f_c?
88. What is the significance of root mean square value?
89. What purpose does an acoustic horn serve?
90. How do you measure the directivity index of a sound source?
91. What is the standard distance at which sound level measurements are made?
92. What is the order of reverberation time in a reverberation chamber?
93. What is the order of reverberation time in an anechoic chamber?
94. Why is a power amplifier to be used with an electrodynamic loudspeaker?
95. What is the specific characteristic of the A-weighting network?
96. Differentiate between constant bandwidth and percentage bandwidth frequency analyzers.
97. Why are percentage bandwidth analyzers preferred to constant bandwidth analyzers in acoustic measurements?

Fill in the Blanks

98. All precision measurement microphones are mostly _____ sensing condenser microphones.
99. The lower and upper cut-off frequencies of a 1/3 octave band pass filter of centre frequency f_c are _____.
100. The centre frequency of the 1/1 octave filter next to that with centre frequency f_c is _____.
101. The reference frequency for audio-acoustics is _____ because _____.
102. The % bandwidths of the centre frequency for a 1/1 octave filter is _____ and for a 1/3 octave filter is _____.

Bibliography

1. American National Standards Institute. (1984). ANSI S1.40-Specifications for Acoustic Cali-brators, R1997.
2. Barron, R. F. (2003). *Industrial noise control and acoustics*. Boca Raton, Florida: CRC Press.
3. Beranek, L. L. (1988). *Acoustic measurements*. New York: Wiley.
4. Beranek, L. L. (1993). *Acoustics*. Woodbridge, New York: Acoustical Society of America.
5. Beranek, L. L. (1993). *Acoustical measurements*. American Institute of Physics.
6. Bies, D. A., & Hansen, C. H. (2003). *Engineering noise control: Theory and practice*. New York, USA: Taylor & Francis.
7. Brüel & Kjær. (1996). *Microphone handbook, vol. 1, theory*, BE 1447-11. Denmark.
8. Brüel & Kjær. *Acoustic noise measurements*, BT 0010-12. Denmark.
9. Brüel & Kjær. *Frequency analysis of sound*, BA 7660-06. Denmark.
10. Brüel & Kjær. (1986). *Noise control - principles and practice*. Denmark.
11. Burns, W., & Robinson, D. W. (1970). *Fundamentals of acoustics hearing and noise in industry*. London: Her Majesty's Stationery Office.
12. Crocker, M. J. (1997). *Encyclopedia of acoustics*. New York: Wiley-Interscience.
13. Crocker, M. J. (1998). *Handbook of acoustics*. New York: Wiley.
14. Crocker, M. J. (2007). *Handbook of noise and vibration control*. Hoboken, New Jersey: Wiley.
15. Davis, D., & Davis, C. (1997). *Sound system engineering*. Boston, USA: Focal Press.
16. Diehl, G. M. (1973). *Machinery acoustics*. New York: Wiley.
17. Everest, F. A., & Pohlman, K. C. (2000). *Master handbook of acoustics*. New York: McGraw-Hill.
18. Fahy F. J. (1989). *Sound intensity*. London, (UK): Elsevier Applied Science.
19. Foreman, J. (1990). *Sound analysis and noise control*. New York: Van Nostrand Reinhold.
20. Garrett, S. L. (2017). *Understanding acoustics: An experimentalist's view of acoustics and vibration (graduate texts in physics)*. Springer.
21. Groves, I. D., Jr. (1981). *Acoustic transducers (benchmark papers in acoustics)*. Stroudsberg, Pennsylvania, USA: Hutchinson Ross Publishing Company.
22. Harris, C. (1998). *Handbook of acoustical measurements and noise control*. New York: LLC, Springer.
23. Miller, H. B. (Ed.). (1982). *Acoustical measurements: Methods and instrumentation (benchmark papers in acoustics)*. Hutchinson Ross Publishing Company.
24. Hemond, C. J. (1983). *Engineering acoustics and noise control*. Englewood Cliffs, New Jersey: Prentice Hall.
25. Huber, D. M., & Williams, P. (1998). *Professional microphone techniques*. Emeryville, CA: Mix Books.
26. International Electrotechnical Commission. (1988). IEC 942, Sound Calibrators, Class 1.
27. Kamichik, S. (1999). *Practical acoustics*. Edinburgh: Prompt Press.
28. Kinsler, L. E., & Frey, A. R. (1982). *Fundamentals of acoustics*. New York: Wiley.
29. Kinsler, L. E., Frey, A. R., Coppens, A. B., & Sanders, J. V. (2000). *Fundamentals of acoustics*. New York: Wiley.
30. Kleppe, J. A. (May, 1989). *Engineering applications of acoustics*. Boston, Massachusetts: Artech House Publishers.
31. Miyara, F. (2017). *Software-based acoustical measurements (modern acoustics and signal processing)*. Springer.
32. Moore, B. C. J. (2003). *An introduction to the psychology of hearing*. Orlando, Florida: Academic Press.
33. Peters, R. J., & Reeves, C. W. (1998). *The noise and acoustics handbook*. UK: Coxmoor Publishing Co.
34. Reynolds, D. D. (1981). *Engineering principles of acoustics: Noise and vibration*. Boston, USA: Allyn & Bacon.
35. Ristic, V. M. (1983). *Principles of acoustic devices*. New York: Wiley.

36. Rossi, M., & Roe, P. R. W. (1988). *Acoustics and electroacoustics*. Boston, Massachusetts: Artech House Publishers.
37. Sujatha, C. (2010). *Vibration and acoustics: Measurement and signal analysis*. Tata McGraw-Hill Education.
38. Talbot-Smith, M. (2001). *Audio engineer's reference book*. Boston, SA: Focal Press.
39. Wilson, C. E. (1994). *Noise control: Measurement, analysis and control of sound and vibration*. Malabar, Florida: Krieger Publishing Company.
40. http://www.bksv.com
41. http://www.hyperphysics.phy-astr.gsu.edu
42. https://edisontechcenter.org

Chapter 6
Theory of Stress Analysis

INTERESTING FACTS: The field of solid mechanics may be said to have developed after Newton's studies on the laws of motion, although it has much earlier roots. Leonardo da Vinci apparently had made some sketches regarding a possible test of the tensile strength of a wire. Galileo had studied the breaking loads of rods under tension and concluded that the load did not depend on the length and was proportional to the cross-sectional area. This was a first step towards a concept of stress. He also studied the breaking loads on beams suspended horizontally from a wall in which they were fixed. This subject has developed considerably due to contributions from outstanding investigators like Coulomb, Poisson, Navier, St. Venant, and Cauchy.

6.1 Introduction

There are numerous engineering applications involving analysis and design of machines, bridges, buildings, aircraft, and load bearing structures. Individual members of real-life structures are subjected to external forces due to operating or environmental conditions. Even when a member is in equilibrium because the resultant of the external forces is zero, there is a reaction in the form of internal forces within the material causing it to deform. During design, it is required to compute the deformations caused by the loads acting on a structure. It has to be ensured that the deformations are small enough to allow the structure to function as per its design. The behaviour of a member under load is decided not only by the fundamental laws

of Newtonian mechanics governing the equilibrium of the forces, but also by the mechanical characteristics of the material of which the member is made and these are got from experimental investigations.

The strength of a material depends on its ability to withstand forces without breaking or failing. It depends on many factors like microstructure, temperature, age, fluid content, etc. It also depends on the type of load, as well as its velocity and direction. When a material is loaded, it will undergo a change in shape. This change in shape or deformation is called strain (ε), which is typically given in relative units or percent change.

There are various ways in which a member can be loaded. They are: (i) static or dead load, due to its own weight; (ii) dynamic or time-varying loads such as fatigue, fluctuating, or alternating loads; and (iii) impact or shock loads caused by impulsive phenomena. Loads can also be classified based on the direction of loading. They are (i) axial compressive loading (Fig. 6.1a); (ii) tensile loading as shown in Fig. 6.1b, in which the forces act along the longitudinal axis of the bar; (iii) torsional loading as shown in Fig. 6.1c, in which the forces act in a tangential direction and give rise to moments; and (iv) bending in which the loads act perpendicular to the beam (Fig. 6.1d).

Fig. 6.1 Loading on bar: **a** compression, **b** tension, **c** torsion, and **d** bending
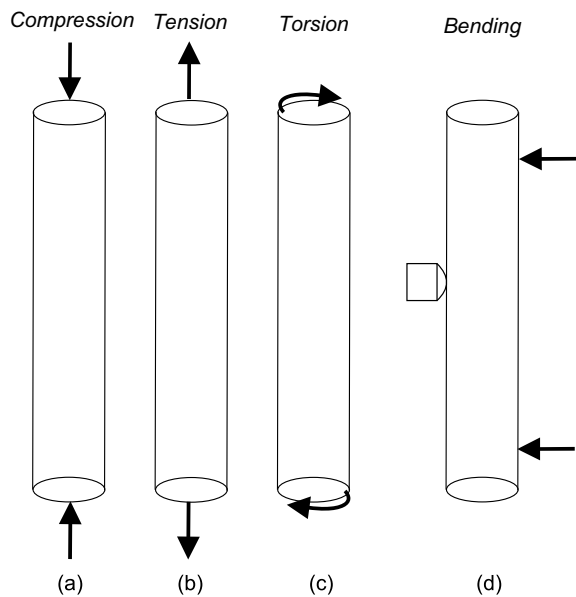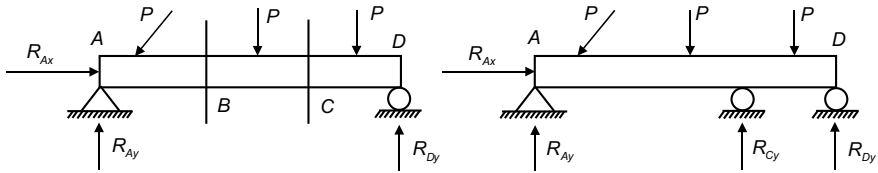
Fig. 6.2 a Statically determinate beam, **b** Statically indeterminate beam

6.1.1 Statically Determinate and Indeterminate Structures

For a three-dimensional body in space to be in equilibrium, the equations of statics require that the sum of all the forces acting on the body in the x-, y-, and z-directions be zero. Also the sum of all moments due to all the forces about any axis parallel to the x-, y-, or z-directions should be zero. These conditions can be expressed by the following equations:

$$\sum F_x = 0 \qquad \sum M_x = 0$$
$$\sum F_y = 0 \qquad \sum M_y = 0$$
$$\sum F_z = 0 \qquad \sum M_z = 0 \qquad (6.1)$$

The above equations can be applied to deformable solid bodies, the deformations typically being very small compared to the overall dimensions of the body. In order to obtain the forces in a member, its original undeformed dimensions are used for calculations. A statically determinate structure is one for which the external reactions and the internal stress resultants may be determined from the equations of statics alone. Figure 6.2 shows statically determinate and indeterminate structures. It can be seen that the number of independent equations in statics is sufficient to solve for the reactions in Fig. 6.2a, but is insufficient in Fig. 6.2b. For statically indeterminate structures, a knowledge of material properties is required for analysis of deformations. In general, a problem is statically indeterminate if a structure is held by more supports than are required to maintain equilibrium. This condition will lead to more unknown reaction forces than the number of equilibrium equations. The degree of indeterminacy is the number of unknown reactions minus the number of equations of equilibrium.

6.2 Axial Loading: Normal Stress and Strain

DID YOU KNOW that it was the English scientist Robert Hooke who first discovered the notion of linear elasticity, though not expressible in the form of stress and strain? It was the Swiss mathematician Jakob Bernoulli who

observed that the appropriate way of describing deformation of a material under
tension was to give force per unit area, or stress, as a function of the elongation
per unit length, or strain. The Swiss mathematician Leonhard Euler, in 1727,
proposed a linear function between stress σ and strain ε, as $\sigma = E\varepsilon$, where the
coefficient E is nowadays referred to as Young's modulus of elasticity after
the British scientist Thomas Young who put forth a related idea in 1807.

Axially loaded members are structural components subjected only to tension or
compression and are very common in civil engineering structures. Examples of such
members are truss members, bracings for buildings and bridges, cables in suspended
roof systems, connecting rods, towing ropes, shafts, columns, etc. Bungee cords
used for entertainment as well as support for free boundary condition in experimental
modal analysis are also subjected to axial load alone. Examples of compressive forces
are those in the support pillars of buildings or in the legs of a chair. One important
difference between tension and compression members is that the former are held
straight by means of tensile loads, while the compressive loads in the latter tend to
bend the member out of the plane of loading.

If a cylindrical rod is subjected to a compressive or tensile force P uniformly
applied across its cross-sectional area A as shown in Fig. 6.1a or b, then the internal
forces which are set up are also distributed uniformly across its cross section. The
internal resistance of the material is said to be direct or normal stress or axial stress
which is defined as

$$\text{Stress } \sigma = \frac{\text{Load}}{\text{Area}} = \frac{P}{A} \tag{6.2}$$

The SI unit of stress is Pa or N/m^2. In situations where the stress varies across any
given section, the stress at any point is given by the limiting value. Axial stress can
cause compression, elongation, buckling or failure in a member.

6.2.1 Stress–Strain Relation for Steel

The strength of a material and its stress–strain curve are directly related. Many mate-
rials, when subjected to axial loading, show a linear stress–strain relationship in the
elastic region of the stress–strain curve. The axial stiffness of the material in this
region is called Young's modulus or modulus of elasticity. For uniform or isotropic
materials such as a solid copper or aluminium rod having identical mechanical
properties in all directions, Young's modulus is constant. For anisotropic materi-
als, on the other hand, Young's modulus is not constant. Such materials exhibit a
different mechanical response when the direction of loading is changed. This is due
to impurities within the material or the direction of fibres within the material.

Figure 6.3 shows a stress–strain curve for steel. Here point A shows the propor-
tional limit. Beyond this limit, the relationship between tensile stress and strain is

Fig. 6.3 Stress–strain curve
for steel
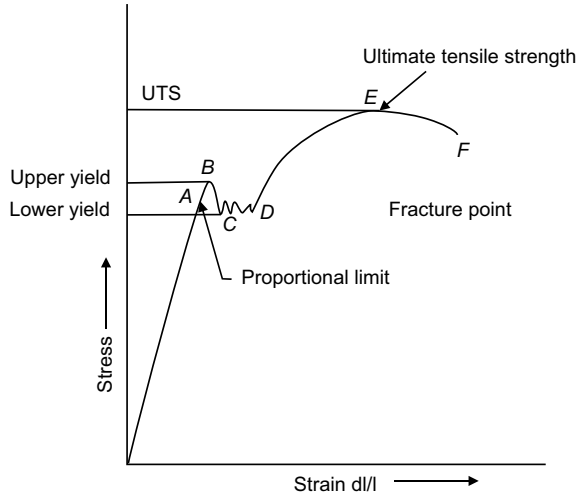
no longer proportional and becomes a little complicated. The maximum amount of
deformation that a material can withstand and still regain its original shape is called
the upper yield point and is depicted by point B, which is the end of the elastic region.
In this region, the material will deform during loading and return to its original shape
when unloaded. When the material is loaded beyond point B, some of the deforma-
tion becomes permanent and is non-reversible. This region of the stress–strain curve
is called the plastic region. At C, a sudden elongation is observed with hardly any
increase in the tensile force. The material is said to have yielded and point C is called
the lower yield point. At this point, a minimum load or stress is required to maintain
the plastic behaviour of the material. If the force is applied beyond point D, the
material recovers its resistance and the strain increases with the tensile stress up to
point E, where the stress reaches its maximum value called the ultimate strength of
the material. When loaded beyond this point, the strain increases with diminishing
stress and the material fractures when point *F* called fracture point is reached.

6.2.2   Different Types of Materials

6.2.2.1   Elastic Material

The elastic response of a material can be modelled by a spring. Figure 6.4 shows a
typical stress–strain curve for an elastic material. It may be observed that this curve
traces the same line for both the loading and unloading phases. When the spring
in this model is loaded, it will deform linearly with respect to the applied force.
The relation between the force *P* necessary to deform the spring and the resulting
deformation *x* is linear and is described by Hooke's law:

$$P = kx \tag{6.3}$$

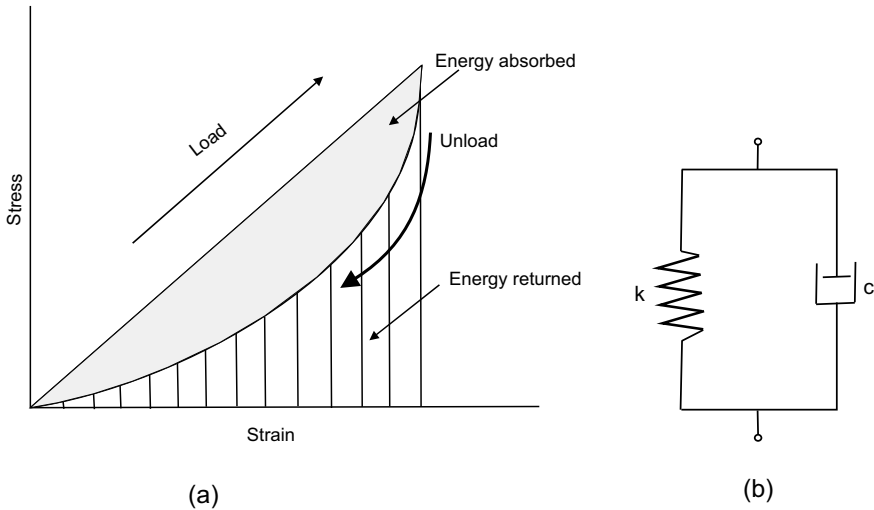**Fig. 6.4** Stress–strain
diagram of an elastic
material
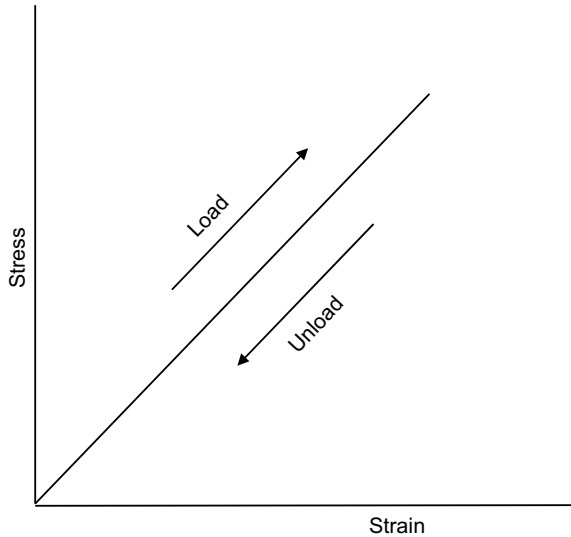
Here k is the spring constant or stiffness of the spring. The load and deflection are
in phase or, in other words, there is no delay of the resulting deflection with respect
to the applied force. When the load is removed, the spring returns the elastic energy
that was stored in it during deformation.

6.2.2.2 Viscoelastic Material

When an elastic material containing a fluid is loaded and unloaded, the material
returns to its original shape, but with a time delay. Such a material is called a vis-
coelastic material and exhibits the properties of both a viscous damper and an elastic
spring when loaded. A viscoelastic material absorbs some energy during deformation,
unlike a spring. Hence, the stress–strain curve for the unloading phase is different
from that for the loading phase. Energy is lost during the deformation of a viscoelas-
tic material primarily due to heat. Figure 6.5a shows the stress–strain curves for
the loading and unloading of the material. The amount of energy absorbed by the
material is given by the shaded area between the loading and unloading curves and
is called a hysteresis loop. The energy returned is the area below the unloading curve
and is shown by the vertical lines.

  A typical example of a viscoelastic element is a shock absorber in a car. Figure
6.5b shows the Kelvin–Voigt model of a spring and dashpot in parallel; such a model
is used to describe the behaviour of a viscoelastic material. The dashpot is a piston in
the fluid-filled cylinder and models the viscosity, which is a measure of the internal
resistance to the fluid flow. The viscosity of the fluid changes the damping provided
by the dashpot and causes a time delay in the mechanical response of the material with

Fig. 6.5 Viscoelastic material: **a** stress–strain curve and **b** Kelvin–Voigt model

respect to the applied force. The response of the damper depends on both velocity and temperature. A viscoelastic material is stiffer when loaded fast and less stiff when loaded slowly.

6.3 Pure Shear

INTERESTING FACTS: Shear stress is caused by forces parallel to the area resisting the force and is also known as tangential stress. It causes deformation of a material by slippage along a plane or planes parallel to the imposed stress. It is different from tensile and compressive stresses, which are caused by forces perpendicular to the area on which they act. An example of shear stress is the stress exerted on a pipeline by a flowing fluid. Shear is what causes a sand castle to collapse when someone steps on it. During an earthquake, there is an increase in shear stresses caused by seismic loading. As a result of this, landslides in dam slopes and slumping of dam embankment may occur, causing calamities. Although seismic shear forces may develop on structures in any direction, they are more critical in the horizontal direction. Hence, structures are analysed for seismic withstand using horizontal forces in two mutually perpendicular horizontal directions of a structure.
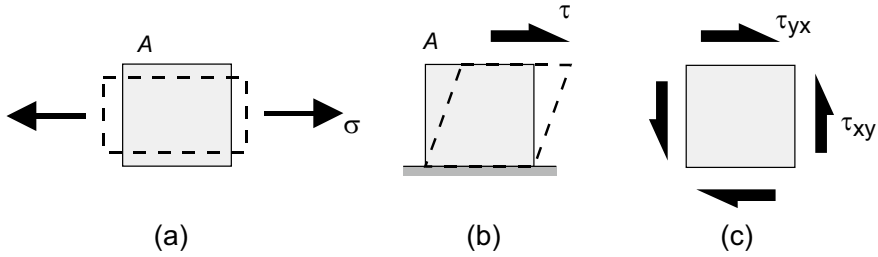
Fig. 6.6 **a** Normal deformation, **b** Shearing deformation, **c** Shear stress

Many of the deformations in civil engineering structures or in machine elements are due to 'shearing', or 'distortional', effects, and not due to elongational or compressive effects only. To understand the concept of shearing distortions, one may consider a rectangular grid drawn on a specimen subjected to tensile loading as shown in Fig. 6.6a. If a uniaxial load were to be applied, the grid would deform such that its length would be increased along the direction of the load and decreased perpendicular to the loading direction. It has already been seen in Sect. 6.2 that these deformations result in normal or axial strains. In Fig. 6.6b, the load is applied parallel to the upper surface of the specimen, causing the horizontal lines to slide relative to one another; however, the lengths of the lines of the originally rectangular grid would remain unchanged. The vertical lines tilt to accommodate this motion; hence, the right angles between the lines are no longer so. The shear stress due to such a loading P parallel to the surface is called direct shear stress.

$$\tau = \frac{P}{A} \tag{6.4}$$

Here A is the area resisting the force and τ is the shear stress. Though this equation is identical to the expression for normal stress, the symbol τ used for shear stress is a reminder that the loading is not axial. The effect of normal or tensile or axial stresses is to pull parallel planes within the material apart or push them nearer. Shear stresses, on the other hand, act so as to slide planes along one another. Shear stress is typically depicted in the form of a stress square shown in Fig. 6.6c, half-arrowheads being shown to differentiate shear stress from normal stress. The subscript yx indicates that the stress is on the $+y$-plane in the x-direction. For maintaining horizontal equilibrium, the τ_{yx} arrow on the $+y$-plane should be accompanied by one with an opposite sign on the $-y$-plane. In order to prevent the clockwise rotation that would be caused by these two arrows themselves and to maintain moment equilibrium, two vertical arrows should be added as shown in Fig. 6.6c; these are labelled τ_{xy}, since they are on the $+x$-plane in the y-direction.

For rotational equilibrium, the magnitudes of the horizontal and vertical stresses must be equal, implying that

$$\tau_{xy} = \tau_{yx} \tag{6.5}$$

Fig. 6.7 Shear strain
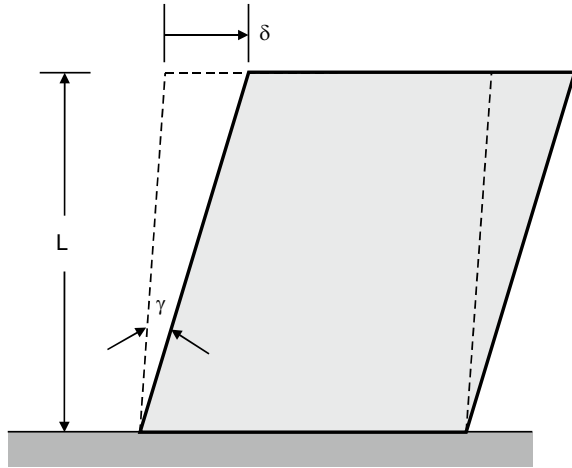
This implies that any shearing causing tangential sliding of horizontal planes should be accompanied by an equal sliding of vertical planes as well. The sign convention used here is that + arrows on + faces are positive, which means that a positive state of shear stress has arrows meeting at the upper right and lower left of the stress square. Likewise, a negative state of shear stress has arrows meeting at the lower right and upper left. The strain due to the shear stress τ_{xy} is denoted by γ_{xy}. Though this is also a deformation per unit length like the normal strain ε, it is tangential to the length over which it is distributed (Fig. 6.7). This is also equal to the distortion or change in the right angle:

$$\frac{\delta}{L} = \tan \gamma \approx \gamma \tag{6.6}$$

Experiments conducted in the past by many researchers have shown that at sufficiently small loads, this angular distortion is proportional to the shear stress, and the shearing counterpart of Hooke's Law can be written as

$$\tau_{xy} = G\gamma_{xy} \tag{6.7}$$

where G is a material property called shear modulus. For isotropic materials, Poisson effect need not be considered in shear; hence, shear strain is not affected by the presence of normal stresses. Likewise, application of a shearing stress does not have any influence on the normal strain. For plane stress scenario, i.e., no normal or shearing stress components in the z-direction, the constitutive equations can be written as follows:

$$\epsilon_x = \frac{1}{E}(\sigma_x - \upsilon\sigma_y) \tag{6.8a}$$

$$\epsilon_y = \frac{1}{E}(\sigma_y - \upsilon\sigma_x) \tag{6.8b}$$

$$\gamma_{xy} = \frac{1}{G}\tau_{xy} \tag{6.8c}$$

where υ is Poisson's ratio. It can easily be shown that

$$G = \frac{E}{2(1 + \upsilon)} \tag{6.9}$$

For isotropic materials, since only two of the material constants are independent, if any of the two properties E, G, or υ is known, the third can be determined.

6.4   Two-Dimensional Stress–Strain

INTERESTING FACTS: There are many examples of biaxial loading in solid mechanics. The effect of such loading on circular and elliptical holes has been well studied for many years. Tests are carried out under biaxial loading conditions in order to find the material properties of sheet metals. Tests under biaxial loading or pressurization are also conducted on many coated fabric applications such as inflatables, airbags, and geotextiles to understand their behaviour. Human and animal tissues are generally subjected to multiaxial loading in vivo.

Consider a plate made of a material which is isotropic and is pulled in two directions x and y. This problem is one of plane stress since σ_z is zero as shown in Fig. 6.8. The elongation in any one of the directions x or y will depend on the tensile stress in that direction and the stress in the perpendicular direction. The strain in the direction of the x-axis due to tensile stress σ_x is σ_x/E. The tensile stress σ_y will produce a lateral contraction in the x-direction given by $\upsilon\sigma_y/E$. Now if there are normal stresses σ_x and σ_y acting simultaneously, then using the principle of superposition, the total normal strain in the x-direction can be written as follows:
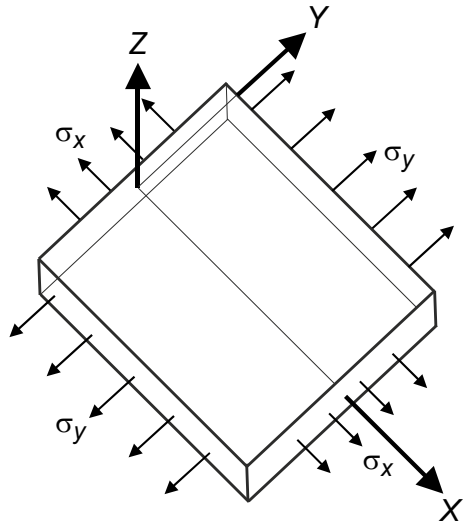
$$\epsilon_{x\,\text{total}} = \epsilon_{x\,\text{due to }\sigma_x} + \epsilon_{x\,\text{due to }\sigma_y} = \frac{\sigma_x}{E} - \frac{\upsilon\sigma_y}{E}$$

or

$$\epsilon_x = \frac{\sigma_x}{E} - \frac{\upsilon\sigma_y}{E} \tag{6.10}$$

Similarly the normal strain in the y-direction is

Fig. 6.8 Plate subjected to
biaxial loading

$$\epsilon_y = \frac{\sigma_y}{E} - \frac{\upsilon\sigma_x}{E} \tag{6.11}$$

From Eqs. (6.10) and (6.11), the stresses σ_x and σ_y can be obtained as functions of ε_x and ε_y as shown in the equations that follow:

$$\sigma_x = \frac{(\epsilon_x + \upsilon\epsilon_y)E}{1 - \upsilon^2} \qquad \sigma_y = \frac{(\epsilon_y + \upsilon\epsilon_x)E}{1 - \upsilon^2} \tag{6.12}$$

However, Hooke's law also defines the relationship between shear strain and shear stress. If the shear stress and strain occur in a plane, then they are related as

$$\tau = G\gamma = \frac{E\gamma}{2(1 + \upsilon)} \tag{6.13}$$

or

$$\gamma = \frac{\tau}{G} = \frac{2(1 + \upsilon)\tau}{E} \tag{6.14}$$

where G is the shear modulus as seen in Sect. 6.3 and γ is the shear strain.

The two-dimensional stress–strain relationships are given in matrix form as

$$\begin{Bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma \end{Bmatrix} = \begin{bmatrix} 1/E & -\upsilon/E & 0 \\ -\upsilon/E & 1/E & 0 \\ 0 & 0 & 1/G \end{bmatrix} \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau \end{Bmatrix} \tag{6.15}$$

or in the matrix form

$$\begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau \end{Bmatrix} = \begin{bmatrix} \frac{E}{1-v^2} & \frac{vE}{1-v^2} & 0 \\ \frac{vE}{1-v^2} & \frac{E}{1-v^2} & 0 \\ 0 & 0 & G \end{bmatrix} \begin{Bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma \end{Bmatrix} \tag{6.16}$$

6.5 Thin-Walled Pressure Vessel

DID YOU KNOW that pressure vessels are air-tight containers which are used in the process industry, refineries, and in petrochemical plants to carry or hold liquids, gases, or process fluids? The pressure vessels most commonly used in industry are heat exchangers, tanks, towers, boilers, reactors, drums, condensers, pipings, and columns. They are also used in homes and hospitals for storage of hot water and gases. Many of these vessels are subjected to very high-pressure loading as compared to the ambient pressure. The larger the vessel and the higher the operating pressure, the higher is the energy stored in it and the higher is the energy released if there is a rupture. A very famous catastrophe in recent times is that of a pressure vessel weighing 22680 kg which exploded at Marcus Oil, a chemical plant in Houston, Texas in 2004. The explosion threw heavy fragments of the vessel all around, damaging a church and nearby cars. Buildings in the vicinity experienced significant structural and interior damage. Pieces of shrapnel metal with mass over 450 kg travelled at extremely high speeds up to 800 m away and fell on nearby highways and railways.

In this section, the theory behind a thin-walled pressure vessel is discussed. For analysing the elastic deformations of such vessels, generalized Hooke's law is used. The walls of an ideal thin-walled pressure vessel resist bending and act as a membrane. Though spheres are also widely used, only cylindrical pressure vessels are analysed, the main difficulty in the latter being the junctions at the ends of the cylinder. A segment of this vessel is isolated as shown in Fig. 6.9.

Let the vessel have the following dimensions:

$L =$ length,
$D =$ internal diameter,
$t =$ wall thickness,
$r_i =$ inside radius,
$r_o =$ outside radius,
$p =$ fluid pressure inside the vessel.

The assumptions made for this analysis are that along the sections of this cylindrical free body, there are only normal stresses and no shear stresses are present in the planes of the sections as they could cause an incompatible distortion of the cylinder. The two normal stresses that occur are the circumferential or hoop stresses σ_h and
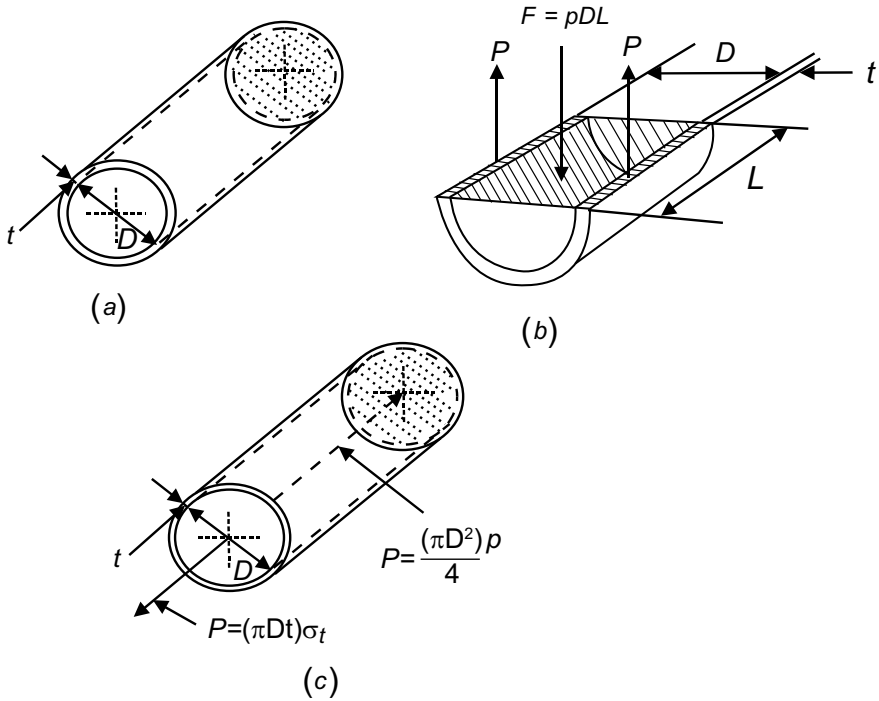
Fig. 6.9 Thin-walled pressure vessel: **a** vessel, **b** section, and **c** Forces

the longitudinal stresses σ_l identified as $\sigma_1 = \sigma_h$ and $\sigma_2 = \sigma_l$. The forces obtained by multiplying these stresses by their respective areas keep the cylindrical element in equilibrium with the internal fluid pressure. The internal pressure p multiplied by the projected area $2r_i L$ generates the force acting on the cylindrical element. This force is balanced by the two forces P developed by the hoop stresses $\sigma_h = \sigma_l$ multiplied by their respective areas $L(r_o - r_i)$. By examining the free body diagram of the lower half of the cylinder shown in Fig. 6.9b, the summation of forces acting normal to the mid-plane is given by

$$\sum F = 0$$
$$F = 2pr_i L = pDL = 2P = 2\sigma_1(r_i - r_o)L = 2\sigma_1 t L$$
$$\sigma_1 = \sigma_h = \frac{pr_i}{t} \tag{6.17}$$

This equation is valid only for thin-walled cylinders, as it gives the average stress in the hoop. Even with a wall thickness one-tenth the inner radius, the error in the above equations would still be small.

The other normal stress $\sigma_2 = \sigma_l$ is determined by solving for equilibrium of the axial forces. The force developed by the internal pressure is equated to the force developed by the longitudinal stress σ_2 in the walls as shown below:

$$p\pi r_i^2 = \sigma_2(\pi r_o^2 - \pi r_i^2)$$

$$\sigma_2 = \frac{p\pi r_i^2}{(\pi r_o^2 - \pi r_i^2)} = \frac{pr_i^2}{(r_o - r_i)(r_0 + r_i)} \tag{6.18}$$

Since $r_o \approx r_i \approx r$

$$\sigma_2 = \frac{pr}{2t} \tag{6.19}$$

It may be noted that for thin-walled cylindrical pressure vessels, $\sigma_2 \approx \sigma_1/2$.

6.6 Beams in Bending

DID YOU KNOW that it took many years to understand the stresses induced in beams by bending loads? Galileo Galilei is often given credit to the first published work on the strength of beams in bending, but with the discovery of 'The Codex Madrid' in the National Library of Spain in 1967, it became clear that Leonardo da Vinci's work published in 1493 had preceded Galileo's work by over a century. Besides, unlike Galileo, da Vinci had correctly identified the stress and strain distribution across a section of a beam in bending. Galileo worked on this problem, but the theory as is used today with concepts of flexural strength and stiffness of beams is now the result of the Euler–Bernoulli beam theory or classical beam theory. It was first proposed around 1750, but was applied on a large scale only after the development of the Eiffel tower in Paris and the Ferris wheel in Chicago in the late nineteenth century. In 1922, Timoshenko corrected this equation to take into account shear deformation and rotational bending effects, making it suitable for describing the behaviour of thick beams as well.

Let us consider the case of bending of a beam. A book shelf sagging under the weight of books is an example of a beam experiencing bending. Figure 6.10 shows a case of pure bending. A beam subjected to a positive bending moment will develop normal stresses in the lengthwise direction with the upper region of the beam being under compression and the lower region in tension. In the transition zone between the compressive and tensile regions, the stress becomes zero; this is the neutral axis of the beam. Figure 6.11a shows a small section of the beam before bending, Fig. 6.11b shows the same section after bending and Fig. 6.11c shows the variation of stress across the cross section. Figure 6.11d shows a cross section of the beam.
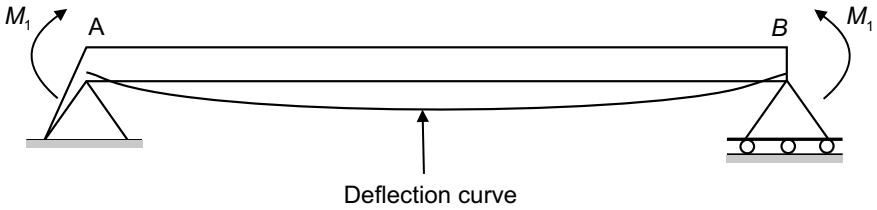
Fig. 6.10 Beam under pure bending
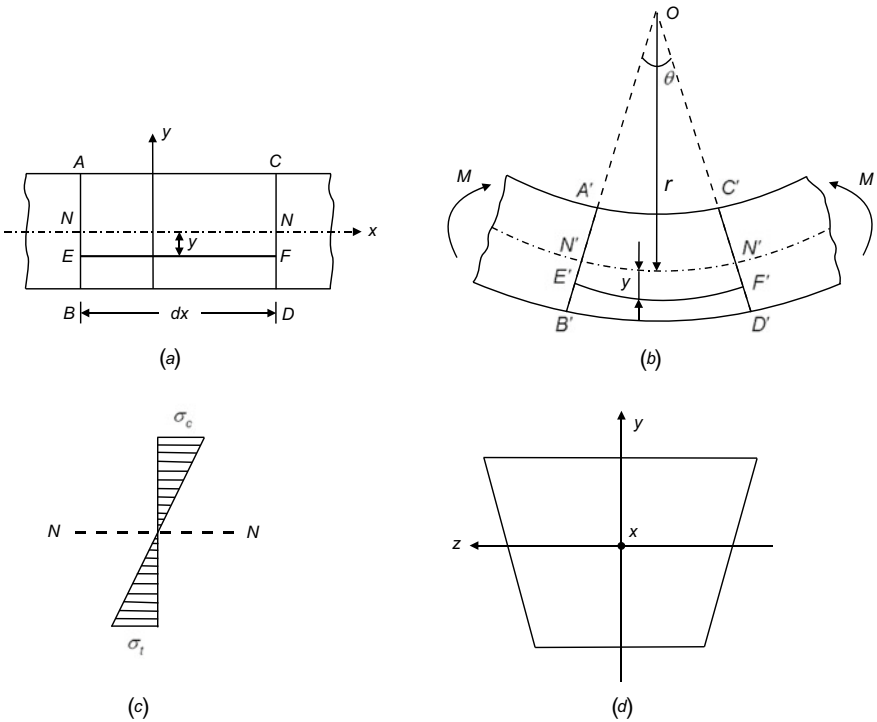
Fig. 6.11 Bending of beam: **a** beam before bending, **b** beam after bending, **c** stress diagram, and **d** beam cross section

The assumptions made for studying the pure bending of a beam are as follows:

1. The beam is straight before the application of loads and has a constant cross-sectional area.
2. Due to symmetry of loading and symmetry of the cross section, the beam bends uniformly into an arc of a circle.
3. The beam is subjected to pure bending, i.e., the bending moment does not change along the length.
4. The material of the beam is homogeneous and isotropic.

The same basic concepts of engineering mechanics of solids that are used for developing theories for axially loaded bars and circular shafts in torsion are used for deriving expressions for bending of a beam as well. They are as follows.

1. The equilibrium relations in statics are used to obtain the internal resisting bending moment at a section.
2. The geometry of deformation (based on experiments) clearly shows that when beams are subjected to pure bending, they deform in such a way that plane sections remain plane. This implies that planes perpendicular to the longitudinal axis before loading remain plane and perpendicular to the axis after loading, but rotate through an angle θ about points on the neutral axis as shown in Fig. 6.11b.
3. From kinematics it is clear that the strains are zero at the neutral axis, i.e., $y = 0$, negative (compressive) above the axis and positive (tensile) below. They increase in magnitude linearly with y, in much the same manner as shear strains increase linearly with the radius in a torsionally loaded circular shaft.
4. Constitutive equations in the form of Hooke's law are used to obtain the stresses. This restricts the applicability of this derivation to linear elastic materials.

Let the x-axis be attached to the neutral axis of the beam. Let r denote the radius of curvature of the deflected axis of the beam. The elongation of any fibre EF at a distance y from the neutral axis may be computed to be

$$\epsilon_x = -\frac{y}{r} \tag{6.20}$$

The strains of the longitudinal fibres are proportional to the distance from the neutral axis. The longitudinal expansion of the fibres on the lower (convex) side of the beam is accompanied by lateral contraction; likewise, longitudinal compression on the upper (concave) side is accompanied by lateral expansion of the same value. Hence, the cross section changes, with the vertical sides becoming inclined to each other as seen in Fig. 6.11d.

The unit strain in the lateral direction is

$$\epsilon_z = -\nu\epsilon_x = \nu\frac{y}{r} \tag{6.21}$$

where ν is Poisson's ratio.

From the strains of the longitudinal fibres, the corresponding stresses may be obtained using Hooke's law.

$$\sigma_x = -\frac{Ey}{r} \tag{6.22}$$

Figure 6.11c shows the distribution of these stresses. The stress in any fibre is proportional to its distance from the neutral axis NN. In Eq. (6.22), the two unknowns are the position of the neutral axis and the radius of curvature. They can be found out from the condition that the forces distributed over any cross section of the beam should give rise to a resisting couple which balances the external couple M as shown in Fig. 6.11b.

Consider an elemental area dA at a distance y from the neutral axis. The small force acting on this elemental area is $(Ey/r)dA$. All such elemental forces distributed over the cross section represent a system equivalent to a couple; hence, the resultant of these forces in the x-direction must be equal to zero. Therefore, we get

$$\int \frac{Ey}{r}dA = \frac{E}{r}\int ydA = 0 \tag{6.23}$$

This implies that the area of the cross section with respect to the neutral axis is equal to zero or, in other words, the neutral axis passes through the centroid of the section. The moment of the force acting on the element dA about the neutral axis is $(Ey/r)dA.y$. Summing up all such moments over the cross section and equating the resultant to the moment M of the external forces, the radius of curvature may be obtained as follows:

$$\int \frac{Ey^2}{r}dA = \frac{EI_z}{r} = M$$

or

$$\frac{1}{r} = \frac{M}{EI_z} \tag{6.24}$$

This is because the moment of inertia of the cross section with respect to the neutral axis z is given by

$$I_z = \int y^2 dA \tag{6.25}$$

Using Eqs. (6.22) and (6.24), r may be eliminated to give the following equation:

$$\sigma_x = -\frac{My}{I_z} \tag{6.26}$$

This discussion will hold good for a bar of any cross section with a longitudinal plane of symmetry and which is bent by end couples acting in this plane. Bending thus takes place in the plane of the couples. The maximum compressive and tensile stresses for a rectangular cross section with its centroid at the middle of the depth h occur at $y = +h/2$ or $y = -h/2$. Then for positive M, we get

$$(\sigma_x)_{\text{max}} = \frac{Mh}{2I_z} \quad \text{and} \quad (\sigma_x)_{\text{min}} = -\frac{Mh}{2I_z} \tag{6.27}$$

If a material, like glass or chalk, has a tendency to fail in tension, crack initiation, and growth start from the lower tensile surface. If the material is strong in tension, but weak in compression such as wood, it will fail at the top compressive surface due to buckling of the outer fibres. Shear stresses also develop, but these are often negligible compared to the normal stresses when the length-to-height ratio of the beam is large.

6.7   Torsional Stresses and Displacements

INTERESTING FACTS: A slender member subjected to a load that tends to twist it about its longitudinal axis is said to be in torsion. Torsionally loaded shafts are commonly used structures in mechanical engineering. For example, the drive shaft of a standard rear-wheel drive automobile helps transmit torsion. These shafts are almost always hollow and circular in cross section, transmitting power from the power source or engine to the differential joint from where the rotation is conveyed to the drive wheels. Simple tools like a screwdriver develop a torsional stress and a twist of the screwdriver when a screw is driven into wood. Similarly, a torque wrench applies a torque to the head of a bolt causing twisting of the bolt and internal torsional stresses. When we turn a key in a lock, the handle or shank of the key is in torsion. Another example of a torsionally loaded member is the twist drill. One of the simplest examples of a torsional force in action is twisting a simple piece of a blackboard chalk between one's fingers till it breaks.

6.7.1   Theory

Analysis of the stresses and angular deformations induced in a shaft of circular cross section by a twisting moment or torque can be determined using the so-called direct method of stress analysis. In this method, initially, an expression for the geometrical form of displacement is proposed. Subsequently, the kinematic, constitutive, and equilibrium equations are applied sequentially to obtain equations for the strains and stresses. In the case of simple twisting of a circular shaft, the geometric implication is that the circular symmetry of the shaft is maintained, and therefore plane cross sections remain plane, without warping. As depicted in Fig. 6.12, the deformation is like a stack of coins that rotate relative to one another while remaining flat. Direct analysis then has the following steps:

1. **Geometrical statement:** To obtain the angular deformation, a very small length dz of a circular shaft is considered as shown in Fig. 6.12; here, the top surface rotates relative to the bottom surface by an incremental angle $d\theta$. The tangential displacement of the top of a vertical line relative to the bottom at a distance r from the centre is given by

$$\delta = r\,d\theta \tag{6.28}$$

2. **Kinematic or strain–displacement equation:** The geometry of deformation matches the description of shear strain; hence,

Fig. 6.12 Deformation in torsion
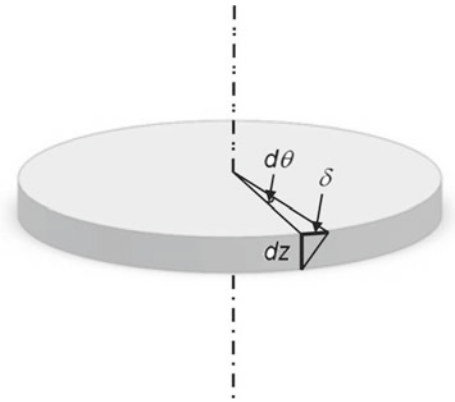
$$\gamma_{z\theta} = \frac{\delta}{dz} = r\frac{d\theta}{dz} \tag{6.29}$$

The subscript $z\theta$ indicates a shearing of the plane normal to the z-axis, (i.e., the z-plane) in the θ-direction. As with the shear stresses, $\gamma_{z\theta} = \gamma_{\theta z}$, so the order of subscripts is not important.

3. **Constitutive equation:** The assumption behind the derivation is that the material is in its linear elastic region of operation. Hence, the shear stress is got from Hooke's law as

$$\tau_{\theta z} = G\gamma_{\theta z} = Gr\frac{d\theta}{dz} \tag{6.30}$$

The sign convention followed here is that positive twisting moments (moment vector along the $+z$-axis) produce positive shear stresses and strains. Hence, in Fig. 6.13, the upper surface $(+z)$ is being twisted clockwise relative to the lower surface $(-z)$ and therefore the upper arrow points to the right.

4. **Equilibrium equation:** For rotational equilibrium to be maintained, the sum of the moments contributed by the shear stress acting on each elemental area dA on the cross section should balance the applied moment T as shown in Fig. 6.13:

$$T = \int_A \tau_{\theta z} r \, dA = \int_A Gr\frac{d\theta}{dz} r \, dA = G\frac{d\theta}{dz} \int_A r^2 dA \tag{6.31}$$

The quantity $I_p = \int_A r^2 dA$ is the polar moment of inertia. The quantity $d\theta/dz$ can now be found as

$$\frac{d\theta}{dz} = \frac{T}{GI_p} \rightarrow \theta = \int_z \frac{T}{GI_p} dz \tag{6.32}$$

Fig. 6.13 Torque balance
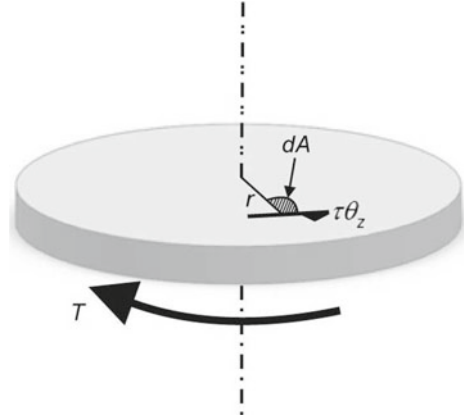
In the simple example of torsion being considered, the quantities T, I_p, and G are constant along the z-direction. Hence, the angle of twist can be written as

$$\frac{d\theta}{dz} = \text{constant} = \frac{\theta}{L} \tag{6.33}$$

or

$$\theta = \frac{TL}{GI_p} \tag{6.34}$$

It is clearly seen that this expression is analogous to the equation $\delta = PL/AE$ for the expansion of a uniaxial tensile specimen.

5. **Final expression:** The formula for the stress can be obtained by using this in Eq. (6.31):

$$\tau_{\theta z} = Gr\frac{d\theta}{dz} = Gr\frac{\theta}{L} = \frac{Gr}{L}\frac{TL}{GI_p} \tag{6.35}$$

or

$$\tau_{\theta z} = \frac{Tr}{I_p} \tag{6.36}$$

It may be noted that the material property G has got cancelled from the final expression for stress; hence, the stresses are independent of the choice of material. Equation (6.36) also explains why most drive shafts are hollow. It is clear that it does not make sense to use material at the centre where the stresses are zero. Also, the designer will intend to maximize the moment of inertia by placing the material as far from the centre as possible (since I_p varies as the fourth power of the radius) with a given quantity of material.
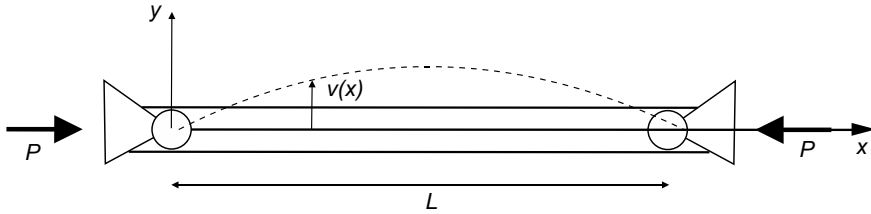
Fig. 6.14   Column with deflection v

6.8   Buckling of Columns

> **INTERESTING FACTS:** There are two major categories leading to the sudden failure of a mechanical component: (i) material failure and (ii) structural instability, which is often called buckling. For failures of the first kind, one needs to consider the yield stress for ductile materials and the ultimate stress for brittle materials. Failure due to buckling happens due to the loss of stability of a component and is usually independent of material strength. This loss of stability generally occurs within the elastic range of the material. Thin-walled components under compressive stress are susceptible to 'Euler buckling', in which the column develops a lateral deflection somewhere along its length and quickly collapses unless the load is removed. This is actually a bending phenomenon, caused by the bending moment that develops if and when the beam undergoes a lateral deflection. Buckling analysis of a column consists of determining the maximum compressive load a column can support before it collapses. But for long columns, the collapse has nothing to do with the material of the column. It is instead governed by the column's stiffness, both material and geometric. The force, P, necessary to induce such a buckling motion typically varies by a factor of four depending only on the boundary conditions at the two ends. Hence, buckling studies are much more sensitive to the component restraints than in a normal stress analysis.

Consider a beam of length L and cross section A loaded in axial compression and pinned at both ends so that it is free to rotate at its ends. Let it be subjected to an axial load P, as shown in Fig. 6.14. Now let the beam be given a transverse deflection, v, perhaps by a sideward load or even an irregularity in the beam's cross section.

Moment equilibrium of a section of the deflected column cut at a point x is applied; using the moment-curvature results in

$$- Pv(x) = M(x) = EI\frac{d^2v}{dx^2} \tag{6.37}$$

Hence, the deflection v satisfies the differential equation

$$\frac{d^2v}{dx^2} + k^2 v(x) = 0 \tag{6.38}$$

where

$$k^2 = \frac{P}{EI} \tag{6.39}$$

The ordinary differential equation (6.38) is linear, homogeneous, and has constant coefficients. Its solution can be obtained from any standard text book on differential equations and is given by (for $k^2 > 0$)

$$v(x) = A_1 \cos(kx) + A_2 \sin(kx) \tag{6.40}$$

where A_1 and A_2 are unknown constants. The boundary conditions for pinned ends are

$$v(x) = 0 \quad \text{and} \quad v(L) = 0 \tag{6.41}$$

The first condition requires that A_1 be zero and the second leads to

$$A_2 \sin(KL) = 0 \tag{6.42}$$

It follows that either:

(a) $A_2 = 0$, in which case $v(x) = 0$ for all x and the column is not deflected or
(b) $\sin(kL) = 0$, which is valid when kL is an integral number of π's, i.e.,

$$k = \frac{n\pi}{L}, \quad n = 1, 2, 3, \ldots \tag{6.43}$$

The solution (a) is governed by the axial deformation theory discussed in Sect. 6.2. Concentrating on (b), the corresponding solution for the deflection is

$$v_n(x) = \sin\left(\frac{n\pi x}{L}\right), \quad n = 1, 2, 3, \ldots \tag{6.44}$$

Knowing parameter k, buckling load is

$$P_n = EI\left(\frac{n\pi}{L}\right)^2, \quad n = 1, 2, 3, \ldots \tag{6.45}$$

It is clear that buckling, i.e., $v \neq 0$, can only occur for a discrete set of applied loads called the buckling loads as given in Eq. (6.45). In reality, the most important buckling load is the first, corresponding to $n = 1$, since this will be the first load reached when the applied load P is increased from zero. This is called the critical buckling load and is given by

Fig. 6.15 Effect of boundary conditions on critical buckling

Case	1	2	3	4	5
Constraints					
n	4	1	0.25	2.046	1

$$P_{cr} = EI \left(\frac{\pi}{L}\right)^2 \tag{6.46}$$

with associated deflection

$$v_1(x) = A_2 \sin\left(\frac{\pi x}{L}\right) \tag{6.47}$$

The above equation is the characteristic equation or eigenfunction of the problem. The critical buckling stress is given by

$$\sigma_{cr} = \frac{P_{cr}}{A} = \frac{EI}{A}\left(\frac{\pi}{L}\right)^2 \tag{6.48}$$

It is seen that the column deforms into a single sine wave, which is called the mode shape of the deflected column. It is to be noted that A_2, the amplitude of the deflection, cannot be found from this model. This is due to the assumption of small deflection obtained by linearizing the problem. The same procedure as described above can be used to determine the critical axial loads for columns with different boundary conditions as shown in Fig. 6.15. In this figure, n is a constant depending on the boundary conditions. Buckling also depends on what is called the slenderness ratio, defined as shown below:

$$S = \frac{L_e}{r_g} = \frac{KL}{r_g} \tag{6.49}$$

where L_e is the effective length of the column and r_g is the smallest radius of gyration about the axis under consideration. The solutions to these problems are very sensitive to the boundary conditions. The buckling patterns for different restraints are shown in Fig. 6.15.

6.9   Stress Concentration for Flat Bars in Tension

DID YOU KNOW that the Saint-Venant principle was named after Adhémar Jean Claude Barré de Saint-Venant who was a great mathematician, scientist, and engineer of the nineteenth century in France? He contributed to the fields of elasticity, mechanics, hydrostatics, and hydrodynamics. He graduated from the École Polytechnique and spent a good part of his life as a civil engineer. He worked on stress analysis and also developed the unsteady open channel flow equations called the Saint-Venant equations, which are popular in hydraulics. He is credited with having given the correct derivation of the Navier–Stokes equations. He also derived solutions for the torsion of non-circular cylinders and extended Navier's work on the bending of beams. In spite of all these contributions, his religious and political views made him unpopular in his own country.

The analysis of axially loaded bars through engineering mechanics is very accurate for bars of constant cross section when the load is uniformly distributed over the entire cross section. For such ideal conditions, the stresses and strains are uniform everywhere along the cross section and are measures of the average stress over a cross section. But in reality, the applied external forces can be concentrated over a small region (they are then called point loads) and the cross sections of members can change abruptly. This causes the stress and strain to be much higher than the average stress and disturbed near the points of loading and locations of changes in cross sections. This results in very complicated deformations due to the complicated states of stress. Such problems can be conveniently studied using the finite element method. There are two important observations to be made here. The first is that the average stress is always correct since it is based on conditions of equilibrium. This principle assumes that the stress is evenly distributed or averaged over the cross section. The second is that the normal stresses at a distance equal to and greater than the width of the member are uniform. This is a statement of the Saint-Venant principle which was proposed in 1855. This principle allows us to quickly identify where in the structure this assumption is true, like bolt holes or locations of changes in cross section, where the maximum normal stresses are finite. The ratio of maximum stress to the average stress is called stress concentration factor and depends on the geometry of a member.
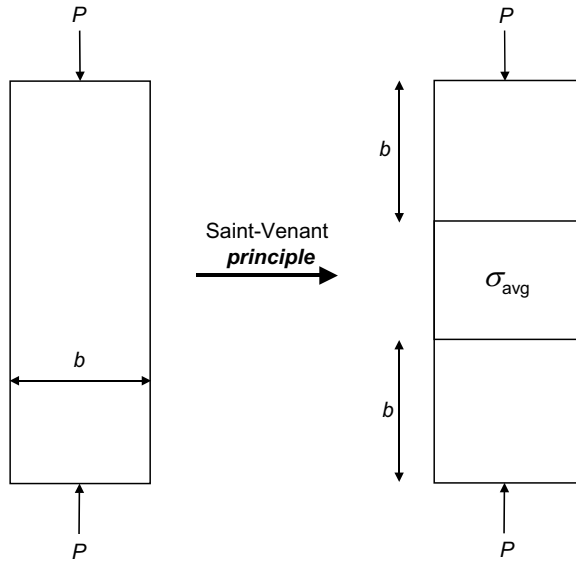
Figure 6.16 depicts the Saint-Venant principle. The maximum normal stress at a section is given by

$$\sigma_{\max} = K\sigma_{av} = K\frac{P}{A} \tag{6.50}$$

where K is called the stress concentration factor.

Chapter 7 discusses strain gauges and other equipment commonly used for strain measurement. Chapter 12 describes basic equipments in strain measurement.

Fig. 6.16 Saint-Venant
principle

List of Symbols

Symbol	Meaning	SI unit
A	area resisting load	m^2
A_1, A_2	constants	
c	viscous damping coefficient	N.s/m
dA	elemental area	m^2
dz	incremental length of shaft	m
$d\theta$	incremental angular rotation of shaft	rad
D	internal diameter of pressure vessel	m
E	Young's modulus of elasticity	Pa
F	force	N
F_x	force in x-direction	N
F_y	force in y-direction	N
F_z	force in z-direction	N
G	shear modulus	Pa
h	depth of beam	m
I_p	polar moment of inertia	m^4

Symbol	Meaning	SI unit
I_z	area moment of inertia about z axis	m^4
k	spring constant or stiffness	N/m
K	stress concentration factor	
L	length of block, pressure vessel, column	m
L_e	effective length of column	m
$M, M(x)$	moment	N.m
M_x	moment about x-axis	N.m
M_y	moment about y-axis	N.m
M_z	moment about z-axis	N.m
p	fluid pressure inside pressure vessel	Pa
P	load	N
P_{cr}	critical buckling load	N
r	radial distance from centre, radius of curvature	m
r_i	inside radius of pressure vessel	m
r_g	radius of gyration	m
r_o	outside radius of pressure vessel	m
S	slenderness ratio	
t	wall thickness of pressure vessel	m
T	torque	N.m
x	deformation	m
y	distance of fibre from neutral axis	m
u, v, w	deflections along $x-$, y-, and z-axes	m
$\gamma, \gamma_{xy}, \gamma_{yx}$	shear strains	m/m
$\gamma_{z\theta}, \gamma_{\theta z}$	shear strain in $z\theta$ plane	m/m
δ	tangential displacement	m
$\varepsilon, \varepsilon_x, \varepsilon_y$	normal strains	m/m
ν	Poisson's ratio	
σ	normal stress	Pa
σ_1, σ_2	principal stresses	Pa
σ_{av}	average stress	Pa
σ_c	compressive stress	Pa
σ_{cr}	critical buckling stress	Pa
σ_h	hoop or circumferential stress	Pa
σ_l	longitudinal stress	Pa
$\sigma_{\max}, (\sigma_x)_{\max}$	maximum normal stress	Pa
σ_t	tensile stress	Pa
σ_x, σ_y	normal stresses	Pa
$(\sigma_x)_{\min}$	minimum normal stress	Pa
$\tau, \tau_{xy}, \tau_{yx}$	shear stresses	Pa

Questions

1. How do you define the strength of a material?
2. Even when a member is in equilibrium because the resultant of the external forces is zero, why does it deform?
3. What factors influence the strength of a material?
4. What are the various types of loads that can act on a member?
5. How do you define Young's modulus of elasticity?
6. Draw a figure of a bar under axial compressional loading.
7. For a three-dimensional body in space to be in equilibrium, what are the equations to be satisfied?
8. What is meant by a statically indeterminate structure?
9. What is the contribution of Leonhard Euler to elasticity?
10. Give practical examples of axially loaded members.
11. Define the loads acting on a bungee cord used in outdoor activities.
12. What is meant by a statically determinate structure?
13. How is shear stress different from normal stress?
14. What equations are required for the solution of statically indeterminate structures?
15. What is an isotropic material?
16. Draw the stress–strain curve for mild steel.
17. Give examples of isotropic materials.
18. How do you model the elastic response of a material?
19. What is meant by the upper yield point of a material?
20. What is the expression for shear strain?
21. What is the expression for normal/axial stress?
22. What is Euler–Bernoulli beam theory?
23. What are the assumptions made in Euler–Bernoulli beam theory?
24. What are anisotropic materials?
25. Explain the stress–strain curve for a viscoelastic material.
26. What is the lower yield point?
27. What is the fracture point?
28. What causes shear stress?
29. What is meant by biaxial loading? Give examples.
30. What phenomenon causes landslides in dams during an earthquake?
31. What causes buckling?
32. What is a viscoelastic material?
33. Why do long columns collapse?
34. How do you define stress concentration factor?
35. What is the difference between elastic and viscoelastic materials?
36. What is Kelvin–Voigt model?
37. How do you define shear modulus?
38. What is meant by Poisson effect?
39. Name three important material constants for isotropic materials.

40. How are walls of an ideal thin-walled pressure vessel that resist bending represented?
41. What is meant by hoop stress?
42. What is meant by PURE bending of a beam? Give examples.
43. Give an example of a torsionally loaded material.
44. What stresses act on a thin-walled pressure vessel?
45. How do you define a pressure vessel as being thin- or thick-walled?
46. What is the Saint-Venant principle?
47. What is the meaning of a mode shape?
48. How do you define the neutral axis of a beam?
49. Describe the general relationship between the size of a pressure vessel and operating pressure.
50. Give an example of biaxial loading in solids.
51. What assumptions are made in the analysis of the pure bending of beams?
52. What is Hooke's law for stress and strain?
53. What are torsional stresses?
54. Show the distribution of stresses for a rectangular beam under bending.
55. Give examples of materials that typically fail under tension.
56. Give examples of materials which are weak in compression.
57. Give a practical example of a dashpot.
58. What is meant by plastic behaviour of a material?
59. What is polar moment of inertia? What is its physical significance?
60. Why are most drive shafts made hollow?
61. Give the constitutive equations for a material that is torsionally loaded.
62. What is meant by degree of indeterminacy?
63. Give examples of anisotropic materials that are used in daily lives.
64. What does buckling analysis of a column involve?
65. What is meant by Euler buckling of a column?
66. Give an example of a shock load.
67. What are truss members?
68. How do boundary conditions of a column determine its critical buckling load? Explain qualitatively.
69. When can you use the principle of superposition in stress analysis?
70. Give the expression for the bending stress in a beam.
71. What is the difference in behaviour between short and long columns subjected to compressive loads?
72. What concepts of engineering mechanics have to be used to understand the bending of beams?
73. Which stress is higher in a thin-walled pressure vessel, longitudinal or hoop?
74. If two bars of identical dimensions, but different materials are subjected to the same axial load, comment on the stresses and strains produced.
75. In beams under pure bending, what is the value of strain at the neutral axis?
76. What is the effect of point and distributed axial loads on a bar?
77. What does the radius of curvature of a beam in bending depend on?

78. Explain the analogy between the expressions for the angle of twist of a circular shaft in torsion and the compression/expansion of a uniaxially loaded bar.
79. The shear stress in a circular shaft under torsion depends on its shear modulus. State whether true or false with reasons.
80. In the buckling of a column, loss of stability usually occurs within the elastic range of the material. Why?
81. How are the shear modulus (G), modulus of elasticity (E), and Poisson's ratio (v) of a material related?
82. What is the difference between point loads and distributed loads?

Fill in the Blanks

83. The deformation per unit length is called _____.
84. When a rectangular beam is loaded transversely, the maximum compressive stress is developed on the _____ layer.
85. The bending stress in a beam is _____ proportional to the section modulus.
86. The stress at which the extension of the material takes place more quickly as compared to the increase in load is called _____.
87. A vertical column has two moments of inertia (i.e., I_{xx} and I_{yy}). The column will tend to buckle in the direction of the _____.
88. Euler's formula holds good only for _____ columns.
89. The maximum value of Poisson's ratio for an elastic material is _____.
90. The number of independent elastic constants for a linear, elastic, isotropic, and homogeneous material is _____.
91. If the shear force at a section of beam under bending is equal to zero, then the bending moment at the section is _____.
92. The first moment of area about the axis of bending for a beam cross section is _____.
93. The 'plane section remains plane' assumption in bending theory implies _____.
94. A viscoelastic material is _____ when loaded fast and _____ when loaded slowly.
95. The maximum stress for a beam with rectangular cross section subjected to bending is _____.
96. To understand material failure in ductile materials, one has to consider the _____ stress.
97. The hoop stress in a thin-walled cylinder is a function of _____.
98. The effect of axial stresses is to pull parallel planes within the material apart or push them closer together, shear stresses act so as to _____ planes along one another.
99. The _____ stress determines the failure of brittle materials.

Bibliography

 1. Beer, F. P., Johnston, E. R., Jr., Dewolf, J. T., Mazurek, D. F., & Sanghi, S. (2016). *Mechanics of materials (SI units)* (Special Indian ed.). New Delhi: McGraw-Hill Education.
 2. Boresi, A. P., & Schmidt, R. J. (2009). *Advanced mechanics of materials* (Wiley Student ed.). New Delhi: Wiley.
 3. Budynas, R. (2010). *Advanced strength and applied stress analysis*. Boston: McGraw-Hill.
 4. Crandall, S. H., & Dahl, N. C. (1978). *An introduction to the mechanics of solids*. New York: McGraw-Hill.
 5. Crandall, S. H., Dahl, N. C., & Lardner, T. J. (1999). *An introduction to the mechanics of solids*. New York: McGraw-Hill Science/Engineering/Math.
 6. da Silva, V. D. (2006). *Mechanics and strength of materials*. Berlin: Springer.
 7. Fenner, R. T., & Reddy, J. N. (2012). *Mechanics of solids and structures*. Boca Raton: CRC Press.
 8. Gere, J. M., & Goodno, B. J. (2013). *Mechanics of materials*. Boston: Cengage Learning.
 9. Gere, J. M., & Timoshenko, S. P. (2004). *Mechanics of materials*. New Delhi: CBS Publishers.
10. Hearn, E. J. (1997). *The mechanics of elastic and plastic deformation of solids and structural materials*. Oxford: Butterworth-Heinemann.
11. Hibbeler, R. C. (2018). *Mechanics of materials*. Hoboken: Pearson.
12. Huston, R. L., & Josephs, H. (2008). *Practical stress analysis in engineering design*. Boca Raton: CRC Press.
13. Javidinejad, A. (2010). *Essentials of mechanical stress analysis*. Boca Raton: CRC Press.
14. Karnovsky, I. A., & Lebed, O. (2010). *Advanced methods of structural analysis*. Spain: Springer.
15. Mase, G. E. (1970). *Theory and problems of continuum mechanics*. Schaum's outline series. New York: McGraw-Hill Book Company.
16. Megson, T. H. G. (2014). *Structural and stress analysis*. India: Elsevier.
17. Muvdi, B. B., & Elhouar, S. (2016). *Mechanics of materials: With applications in excel*. Boca Raton: CRC Press.
18. Nash, W. A., & Potter, M. C. (2011). *Schaum's outline of strength of materials*. New York: The McGraw-Hill Companies, Inc.
19. Philpot, T. A. (2017). *Mechanics of materials*. New York: Wiley.
20. Popov, E. P. (1968). *Introduction to mechanics of solids*. Englewood Cliffs: Prentice-Hall.
21. Popov, E. P. (1990). *Engineering mechanics of solids*. Prentice-Hall international series in civil engineering and engineering mechanics. Englewood Cliffs: Prentice Hall.
22. Popov, E. P. (2015). *Mechanics of materials*. New Delhi: Pearson Education India.
23. Prusty, G., & Chowdhury, M. (2008). *Mechanics of solids*. Melbourne: Pearson Education.
24. Shames, I. H., & Pitarresi, J. M. (1999). *Introduction to solid mechanics*. London: Pearson.
25. Singer, F. L., & Pytel, A. (1998). *Strength of materials*. New York: Harper & Row.
26. Srinath, L. S. (2010). *Advanced mechanics of solids*. New Delhi: Tata McGraw-Hill Education Pvt. Ltd.
27. Steif, P. S. (2012). *Mechanics of materials*. Upper Saddle River: Pearson.
28. Timoshenko, S. (2002). *Strength of materials part 1: Elementary theory and problems*. New Delhi: CBS Publishers and Distributors Pvt. Ltd.
29. Timoshenko, S. (2002). *Strength of materials part 2: Advanced theory and problems*. New Delhi: CBS Publishers and Distributors Pvt. Ltd.
30. Timoshenko, S., & Young, D. H. (1968). *Elements of strength of materials*. Princeton: Van Nostrand.
31. Timoshenko, S., Young, D. H., Pati, S., & Rao, J. V. (2017). *Engineering mechanics*. New Delhi: McGraw-Hill Education.
32. Timoshenko, S. P. (1983). *History of strength of materials*. New York: Dover Publications Inc.
33. Timoshenko, S. P., & Goodier, J. N. (2010). *Theory of elasticity*. New Delhi: Tata McGraw-Hill Education Pvt. Ltd.
34. Timoshenko, S. P., & Young, D. H. (2003). *Elements of strength of materials*. New Delhi: Affiliated East-West Press Pvt. Ltd.
35. Wonsiri, P., & Hsu, C. T. T. (2012). *Mechanics of materials: Laboratory and experiments*. Saarbrücken: Lambert Academic Publishing.

Chapter 7
Strain Gauge-Based Equipment

INTERESTING FACTS: The bonded wire strain gauge, which is the forerunner of the present-day metallic foil strain gauge, was developed independently and almost simultaneously by two different researchers situated at widely separated places in the United States of America who did not have any contact with each another. Edward E. Simmons at the California Institute of Technology (Caltech) developed the sensor while investigating the stress-strain behaviour of metals under shock loads in 1936. Arthur C. Ruge of the Massachusetts Institute of Technology invented the device in 1938 to help his graduate student John Meier complete his investigation of earthquake-induced stress on elevated water tanks. Although Simmons received official credit for inventing the wire resistance strain gauge, it was Ruge, who was instrumental in introducing many practical forms of the gauge and a variety of transducers that incorporate it as the input sensor. In 1939, Ruge started a business to manufacture the SR-4 gauge. From such humble beginnings grew a highly lucrative industry that laid the foundation for the present-day foil strain gauge-based transducers.

7.1   Introduction

Mechanical strain is the ever-present companion of physical deformation. Stress analysis and vibration analysis often go hand in hand in the study of dynamics. In engineering design, anything built from available materials must be tested to see if, regardless of its intended function, its form distributes the forces in acceptable concentrations of strains. This chapter describes various types of strain gauges and transducers used for the measurement of physical quantities that often accompany vibration such as force and strain. Besides, other measuring devices and equipment like stroboscopes, slip ring assemblies, and telemetry units that facilitate vibration and strain measurements on rotating machines are discussed. Details of operation of the signal conditioning amplifier called carrier frequency amplifier that is to be used with inductive vibration pick-ups and strain gauge based transducers are also described.

7.2   Strain Gauges

DID YOU KNOW that in 1952, Peter Jackson working in a company called Saunders-Roe in the United Kingdom developed the idea of making a foil strain gauge when he was doing stress analysis of rotors of helicopters? He faced problems with the then available bonded wire strain gauges due to their fatigue failures, slip ring noise problems and poor sensitivity. He made the first few foil gauges, but it took a few years before they were perfected. Foil gauges are manufactured from a thin foil using photo-etching technology, a process similar to that used in printed circuit technology. Soon these gauges began to replace bonded wire strain gauges owing to their improved performance and ability to withstand fatigue in more demanding environments. These foil gauges had a considerable reduction in size and production costs and hence are the most widely used gauges today. This fabrication technique allows production of almost any size and shape of strain gauge. It has better heat dissipation characteristics, simplified bonding techniques, reduced creep effects, lower cost and allows greater flexibility in providing gauges to adapt to complex transducer geometries and sizes.

Resistance strain gauges are very often used for measuring displacement, force, torque, pressure, acceleration, etc. The effect of elongation on the resistance of the metallic wire was first found by Lord Kelvin in 1856. The strain gauge is essentially a simple electrical resistor, specially designed to be easily bonded to the surface of a solid object and to undergo a change in resistance when a strain is present in the direction of its sensing grids. For uniaxial loading, there is a single non-zero principal stress in the direction of application of load and zero principal stress in all other directions perpendicular to the axis of load. The resistance R of the wire is

given by

$$R = \frac{\rho L}{A} \tag{7.1}$$

where ρ is the specific resistance or resistivity of the conductor material, L is the length of the conductor and A is its (uniform) cross-sectional area. Taking logarithm on both sides of the equation yields

$$\log R = \log \rho + \log L - \log A \tag{7.2a}$$

Differentiating the above equation,

$$\frac{dR}{R} = \frac{dL}{L} + \frac{d\rho}{\rho} - \frac{dA}{A} \tag{7.2b}$$

Since

$$A = \frac{\pi D^2}{4} \quad \text{and} \quad dA = \frac{\pi D dD}{2}$$

$$\frac{dA}{A} = \frac{2dD}{D} \tag{7.3a}$$

where D is the diameter of the conductor. Also

$$\frac{dD}{D} = -\nu \frac{dL}{L} \tag{7.3b}$$

where ν is Poisson's ratio. Hence

$$\frac{dA}{A} = -2\nu \frac{dL}{L} \tag{7.4}$$

Substituting for dA/A from Eq. (7.4) into Eq. (7.2), the following equation is obtained.

$$\frac{dR}{R} = \frac{dL}{L} + \frac{d\rho}{\rho} + 2\nu \frac{dL}{L} \tag{7.5}$$

Dividing the above equation throughout by dL/L

$$\frac{dR/R}{dL/L} = 1 + 2\nu + \frac{d\rho/\rho}{dL/L} \tag{7.6}$$

The term $(dR/R)/(dL/L)$ often denoted by F is called gauge factor and the term $(d\rho/\rho)/(dL/L)$ is called the piezoresistance effect of the material.

Requirements of a good strain gauge are as follows:

- High sensitivity.
- Good stability, repeatability and linearity over a large range of strains.

- Freedom from (or ability to compensate for) effects of temperature and other environmental conditions.
- Low 'drift' or 'null displacement' or 'zero shift', which is a change in the transducer's no-load output signal over time.
- Small size and mass.
- Ease of fixing to the test specimen.
- Ease of production over a range of sizes.
- Robustness.
- Suitability for static and dynamic measurements as well as remote recording.
- Low cost.

7.2.1 Electrical Wire Gauge

In an electrical wire gauge, a single length of wire is used as the sensing element. Strain gauge circuits used for measuring resistance changes impose certain restrictions on the minimum resistance that the strain gauge should have. Absolute resistance of a strain gauge relates to the need to initially balance the Wheatstone bridge (to be discussed in Sect. 7.3). Large resistance imbalances in one or more arms of the bridge may exceed the bridge balance capability of the strain indicator being used. The larger the gauge resistance, larger is the possible change in the resistance for a given gauge factor, lower is the current and less the heat dissipation at a given bridge excitation or in other words, higher is the signal-to-noise ratio (SNR) at a given power level. A resistance of the order of $60–1000\,\Omega$ is normally chosen for optimum performance and gauges are available with nominal resistances of 120, 350 and $1000\,\Omega$. To achieve these values, a grid pattern is formed, thereby increasing the length of the wire, and at the same time keeping the gauge length and width minimal.

The materials used for the fabrication of strain gauges should have a few basic qualities in order to achieve high accuracy, good sensitivity, stability, insensitivity to temperature, long life and the ability to operate under the required environmental conditions. Unfortunately, the most desirable strain gauge materials are also temperature sensitive and their resistance also tends to change with age. Though this may not be a matter of concern for tests of short duration, temperature and drift compensation must be provided for continuous industrial measurement. Materials with high specific resistance, low-temperature coefficient of resistance, constant and preferably high gauge factor and constant strain sensitivity over a wide range of strains are preferred. The most popular and widely used alloys are copper–nickel alloys and nickel–chromium alloys. Typical materials for strain gauge applications include Constantan (copper–nickel alloy), Nichrome V (nickel–chrome alloy), platinum alloys (usually tungsten), Isoelastic (nickel–iron alloy), or Karma (nickel–chrome) alloy wires, foils or semiconductor materials. The gauge factor typically varies between 2 and 4 for metallic wire strain gauges. Wire gauges can be classified into bonded and unbonded gauges and are fabricated in four basic varieties, viz. grid, wrap around, single wire and woven.

7.2.1.1 Bonded Wire Gauge

In this type of gauge, the strain gauge is bonded directly to the surface of the specimen being tested with a thin layer of adhesive cement, which serves to transmit the strain from the specimen to the gauge wires. The stress analyst has to ensure that the bonding is perfect so that the strains are fully transmitted from the surface of the object to the gauges, resulting in $\mu\Omega$ resistance changes. Such strain gauges are consumables; once bonded, they cannot be removed and re-bonded. The material used for backing and the bonding cement should offer a high insulation resistance and excellent transmissibility of strain from the base to the wires and must be immune to moisture effects. The most common materials used for wire strain gauges are Advance or Constantan alloys. Epoxy paper or bakelite is often used as the backing material to which the grid structure is bonded. The most popular wire gauge is the flat grid type where the wire is wound back and forth as a grid, as illustrated in Fig. 7.1.

The grids of these gauges with linear patterns are meant to measure strains in a single (longitudinal) direction. However, since the ends of each section of the wire are looped around, grids are also sensitive to the component of strain acting perpendicular to their length. Transverse strains also cause changes in resistance. Transverse sensitivity K_t is the ratio of the resistance change of the gauge produced by uniaxial transverse strain to the resistance change produced by uniaxial longitudinal strain of the same magnitude and is expressed as a percentage. In order to reduce transverse sensitivity, loop lengths in the transverse sections of the wire should be minimized or joined through a material having a lower sensitivity to strain. Fortunately, this sensitivity to transverse strain has been greatly reduced in foil gauges through the design of the grids. It is preferable not to have transverse sensitivity greater than 2% of the sensitivity along the major axis. The wire grid plane should be kept as close to the specimen surface as possible to achieve best transfer of strain.

Fig. 7.1 Wire strain gauge
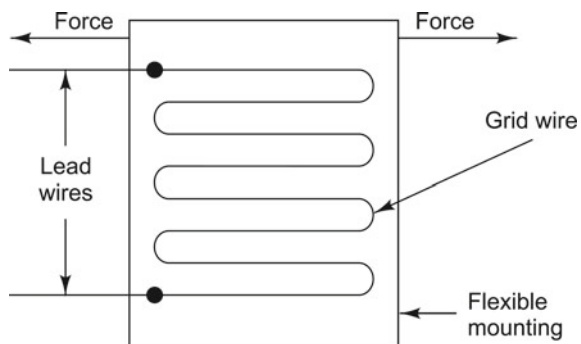
Fig. 7.2 Unbonded gauge
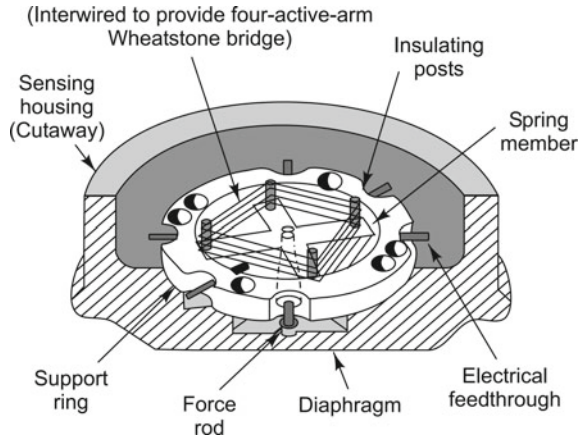
7.2.1.2 Unbonded Strain Gauge

The unbonded strain gauge (Fig. 7.2) consists of a free filament-sensing element and strain is transferred to the resistance wire directly without any backing. The sensing element is made by stretching relatively long strands of high tensile strength wire (typically 25 μm in diameter) around an array of posts that are linked to a diaphragm, such that it experiences relative motion when the diaphragm deforms under pressure. One of these posts is generally attached to a stationary frame and the other to the moveable frame of a linkage, which is designed such that when pressure increases, one half of the wire is stretched more than the other half.

The main applications of such strain gauges are in displacement, pressure and acceleration transducers. The primary advantage of unbonded strain gauges over bonded gauges is that they have a higher gauge factor, of the order of 3. Since no adhesives are required, they can be designed and fabricated for operation at higher temperatures. Owing to the radial symmetry of the transducer, spurious signals from transverse forces are cancelled. Besides, they have very low hysteresis and creep (due to the absence of backing and bonding). Unbonded strain gauge transducers, however, tend to be large.

7.2.2 Foil Gauge

The foil gauge was born out of a need to minimize the transverse sensitivity of a wire gauge. This gauge is basically an extension of the wire gauge, the main difference being in its construction. In a foil gauge, the required grid pattern is produced by printed circuit technique using a very thin foil of the same material as that used for the wire gauges, with a plastic backing. Owing to the larger surface area, this gauge has higher heat dissipation capability and better thermal stability. Besides, transverse

sensitivity is reduced in these gauges by making the perpendicular sections of the foil wide. For the strains to be faithfully transmitted from the specimen to the gauge, the foil diaphragm and the adhesive bonding agent must work together; the adhesive should also electrically insulate the foil grid from the surface of the specimen. The construction of these gauges ensures that there is no stress concentration at the terminals, thereby extending their life. Since photochemical etching processes are used for manufacturing these gauges, they can be made in any complex pattern or size, such as circular gauge or diaphragm gauge, as well as rosettes for two-dimensional stress analysis. Three element rosettes are also available and can be used to determine the biaxial state of strain. Figure 7.3a shows a linear foil gauge and Fig. 7.3b various types of rosettes, and a circular gauge.

Gauges are available in lengths of 0.25–50 mm, the most commonly used lengths being 5–10 mm. Smaller gauges tend to have degraded performance, since the strain

(a)

(b)
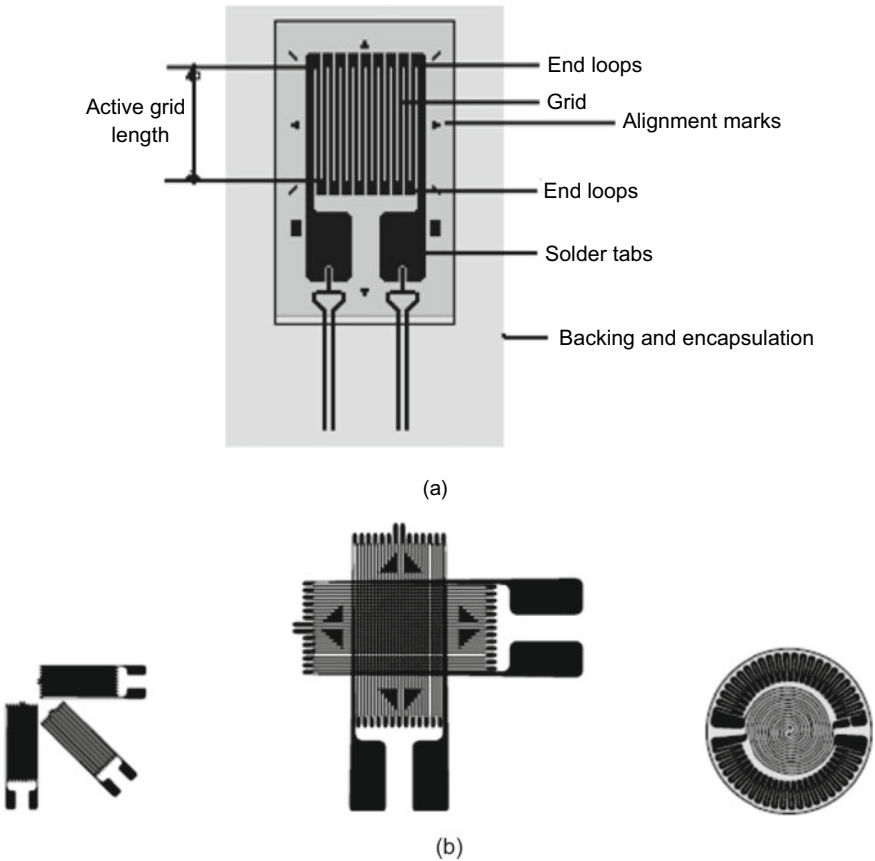
Fig. 7.3 Foil gauges: **a** linear foil gauge, **b** various types of strain gauges (Courtesy of http://www.omega.com)

is not uniform along the length; besides it is difficult to handle these gauges. Gauge factors of most metal foil strain gauges are well behaved: typically 95% (2σ confidence limits) of all gauges constructed have a gauge factor within the limits of the tolerance given. A gauge factor tolerance of $\pm 5\%$ indicates that if the indicated strain is 1000 μm/m, the actual strain lies between 995 and 1005 μm/m.

Strain Gauge Specifications

Nominal resistance: $120 \pm 0.4\,\Omega$.
Gauge factor: $2.0 \pm 5\%$.
Operating temperature range: -20 to $+100\,^{\circ}$C.
Maximum strain: 5% or 50,000 μm/m.
Active length: 6 mm.

7.2.3 Thin Film Gauges

Evaporated, thin film gauges have been receiving a lot of attention of late, especially for applications such as in diaphragm type pressure gauges. The force sensing element in these gauges which is subjected to stress is generally in the form of a metallic diaphragm on a ceramic layer. These gauges do away with the need for adhesive bonding of the grid on to the backing. Thin films of metals such as aluminium, gold, nickel, platinum, or palladium are made in desired patterns and molecularly bonded directly to a ceramic layer (substrate) by vacuum deposition (thermal evaporation of the desired material in vacuum) or sputtering techniques. The substrate is bonded to the test specimen in the same manner as that used for other gauges. The major contribution to change in the gauge factor in these gauges is due to resistivity changes, rather than geometrical changes. These gauges can operate over a very wide temperature range, say from -200 to $+400\,^{\circ}$C with good stability. The sensitivity is also high because of a high gauge factor and high gauge resistance. Because the thin film gauge is molecularly bonded to the specimen, it is highly stable and has very little drift in resistance. These gauges are rugged and have very low hysteresis and creep.

7.2.4 Semiconductor Strain Gauges

Important advances in strain gauge technology have been made through studies of the piezoresistive properties of silicon and germanium. These materials were studied in the mid-1950s by researchers at Bell Laboratories, USA and were found to have gauge factors more than 50 times, and sensitivity more than a hundred times that of metallic wire or foil strain gauges. Silicon wafers were also found to be more elastic than metallic ones, returning quite readily to their original shapes after being strained. They have a major drawback though, in terms of material non-linearity
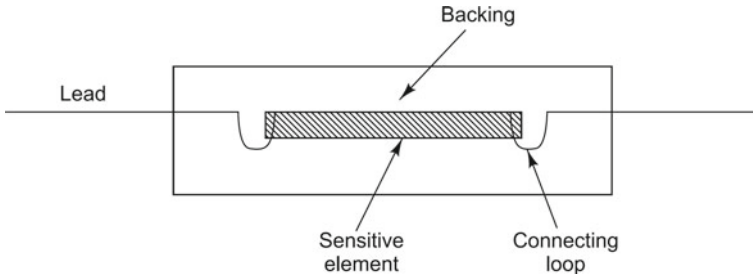
Fig. 7.4   Typical semiconductor gauge construction

(resistance-to-strain relationship is non-linear, varying 10–20% from a straight line equation) and temperature sensitivity. The first silicon strain gauges were developed for the automotive industry around 1970.

In the metal alloy strain gauges described earlier, the strain sensitivity is mainly due to the dimensional change, with a smaller contribution from the resistivity changes. Semiconductor strain gauges, on the other hand, exploit the piezoresistive effects of silicon or germanium and measure the change in resistance with stress as opposed to strain. The construction of a semiconductor strain gauge is different from that of the metallic strain gauge. Figure 7.4 illustrates the general construction of a semiconductor gauge.

The semiconductor gauge is essentially a wafer with the resistive element diffused into a substrate of silicon. It is fabricated from a single crystal of silicon and germanium using integrated circuit technology. It is made in the form of a single rectangular strip with lead wires and is held by a thin paperback with cement. A thin layer of silicon, typically 0.125×0.0125 mm in cross-section is the common transducer element for the semiconductor strain gauge. Effective lengths may range roughly from 1.25 to 12 mm. Essentially, the same type of backing, bonding epoxies and mounting techniques as those used for metallic gauges, are used for semiconductor gauges also. The main characteristics specified for a semiconductor gauge are filament material (p or n type silicon) and gauge factor (positive or negative), gauge length, gauge resistance, temperature coefficient, backing or encapsulation, bonding and lead geometry.

The piezoresistive effect of semiconductor gauges is due to the change of the sensor geometry resulting from an applied mechanical stress T. The following relation describes the geometrical piezoresistive effect on the gauge factor of a semiconductor gauge.

$$F = \frac{\Delta R/R}{\Delta L/L} = 1 + 2v + \frac{\Delta \rho/\rho}{\Delta L/L}$$
$$= \frac{\Delta R/R}{\varepsilon} = 1 + 2v + \pi E \tag{7.7}$$

where F is the gauge factor, ΔR is the change in gauge resistance in Ω, R is the resistance of the unstrained gauge element, ε denotes the strain, ΔL denotes the relative increase of the length of the element, L is the length of the unstrained gauge element, ν denotes the material-dependent Poisson's ratio, π is the coefficient of piezoresistance along the axis of the gauge and E is the Young's modulus of the gauge along its length. $\frac{\Delta\rho/\rho}{\Delta L/L}$ represents the piezoresistance effect of the material.

Typical Specifications

Gauge length: 5 mm.
Resistance: 200 Ω.
Gauge factor: $120 \pm 5\%$.
Strain limit: 6000 μm/m.
Working current <25 mA.
Working temperature: $<80\,°$C.
Temperature coefficient of resistance @ $32°$ C $< 0.15\%$.
Temperature coefficient of gauge factor @ $32°$ C $< 0.15\%$.

Advantages The major advantage of the semiconductor gauge is the high gauge factor, currently approximately 150, but which may be increased through further development. This represents a marked improvement in sensitivity compared to the 2–4 range exhibited by the ordinary metallic element. It is approximately 100 times more sensitive to strain than the metallic gauge. Other advantages include chemical inertness, freedom from hysteresis and creep effects, good fatigue life and a low transverse sensitivity. Besides, these gauges are smaller in size and are suitable for both static and dynamic measurements.

Disadvantages The increased sensitivity and higher unit resistance, however, are also accompanied by comparative disadvantages such as the following.

   (i) The output of the semiconductor is inherently non-linear with strain.
  (ii) Strain sensitivity is markedly temperature-dependent.
 (iii) The output of the semiconductor gauge tends to drift with time.
 (iv) The gauge is somewhat more brittle and fragile than the corresponding wire or foil element, though it can be bent to a radius as small as 3 mm.
  (v) The strain range of the semiconductor gauge is roughly limited from 3000 to 10,000 μm/m (depending on the specific gauge type) as compared to an upper limit of 100,000 μm/m for some metallic resistance gauges.
 (vi) The semiconductor gauge is considerably more expensive than the ordinary metallic gauges.
(vii) Because of the high sensitivity of the semiconductor element, the non-linearity of the simple Wheatstone bridge cannot always be ignored, as is normally done when conventional metallic element gauges are used. This may necessitate special computer-controlled instrumentation in which these drawbacks can be overcome through software compensation.

7.2.5 Error in Strain Gauge Measurements

- The potentially most serious source of error in strain gauge measurements is thermally induced apparent strain due to resistance changes. Datasheets normally provide plots of apparent strain versus temperature. Another source of error is gauge factor variation with temperature. For Constantan alloy gauges, gauge factor typically increases by about 0.01% per °C.

- Another problem is the hysteresis effect. The bonded strain gauge does not sense exactly the same strain as that experienced by the surface to which it is bonded. This gives rise to a hysteresis loop in a plot of relative resistance variation versus strain for cyclic loading. Hysteresis effect can be reduced by repeated cycling to a strain level higher than that required. Faulty bonding can also give rise to large hysteresis, which however cannot be eliminated by cycling.

- Strain gauges are susceptible to creep. If the backing material and the bonding cement are not stronger than the resistive element, the transmission of the strain to the gauge is not faithful, and the indicated strain is less than the true value, becoming increasingly non-linear as strain increases. Creep may be due to a faulty structure in the gauge itself, imperfect bonding or a rise in the operating temperature. It is generally large immediately after gauge installation. Unlike hysteresis, creep is time dependent and has a greater effect on static tests, as well as on small gauges with a small bonding area.

- Fatigue failure is another problem arising in dynamic conditions due to stress reversals. The weakest points in a gauge are the locations where the connecting leads are attached to it. Fatigue failure may be reduced by ensuring that there is no abrupt change in area at the joints.

- Moisture can also have a disastrous effect due to a change in the volume of the backing or bonding adhesives. It leads to a sharp fall in the insulation resistance between the gauge and the specimen.

7.3 The Wheatstone Bridge

DID YOU KNOW that Sir Charles Wheatstone, after whom the Wheatstone bridge is named was an English natural philosopher and inventor, recognized mainly for his contributions to the electric telegraph? At 14, he was made apprentice in the family business of his uncle Charles, who manufactured and sold musical instruments, but he spent more time reading, writing, publishing songs and learning about electricity and sound. At 16, he produced his first keyed musical instrument called the 'flute harmonique'. At 19, he exhibited his 'Enchanted Lyre' or Acoucryptophone, an instrument that appeared to play itself. Charles Wheatstone was one of the most inventive geniuses of his times;

he published scientific papers, applied for patents and did serious research in acoustics and optics. He invented both philosophical toys and scientific instruments. He had varied interests and contributed to the fields of linguistics, photography, optics, cryptography and acoustics, inventing electrical generators, typewriters, the electric telegraph, the concertina, the stereoscope and the polar clock. Despite his lack of a formal education in science, he was made Professor of Experimental Philosophy at King's College, London at the age of 34. There he invented an improved dynamo, the rheostat and an improved version of an electrical bridge invented by Samuel Hunter Christie and what is now known as the Wheatstone bridge. He continued to work at King's College for the remainder of his life, and also was involved in the family business.

The Wheatstone bridge is the signal conditioning device used for a variety of sensors with passive elements such as resistors, capacitors, and inductors. This bridge is a very simple circuit allowing measurements of impedances with very high accuracies. The operation of this bridge with DC excitation for resistive arms is discussed in the following section. For reactive arms such as those constituted by inductances or capacitances, a bridge with AC excitation may be used. The Wheatstone bridge measures resistance changes produced when active resistance strain gauges are exposed to strain. A typical bridge circuit consists of four 'arms', each consisting of an impedance element, which could be resistive, capacitive or inductive. A typical resistive Wheatstone bridge is shown in Fig. 7.5.

Some of the common terms pertaining to a Wheatstone bridge are as follows:

Terminals or **bridge corners** are indicated in Fig. 7.5 as A, B, C, and D; each of these may be either physically close to or far away from any of the others. A bridge corner represents the electrical junction of two bridge arms and either a voltage input or voltage output lead. Voltage input leads are denoted by P (for Power) and signal leads, S (for Signal).

Fig. 7.5 Resistive Wheatstone bridge
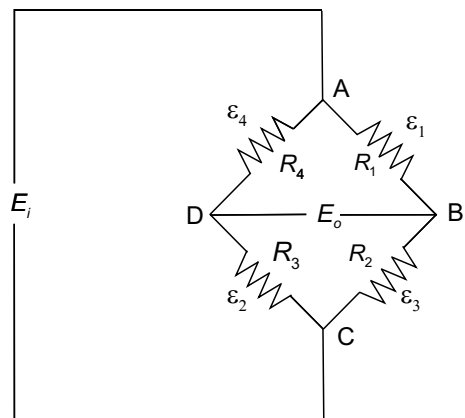
Bridge arm refers to all the resistances between two adjacent corners. It includes both the gauge as well as the lead wires in an active arm, or the bridge-completion resistor and the lead wires in an inactive or dummy arm. For example, in this figure, it could refer to resistance R_2 along with the lead wires.

Adjacent arms refer to any two arms that share a common corner, e.g., AB and BC or AD and DC.

Opposite arms refer to any two arms that do not share a common corner, e.g., AB and CD or AD and BC.

Adjacent corners imply any two corners that are separated by a single bridge arm, say A and B or B and C.

Opposite corners imply any two corners that are separated by two bridge arms such as A and C or B and D.

Branch refers to two adjacent arms across the bridge excitation voltage, e.g., AB and BC or AD and DC.

In most of the present-day strain gauge-based instrumentation, the strain gauges themselves form one or more arms of the bridge. The four impedance elements are arranged in the bridge such that there are two parallel voltage divider circuits, each divider consisting of two elements in series. The bridge has four terminals, A, B, C, and D, with two terminals for voltage input and the remaining two for voltage output. Bridge excitation voltage is applied to the circuit at terminals A and C which are the junctions of the two series circuits, and the output voltage is measured between terminals D and B, which are the junctions of the series elements in each parallel arm. A resistive bridge is provided with a DC excitation a few Volts in magnitude. Care should be taken to minimize the non-strain resistance changes in the gauges and all other conductors within the bridge circuit.

Consider the simple case when just two resistors R_1 and R_2 are connected in series across the constant voltage supply E_i. A portion of the supply voltage is dropped across R_1 and the remaining across R_2.

$$E_i = V_{R_1} + V_{R_2} \qquad (7.8)$$

When two more resistors R_3 and R_4 are also connected across the same voltage supply (Fig. 7.5), they too function as a voltage divider. Such a circuit is called a full bridge circuit. Depending upon the resistances of the four resistors, a differential voltage can also be present between points B and D. This voltage E_o can be measured with a good high impedance voltmeter and can be calculated as shown in the following paragraph.

The output voltage is proportional to the change in resistance; the output of the bridge is in fact proportional to $\Delta R/R$ so that for a given strain $\Delta L/L$ and gauge factor F, the output is independent of the initial value of the absolute resistance R in each arm of the bridge. The bridge output across terminals B and D is given by

$$\frac{E_o}{E_i} = \frac{R_1}{R_1 + R_2} - \frac{R_4}{R_3 + R_4} \qquad (7.9)$$

Corresponding to a non-zero input or bridge excitation voltage E_i, the output voltage E_o is zero when a condition of resistive bridge balance is satisfied, or when

$$\frac{R_1}{R_1 + R_2} = \frac{R_4}{R_3 + R_4} \quad \text{or when} \quad \frac{R_2}{R_1 + R_2} = \frac{R_3}{R_3 + R_4} \tag{7.10}$$

A change in resistance of any of the four gauges will unbalance the bridge, causing the output voltage E_o to change. For instance, output of the bridge with R_1 alone active is given by

$$E_o = E_i \left[\frac{R_1 + \Delta R_1}{R_1 + \Delta R_1 + R_2} - \frac{R_4}{R_4 + R_3} \right] \tag{7.11}$$

If all the strain gauges have the same initial resistance, then Eq. (7.11) becomes

$$E_o = E_i \left[\frac{R_1 + \Delta R_1}{2R_1 + \Delta R_1} - \frac{1}{2} \right] = E_i \left[\frac{\Delta R_1/R_1}{4 + 2(\Delta R_1/R_1)} \right] \tag{7.12}$$

Since the relative change in resistance of a strain gauge, i.e., $\Delta R/R$ is equal to the gauge factor (F) times the strain (ε) producing the change in resistance, the output of the bridge for a single active gauge will take the form

$$\frac{E_o}{E_i} = \frac{F\varepsilon_1}{4 + 2F\varepsilon_1} \tag{7.13}$$

However, in actual practice, a true resistive balance is seldom achieved. Rather a zero balance circuit like that shown in Fig. 7.6 is often externally applied to the bridge in order to eliminate the small initial imbalances by changing the ratio of the resistances in the two adjacent arms. Here R_c is a large fixed (balance limit) resistor and R_V is a variable resistor (across the power supply).

In actual applications, strain gauges may constitute the four arms of a bridge. Typically, the four gauges are chosen such that they are 'matched', i.e., they are of identical resistance values. This helps to easily calculate the output voltage of the Wheatstone bridge, although theoretically it is not necessary that the two series circuits be identical. Besides, this choice allows the bridge circuit to self-compensate for temperature changes and drift in sensor output. The measurement of output voltage is used to calculate the exact impedance of the variable elements, which can then be converted to strain. The bridge is initially balanced, or in other words its output is adjusted to a null or zero voltage state (by adjusting an impedance element). The change in resistance of one element will then induce an output voltage. For instance, if a strain gauge is subjected to a tensile load, its resistance will increase, resulting in a positive output voltage. Therefore, for stress analysis applications, it is the difference in the bridge output before and after the gauges are strained, that is of interest. When the initial resistances and the changes in resistances are known, Eq. (7.14) gives the bridge output after strain is applied.
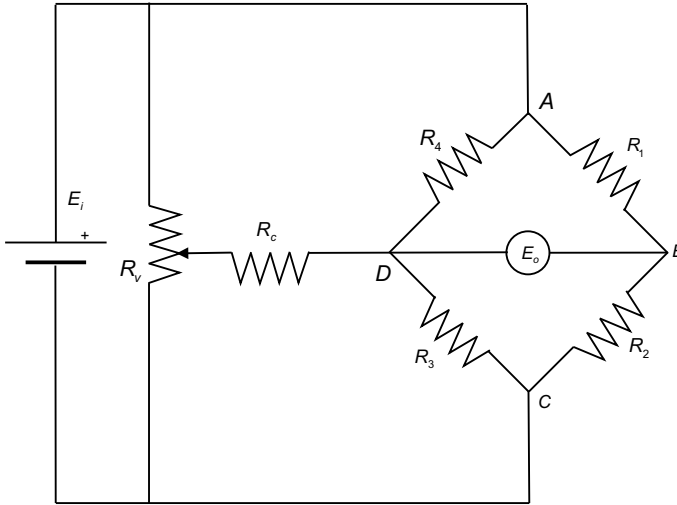
Fig. 7.6 Wheatstone bridge with external zero-balance circuit

$$E_o = E_i \left[\frac{R_1 + \Delta R_1}{R_1 + \Delta R_1 + R_2 + \Delta R_2} - \frac{R_1}{R_1 + R_2} \right.$$
$$\left. - \frac{R_4 + \Delta R_4}{R_4 + \Delta R_4 + R_3 + \Delta R_3} + \frac{R_4}{R_4 + R_3} \right] \quad (7.14)$$

Here, R_1, R_2, R_3, and R_4 are the initial resistances, prior to the application of strain and ΔR_1, ΔR_2, ΔR_3, and ΔR_4 are the resistance changes experienced by the bridge arms when the strains are applied to them. It is to be noted that an increase in either R_1 or R_3 will produce a positive change (increase) in bridge output voltage; conversely an increase in R_4 or R_2 will result in a decrease in bridge output voltage. This leads to the fundamental principle of the Wheatstone bridge: changes of resistance in adjacent arms (R_4 and R_3, for example) have a numerically additive effect on the bridge output when the changes are of opposite signs. When the changes in resistances in adjacent arms are of the same sign, they have numerically subtractive effects.

The important concepts regarding a Wheatstone bridge output that are worth remembering and that may be put to use to increase the bridge output for a particular type of loading when all four arms of the bridge are active strain gauges (subjected to strain) are listed below. These concepts can also be used to achieve temperature compensation.

 (i) Changes of resistance in adjacent arms, (e.g., R_1 and R_2) have a numerically additive effect on the output when the changes are of opposite sign.
 (ii) When changes in adjacent arms of the bridge are of the same sign, they have a numerically subtractive effect.
(iii) When changes in opposite arms, (e.g., R_1 and R_3) are of the same sign, they have a numerically additive effect.

(iv) When changes in opposite arms are of the opposite sign, they have a numerically subtractive effect.

The bridge output can also be expressed in terms of a constant current supplied to the bridge. With such an input, the output of the bridge can be obtained for any combination of input current and bridge arm resistances. For a bridge having a constant-current power supply, the bridge output can be calculated from the initial resistances and the strain-induced resistance changes as shown in Eq. (7.15).

$$E_o = I_i \left(\frac{R_1 \times R_3 - R_4 \times R_2}{R_1 + R_2 + R_3 + R_4} \right) \tag{7.15}$$

When the Wheatstone bridge is used for reactive arms such as capacitors and inductances, the DC excitation voltage is replaced by an AC excitation source. A typical inductive bridge used for measurements involving inductive accelerometers and linear variable differential transformers (LVDTs) is shown in Fig. 7.7. Here the L_s represents inductances.

Different bridge configurations are used, depending on the quantities to be measured. They are:

- Quarter bridge—One arm of the bridge is an active element.
- Half bridge—Two arms of the bridge are active elements.
- Full bridge—All four arms of the bridge are active elements.

In the quarter and half bridges, the arms of the bridge that are not made up of active elements are called dummies and consist of fixed elements that match the original unaltered values of the active elements. They are essentially used for completion of the bridge. In case the active gauge is placed in a high-temperature environment, it

Fig. 7.7 Inductive bridge circuit
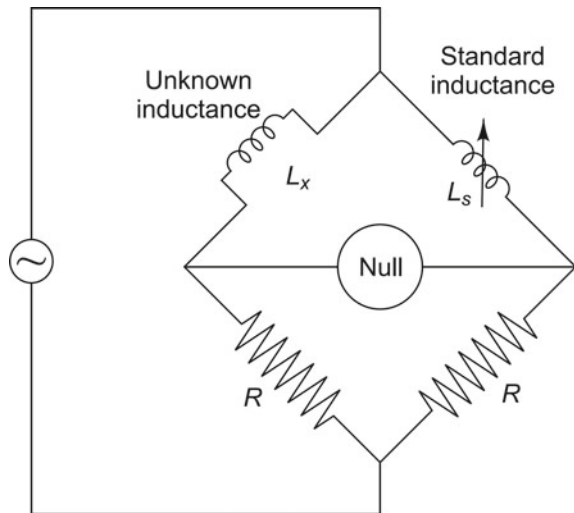
is required that the dummies also be placed in a similar environment to take care of temperature compensation.

The following are the errors that will have to be accounted for:

- *Random drift:* This is generally traceable to a poor strain gauge installation and is virtually impossible to correct after installation. This can be avoided by ensuring proper bonding between the structure and the gauge and avoiding leakage paths to the ground or between leads.
- *Strain averaging:* This is always present when the gauge is installed over a strain gradient. The steeper and more non-linear the gradient, the worse the error. This can be minimized by using a gauge with a smaller grid, but is never completely eliminated.
- *Thermal output:* This is a very dangerous and troublesome error because its magnitude is completely independent of strain. It can be eliminated by using self-temperature compensation techniques.
- *Gauge factor variation with temperature:* This is a very minor error in stress analysis and can usually be ignored.
- *Transverse sensitivity:* This is present any time the gauge is installed in a strain field for which the ratio of the transverse to longitudinal strains is other than -0.285. It is relatively minor for measurements made with longitudinal gauges, but data from rosettes usually require numerical correction because one grid is exposed to a high transverse strain.
- *Lead wire effects:* This error is due to strain gauges being mounted at a distance from the measuring equipment and the problem of varying lead wire temperature. Typically, any change in the lead wire resistance will be indistinguishable from changes in the resistance of the strain gauge itself. Special methods of wiring strain gauges to a bridge are employed to correct for lead wire effects to cancel at least a part of the effect. If the lead wire resistance exceeds 0.1% of the nominal gauge resistance, this source of error becomes significant. Therefore, in industrial applications, these errors are minimized by locating the transmitter directly at the sensor.
- *Wheatstone bridge non-linearity:* This is present in all quarter bridge circuits and some half and full bridge configurations. It amounts to a relatively minor error at strain levels normally encountered in experimental stress analysis, but must be carefully considered for large values of measured strains. Non-linearity is absent only when the resistance changes are such that the currents in the bridge arms remain constant, that is, when $(\Delta R_1/R_1) + (\Delta R_2/R_2) = 0$ and $(\Delta R_4/R_4) + (\Delta R_3/R_3) = 0$. The only configurations producing complete linearity are bridges employing equal and opposite arms, i.e., two fully active arms and four fully active arms.

7.3.1 The Instrumentation Carrier Frequency Amplifier

The instrumentation carrier frequency amplifier is widely used in instrumentation wherever mechanical signals from passive pickups of the resistive or inductive or capacitive type are used. Such transducers consist of strain gauge bridges, differential transformers, capacitive bridges and hence the carrier frequency amplifier serves as the signal conditioning amplifier for the measurement of mechanical quantities such as strains, torques, displacements, accelerations, forces and pressures. It can be used either for static strain or dynamic strain measurements or both or for dynamic measurements of accelerations, pressures, etc. A typical carrier frequency amplifier is shown in Fig. 7.8. It does the following functions: supplies AC excitation to the bridge, provides bridge completion for quarter and half bridge installations, zero balances the bridge, shunt calibrates the system of gauges, conditions and amplifies the signal from the bridge, sets the gauge factor for direct readout of the strain or sets the gain for scaling the output signals from the instrument. The oscillator produces a high-frequency signal and this carrier signal serves as the AC excitation voltage for the transducer. Its frequency should be at least 10 times the highest frequency of the measured input (physical) signal. Carrier frequency amplifiers used in vibration measurement usually have a carrier frequency of 5 kHz. They can easily measure vibration frequencies in the range of 0–500 Hz, i.e., 0–30000 revolutions per minute (RPM).
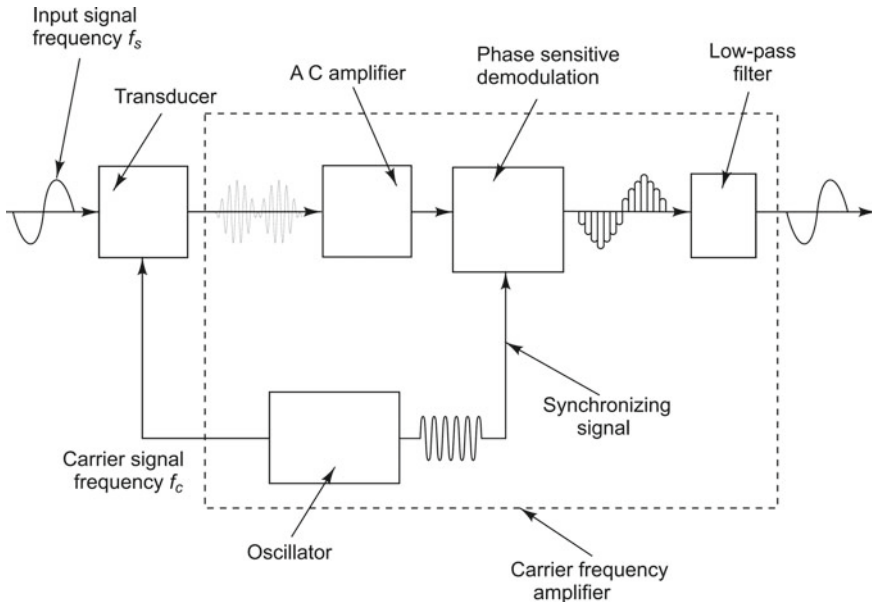
Fig. 7.8 Instrumentation carrier frequency amplifier

The amplitude of the carrier signal is modified by the amplitude of the input signal or in other words, the carrier signal is amplitude modulated by the input signal. If the input modulating signal is represented by the function $A_s \sin \omega_s t$ and the carrier signal by the function $A_c \sin \omega_c t$, then the modulated signal x_m is given by

$$
\begin{aligned}
x_m &= (A_s \sin \omega_s t)(A_c \sin \omega_c t) \\
&= \frac{A_s A_c}{2} \sin\left[(\omega_c - \omega_s)t + \frac{\pi}{2}\right] + \frac{A_s A_c}{2} \sin\left[(\omega_c + \omega_s)t - \frac{\pi}{2}\right]
\end{aligned}
\tag{7.16}
$$

The above equation shows that the modulated signal has two side bands corresponding to two frequencies $(\omega_c + \omega_s)$ and $(\omega_c - \omega_s)$. This amplitude modulated waveform is shown in Fig. 7.9. After amplification, the desired signal is extracted from

(a)

(b)

(c)
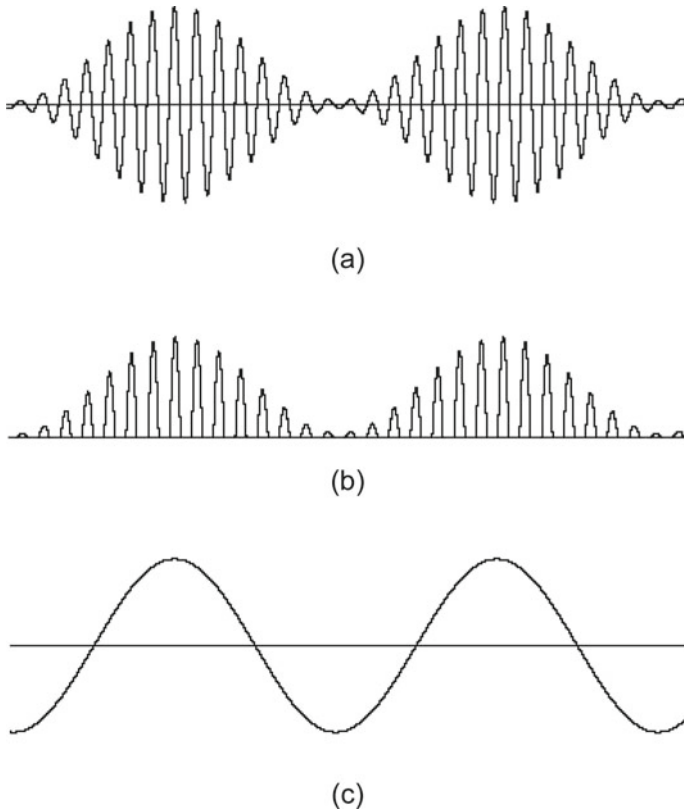
Fig. 7.9 Amplitude modulation: **a** amplitude modulated signal, **b** rectified signal, **c** demodulated signal

the modulated signal by using a demodulator and a low pass filter. All dynamic signals undergo some attenuation. Errors are generally small at low frequencies but can become very large if the frequency bandwidth of the carrier frequency amplifier is insufficient to accommodate the high frequencies present in the signal.

Typical Specifications of a Carrier Frequency Amplifier

Carrier frequency: 5 kHz ±1%.
Strain Gauges: 40–1200 Ω.
Inductive Transducer: 2–20 mH.
Usable frequency range: 0–500 Hz.

7.4   Transducers for Force Measurement

DID YOU KNOW that force transducers are used for static as well as dynamic measurements of forces? They are used to obtain static force versus displacement characteristics. In the context of vibration measurements, they are often used to measure the dynamic force produced by a shaker or impact hammer or for obtaining frequency response functions (FRFs) in modal analysis. Force may be measured either directly, by using an elastic device positioned in the line of action of the force, or indirectly in which method the elastic deformation of any of the force bearing members is measured to deduce the force. Due to the property of elasticity, the force results in a proportional reaction which can be measured either as displacement or as strain. The accuracy of measurement would depend on the accuracy of the sensor used. A load cell is a transducer that converts a force into an electrical signal; the force being sensed deforms a strain gauge through a mechanical arrangement. The strain gauge, in turn, converts the strain to an electrical signal using a Wheatstone bridge and an instrumentation amplifier.

7.4.1   Strain Gauge Load Cell

The load cell may be used for the measurement of static as well as dynamic forces over a wide range of forces with very high reliability. In this transducer, the output is obtained as the deformation of an elastic member with high tensile strength. The elastic column is typically made of homogeneous materials, generally steel alloys manufactured to very close tolerances. The main parameters to be considered in

designing a load cell are maximum force, relative size and shape, material density, modulus of elasticity, strain sensitivity, deflection and dynamic response. With proper design, a linear relationship between a dimensional change and measured force can be achieved. The materials used should result in low strain hysteresis over repeated loadings and very low creep over long periods of loading. The variation in the modulus of elasticity of the column with temperature, its ultimate strength and ease of fabrication are other parameters to be considered.

The column type strain gauge-based load cell is most popular for measuring unidirectional tensile or compressive forces. It consists of a hollow rod made of a material like high carbon steel with chromium and molybdenum or hardened stainless steel which has a large elastic limit and low hysteresis and creep. To facilitate fixing of strain gauges, the middle portion of the rod is made square in cross-section. When a load P acts on a cross-sectional area A, the stress developed is P/A. It is preferable to have a reduced cross-sectional area to obtain large measurable strains. This may be achieved by using a hollow cylinder (tubular column). However, slender columns are susceptible to buckling above a particular load. The cylinder length is hence chosen such that there is no buckling in the given range of loading. Typically, four strain gauges are bonded to the column. All four may be bonded such as to sense axial strains, or they may be bonded with two gauges sensing in the axial direction and two in the circumferential direction also called Poisson configuration. The second arrangement is more popular. Figure 7.10a shows the column load cell with gauges 1 and 3 measuring axial strains and gauges 2 and 4 measuring the circumferential or Poisson strains. The strain gauges are fixed to the column in Poisson configuration and the measured strains are:
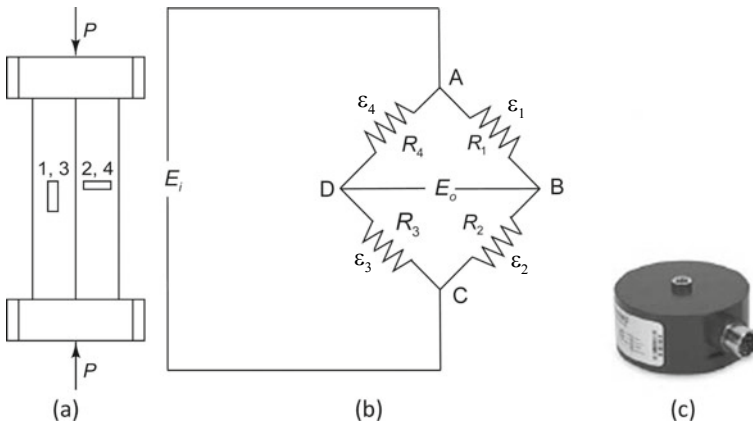
Fig. 7.10 Column load cell: **a** schematic drawing, **b** bridge circuit, **c** photograph of load cell (Courtesy of http://www.pcb.com)

$$\varepsilon_1 = \varepsilon_3 = \frac{\sigma}{E} = \frac{P}{AE}$$
$$\varepsilon_2 = \varepsilon_4 = -v\frac{P}{AE} \tag{7.17}$$

With a Poisson half bridge arrangement, the resulting bridge output is $(1 + v)$ times that of a quarter bridge with axial gauge alone where v is the Poisson's ratio of the column material. With a full Poisson bridge as shown in Fig. 7.10b the input-output relationship for the bridge (excluding resistance R_c is given by)

$$\frac{E_o}{E_i} = \frac{F\varepsilon_1(1 + v)}{2 + F\varepsilon_1(1 - v)} \tag{7.18}$$

where F is the gauge factor as before and E_o and E_i are the bridge output and input, respectively. Such a bridge offers better linearity than a bridge with axial gauges alone. Since all the four gauges experience the same temperature changes, there is temperature compensation inherently built into the bridge circuit. If the applied force is not concentric or if it is at an angle to the axis of the rod, there is a bending moment on the rod. To avoid any error due to this bending strain, the strain gauges are fixed at the centre of the faces. Under this condition, if bending takes place in the rod such that strain gauge 1 is subjected to tensile strain and strain gauge 3 to compressive strain, they produce zero voltage in the bridge circuit since they are connected in opposite arms of the bridge. Thus, the measured load in this transducer is insensitive to any bending load.

As already pointed out, the effect of temperature on the strain gauges is automatically compensated for in the bridge circuit. But to compensate for the effect of temperature on the Young's modulus of the material of the rod, a separate resistance R_c may be connected in series with the bridge as shown in Fig. 7.6. When the atmospheric temperature increases, Young's modulus decreases; therefore, for a given load, the strain increases, increasing the bridge output. This increase is cancelled by the increase in the value of the resistance of resistor R_c, which is kept in the vicinity of the active gauges. The voltage drop across R_c increases due to the increase in temperature; hence the bridge excitation voltage is reduced, which, in turn, reduces the output voltage of the bridge circuit to the extent it would have increased due to change in Young's modulus. Thus, inaccuracy of such load cells is reduced up to 0.1% of the full-scale reading. Strain gauge force transducers are best suited for static measurements. Figure 7.10c shows the photograph of a commercially available load cell.

Typical Specifications

Force: 1500 N Full scale.
Hysteresis: 0.1–0.05%.
Output: 2 mV/V.
Non-repeatability: 0.05%.

7.4.2   Piezoelectric Force Transducer

This transducer is discussed in this chapter though it does not come under the classi-
fication of strain gauge-based sensors for the simple reason that it is equally popular,
especially for smaller force ratings. A piezoelectric crystal can conveniently be used
in a force transducer since it has all the required characteristics. The piezoelectric
force transducer (Fig. 7.11) like the piezoelectric accelerometer works on the princi-
ple that the deformation of a piezoelectric crystal produces a charge or voltage output
proportional to the force acting on the crystal. However, unlike the accelerometer, the
force sensor does not have an inertial mass attached to the transducing crystal. Piezo-
electric force transducers are highly suitable for the measurement of tensile and com-
pressive forces, for which the crystal has to be physically elongated or compressed
for the generation of an output. The transducer responds to forces applied at one or
more points on the face of the crystal by producing a proportional charge across per-
pendicular faces. More than one piezoelectric crystal may be stacked (mechanically
in series, but electrically in parallel) to form what are called bimorphs (two elements)
or multimorphs (a piezopile consisting of multiple elements), enhancing the charge
sensitivity of the transducer in proportion to the number of elements stacked. For
modal analysis using shaker excitation, the shaker applies a force through a stinger
to the force transducer, which is, in turn, attached to the structure. The force applied
to the structure in the axial direction is that transmitted through the piezoelectric
crystal (measured) minus the force that is required to accelerate the base of the force
gauge. For excitation using an impact hammer, the transducer is fixed to the tip of the
hammer and gets compressed when an impact is applied to it. When selecting a force
transducer, the way in which the transducer interacts with the excitation device to
which it is connected, the exciter or hammer tip, is to be carefully considered, since
this can alter the force transducer calibration and cause distortion of the force signal
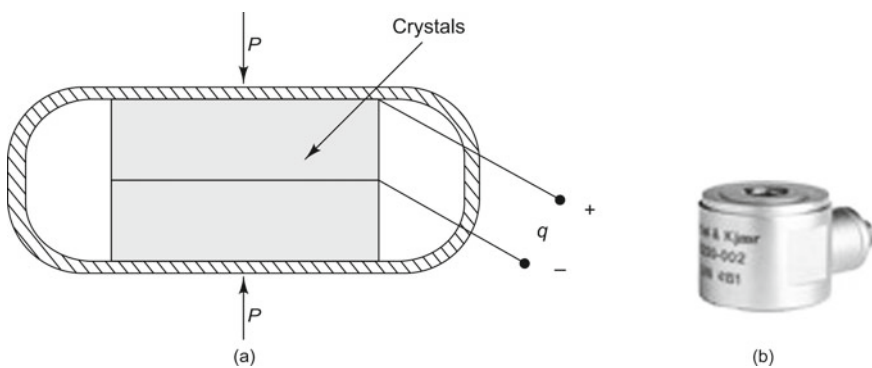
Fig. 7.11 Piezoelectric force transducer: **a** schematic, **b** photograph (Courtesy of Brüel & Kjær,
Denmark)

measured, especially at resonances. The mass of the transducer may also result in sensitivity to bending moments which is not desirable.

The force transducer can also be of the ICP type, ICP being PCB's registered trademark standing for 'Integrated Circuit-Piezoelectric' (discussed in Sect. 3.5.1) and finds wide applications in dynamic structural testing systems for modal analysis and for the prediction of structural response to steady state continuous and impact forces on small lightweight and delicate structures. It is widely used for the measurement of frequency response functions (FRFs) using a dual-channel analyser, in which case the force transducer is used to measure the input force and an accelerometer (or velocity transducer) is used to measure the response of the structure.

The major advantages of a piezoelectric force transducer are that it has good linearity, low weight, robust construction, as well as excellent and long-term stability. It is characterized by a compact design and is especially suited for applications where space is a premium. It can easily be attached to a stinger or hammer and can also be easily calibrated and mounted. It causes minimal mass loading as compared to other transducers and hence minimal change to the dynamic characteristics of a test structure. Compared to strain gauge sensors it provides excellent resolution. The lowest frequency to be measured depends on the signal conditioner used and is in the range 0.1–1 Hz. Piezoelectric force transducers are to be used in conjunction with regular charge amplifiers.

Typical Specifications

Sensitivity: 300 pC/N.
Maximum compression force: 2000 N.
Maximum tensile force: 300 N
Mass: 20 g.

7.5 Transducers for Torque Measurement

DID YOU KNOW that torque measurements are used in process control and monitoring torque is critical to the performance of axles, drive trains, gear drives, electric and hydraulic motors, gas and steam turbines? Torque and speed determine horsepower, which is an indication of system efficiency. There are several methods used for acquiring reliable torque data from rotating objects. The most widely used transducing element for converting torque into an electrical signal is the strain gauge. Other techniques for the measurement of torque use transducers mounted in the machine train or on the rotating shaft; most

of them also utilize strain gauges. Each method has its pros and cons and the best solution is obtained from a thorough understanding of the application. The most popular measurement techniques are: (i) strain gauging the shaft and (ii) using in-line torque transducers. The difficulty in both methods is getting power from the stationary supply to the gauges over the stationary/rotating gap and getting the output signal back to the stationary signal conditioning amplifier. Both contact and non-contact methods are in vogue to bridge the gap.

7.5.1 Static Torque Measurement

Reaction torque is the moment acting on an object which is not free to rotate. A strain gauge is bonded to the structural member that deforms when a torque is applied. The resulting deflection produces a stress, causing the gauge to change its resistance. A Wheatstone bridge may be used to convert the resistance change into a calibrated electrical output. A reaction torque sensor or static torque transducer is used when limited rotation will not damage the power supply and signal output cables. In the design of such a torque transducer, care must be taken to eliminate side loading (bending) and axial loading and to ensure that the transducer is sensitive only to torque loading. Such a transducer finds applications in calibration of torque wrenches, determining the breaking torque of tubes and aircraft fasteners and the opening torque of child-proof containers. Figure 7.12 shows a strain gauge-based static torque transducer.

Fig. 7.12 Static torque transducer (Courtesy of http://www.datum-electronics.co.uk)

7.5.2 Strain Gauge-Based Transducer

In this method, the existing shaft of the machine can be used by bonding strain gauges on to it. In other cases, a rotary torque transducer may be used. The strain gauge based sensor is nothing but a coupled strain gauged shaft that is incorporated into the existing driveline using a keyway, spline or flange shaft. With strain gauges bonded to the shaft, the shaft itself becomes the transducer. This method is suitable provided the applied torque induces at least 150 μm/m. The transducer should be calibrated before it is used. This is done by loading the shaft statically and noting the strains. This is relatively easy in small systems but becomes extremely cumbersome as loads and shaft sizes increase. One should also be careful in selecting the appropriate location for the strain gauges, in mounting them carefully and protecting them.

The strain gauge-based transducer measures the torsional shear strain in a shaft. This makes use of the fact that shear in a plane normal to the axis of the shaft is accompanied by tensional and compressional strains at $\pm 45°$ to the axis. For this reason, torque gauges similar to that shown in Fig. 7.13a are used. The two principal strains occur at $\pm 45°$ to the direction of pure shear strain produced by the torque. These strains are opposite in polarity, though their absolute values are identical. The magnitude of shear strain at $45°$ to the axis of the shaft is given by the equation

$$\varepsilon_{45°} = \frac{16T}{\pi D^3 E}(1 + \nu) \tag{7.19}$$

where T is the torque developed, E is the Young's modulus of the material of the shaft, ν is the Poisson's ratio and D is the diameter of the shaft. In order to eliminate the thrust and bending stresses possibly present in the shaft, a second pair of strain gauges
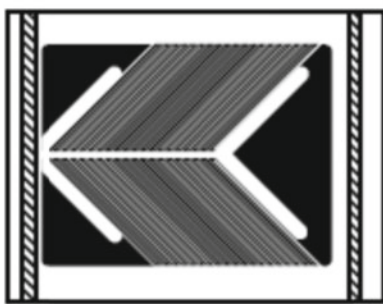
Fig. 7.13 Transducers for torque measurement: **a** torque gauge (Courtesy of http://www.omega.com), **b** rotary torque transducer (Courtesy of http://www.datum-electronics.co.uk)

may be pasted diametrically opposite to the first pair, on the other side of the shaft. These gauges additionally bring about higher sensitivity and provide temperature compensation. The four gauges are generally wired to form a full Wheatstone bridge. With a full bridge configuration consisting of two 45° gauges ε_1 and ε_2 as shown in Fig. 7.13a connected in adjacent arms of a full bridge and two gauges ε_3 and ε_4 bonded 180° away as mirror images of gauges ε_1 and ε_2, respectively, the input-output relationship is given by

$$\frac{E_o}{E_i} = \frac{G\gamma_{max}}{2} \tag{7.20a}$$

where

$$\gamma_{max} = |\varepsilon_1 - \varepsilon_2| = |\varepsilon_3 - \varepsilon_4| = \frac{32T}{\pi D^3 E}(1 + \nu) \tag{7.20b}$$

Figure 7.13b shows a strain gauge based rotary torque transducer. Many of these rotary transducers use standard RS232 digital data to transmit torque, giving fast and effective and easy to use data.

Range of Specifications Available

  Torque rating: 1–500,000 Nm.
  Shaft diameter: 5–1200 mm.
  Rotational speed: 0–20,000 RPM.
  Data sampling rate: 0–20,000 samples per second.
  Torque output: RS232/RS485/0–10 V/0–20 mA/±10 V/±20 mA.
  Shaft options: Flange/flange, keyway shaft, spline shaft.
  Accuracy: 0.1–0.05%.

7.5.2.1 Aids for Making Torque Measurements

The signal from the strain gauges bonded to a rotating shaft can be retrieved from the shaft using slip rings, rotary transformers, or wireless telemetry. Modern torque sensors and torque meters often have additional signal conditioning electronics and a data acquisition system. In the measurement of torque in machinery, three systems are typically employed to feed the excitation voltage to the strain gauge bridge network and to take the output signal: (a) direct contact measurement through slip rings, (b) rotary transformer (non-contacting), or (c) short range telemetry (non-contacting). In all the three cases, accuracy of the order of ±0.5% full scale is achievable. These techniques are discussed in detail in Sects. 7.7.2–7.7.4.

7.6   Pressure Transducers

DID YOU KNOW that a pressure transducer is a device that converts pressure into an analogue electrical signal? There are many types of pressure transducers, but one of the most common is the strain gauge based transducer. Pressure is converted into an electrical signal by the physical deformation of strain gauges bonded to the diaphragm of the pressure transducer. These gauges are wired into a Wheatstone bridge configuration. Pressure applied to the pressure transducer causes a deflection of the diaphragm and induces strain in the gauges. This strain, in turn, produces a change in electrical resistance proportional to the pressure. Pressure transducers are used in many automation applications.

Many electrical output pressure transducers detect pressure using a mechanical sensing element. These elements may consist of a thin-walled elastic member such as a plate or tube, which provides a surface area for the pressure to act upon. When the pressure is not balanced by equal pressure acting on the opposite side of this surface, the element deflects as a result of the pressure. This deflection is then used to produce an electrical output. When another different pressure is allowed on the other side of the pressure-sensing surface, the transducer will measure differential pressure. If the other side of the surface is evacuated well and sealed, absolute pressure is obtained. The transducer will measure gauge pressure if ambient pressure is allowed on the reference side.

Applications These transducers have been used successfully, not only for shock/ vibration testing but also in various critical applications including aerospace and nuclear testing. They have typically been used for model testing, wind tunnel and shock tube instrumentation, landing gear hydraulics, rocketry studies, ejection systems and cutting force studies. They have also found applications in combustion, explosion and detonation studies, gait analysis, engine testing, control systems, reactors, building structures, ship structures, auto chassis structural testing, shock and vibration isolation, dynamic response testing, and machine health monitoring.

7.6.1   Strain Gauge-Based Pressure Transducers

Although there are various types of pressure transducers, one of the most common is the strain gauge-based transducer. In this sensor, the force causing a pressure change gives rise to a change in the length of a member to which strain gauges are attached and hence a resistance change due to mechanical strain. Since the required deflection is small, the pressure sensing element may be either a diaphragm or a straight tube. If a tube is used, the strain gauges are bonded on it and it is sealed at one

Fig. 7.14 Strain gauge pressure transducer (Courtesy of http://www. senso-metrics.com)
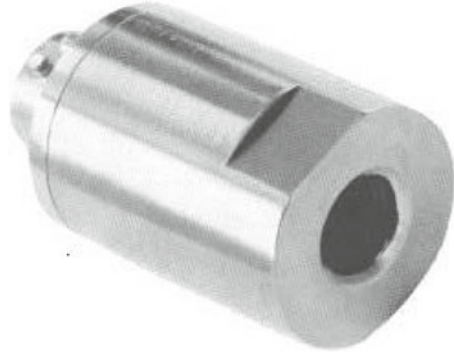
end. The pressure difference causes a slight expansion or contraction of the tube diameter. If a diaphragm is used as the sensing element, strain gauges are bonded onto the diaphragm. Pressure applied to the transducer causes a deflection of the diaphragm and introduces strain in the gauges. The strain in turn will produce an electrical resistance change proportional to the pressure. The gauges are wired into a Wheatstone bridge configuration with one to four active gauges. The types of strain gauge elements used in pressure transducers include unbonded metal wire gauges, bonded metal wire gauges, bonded metal foil gauges, thin-film deposited gauges and bonded semiconductor gauges. The advent of integrated circuit technology has led to the development of composite pressure sensors that are very easy to use. These devices often employ a semiconductor diaphragm on to which a semiconductor strain gauge and sensor for temperature compensation have been grown. Appropriate signal conditioning is also included in the integrated circuit, providing an output DC voltage or current that is linearly proportional to pressure over a specified range. The strain gauge pressure transducer is advantageous in that it offers solutions that meet varying accuracy, size, ruggedness and cost constraints. These sensors can be used for high- and low- pressure applications and can measure absolute or differential pressure. Figure 7.14 shows a typical strain gauge pressure transducer.

Typical Specifications

Pressure range: 0.08–1400 MPa.
Maximum temperature: 300 °C.
Error: ±0.1–1% of total dynamic range.
Resolution: Infinite.

7.7 Equipment that Facilitate Measurements on Rotating Structures

INTERESTING FACTS: A tachometer is one of the most common instruments used on rotating machines. The word tachometer has been formed using two Greek words 'tachos' meaning speed and 'metron' meaning measure and is used to denote a mechanical or electrical instrument employed to measure instantaneous values of running speeds in revolutions per minute. It is widely believed that the first mechanical tachometer was invented by the German engineer, Dietrich Uhlhorn in 1817 to measure the speed of machines. When he developed this device, little did he suspect that within two centuries, this device would be used as a standard feature in every vehicle produced. Uhlhorn's tachometer worked on the principle of a centrifugal governor, utilizing the fact that the centrifugal force on a rotating mass is a function of the speed of rotation and can be used to elongate or compress a mechanical spring. There are tachometers of the resonance, or vibrating-reed type, which use a series of consecutively tuned reeds to determine the frequency of vibration of a machine. Electrical tachometers are very popular today and are of several types. The eddy current tachometer is used in the speedometers of vehicles. In this device, a magnet is fixed on the rotating shaft and produces eddy currents that are proportional to the angular speed. Electric-generator tachometers generate either an alternating or a direct current proportional to the running speed. A stroboscope is an instrument that illuminates rotating objects to make them appear stationary so that they can be inspected; it can also be used to measure speeds knowing the flicker rate of the strobe light.

The need for determining the vibration and stress levels under actual operating conditions is felt for all equipment even at the design stage and they are obtained analytically. But an experimental validation, done after the prototype is ready, not only verifies the analytical method but also enables identification of failure-prone situations. Measurements on stationary components do not pose any difficulty since standard methods are available. However, such measurements on rotating components are cumbersome and pose difficulty in taking out cables from the rotating parts. The tachometer is the most commonly used device for directly measuring the rotating speed of a shaft. The stroboscope is a slightly more complicated piece of equipment used to measure rotating speeds of objects or to make them appear stationary for visual inspection. The slip ring unit comes in handy for vibration and strain measurement on rotating structures and uses contacting slip rings and brushes to take leads out from rotating members to external stationary equipment. The telemetry system is another option in which signals are transmitted from transducers on rotating machines to stationary measuring/analyzing devices.

7.7.1 Stroboscope

The stroboscope utilizes the property of persistence of vision. The human eye responds so slowly to light stimuli that it cannot differentiate between two light impulses reaching the eye within 0.1 s of each other. Impulses reaching the eye in quick succession are observed by the eye as a continuous unbroken sequence caused by this temporal aliasing or the 'stroboscopic effect'. Joseph Plateau and Simon von Stampfer simultaneously and independently invented the stroboscope, which is a device exploiting this property of the eye to make objects with periodically repetitive motions appear to be slow moving, or stationary, thus enabling them to be observed more conveniently. Commercially available stroboscopes are capable of operating at speeds between 600 and 20,000 rpm.

The stroboscope is used for measuring speeds of rotation of shafts and other parts of machinery without stopping the machine. It is also used for the study of reciprocating and oscillating motion of structures and machine parts and in the study of stresses in machinery in motion and is very useful in machinery diagnostics. The stroboscope is also used to measure car engine speeds; such a stroboscope does not have an independent oscillator to control the flash frequency; instead it uses the ignition timing of the engine and is actually meant for measuring the phase angle, (i.e., the timing of the spark relative to the position of the camshaft) rather than the frequency. This device is also used as a diagnostic aid in medicine to view the vocal cords. Flashing strobe lamps are also used in discotheques where they give the impression of dancing in slow motion. In high-speed photography, very short flashes of light are used as a means of producing still photographs of fast-moving objects, such as bullets in flight. The stroboscope, however, cannot be used where the ambient light intensity is high, since it requires a subdued surrounding light for its efficient operation.

7.7.1.1 Mechanical Disk-Type Stroboscope

A mechanical disk type stroboscope consists of a whirling disk attached to a motor, the speed of which can be varied and measured. In its simplest form, a rotating disk with a single opening or a number of evenly spaced holes is placed in the line of sight between the observer and the rotating shaft. A reference mark on the rotating shaft is observed through the openings in the rotating disk. The rotational speed of the disk is adjusted so that it becomes synchronized with the movement of the observed shaft, making the mark on the shaft appear to be stationary. When this happens, the shaft speed is equal to that of the rotating disk, or some integral multiple of this speed and can be expressed as

$$\text{Shaft speed} = \frac{(\text{disk speed}) \times (\text{number of openings in the disk})}{\text{number of images}} \qquad (7.21)$$
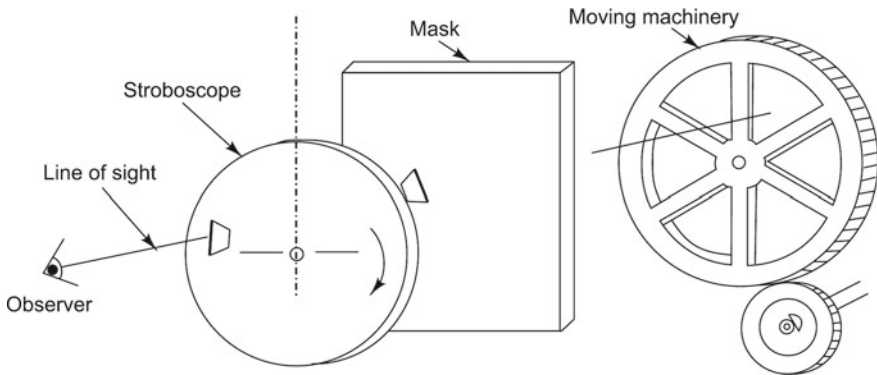
Fig. 7.15 Mechanical disk-type stroboscope

The number of images denotes the number of times a single mark on the shaft appears when viewed through all the openings in the disk when the images appear to be stationary. Figure 7.15 shows a mechanical type disk stroboscope.

7.7.1.2 Electrical Stroboscope

In an electrical stroboscope, the perforated disk is replaced by a neon gas discharge lamp, capable of emitting brief and rapid flashes of light. The rate of flashing is controlled by a variable frequency electronic oscillator, the frequency of which can be read off a dial/display attached to its tuning control. In operation, these flashes of light are directed on to the rotating, reciprocating, or vibrating member and the rate of flashing is adjusted to be equal to the frequency of motion of the target, causing it to appear stationary. A repetitive motion occurring 2400 times a minute will appear to stand still if viewed against repetitive light flashes occurring 2400 times per minute or at 40 Hz. If the frequency of the flashes is less than that of the rotating member or shaft, then the light will flash before the shaft can make a complete revolution, and it will appear as though the mark on the shaft is slowly rotating backwards. Similarly, if the frequency of the light flashes is greater than the shaft's rotational rate, the mark will appear to turn forwards. Figure 7.16 demonstrates the use of an electrical stroboscope. Let us consider the following cases:

Case (i) Single mark on the shaft: A single mark on the rotating shaft gets illuminated in the same position for each flash and hence appears to be stationary when the time between flashes is equal to the time for one rotation, or in other words, when the flashing frequency f_f equals the rotational frequency f_r, i.e., $f_r = f_f$. However, it is to be noted that the mark will also appear to be stationary if the shaft speed is an integral multiple of flashing frequency, i.e., $f_r = n f_f$ where $n = 1, 2, 3, 4, \ldots$, etc. Thus, a single mark on a shaft rotating at 2400 rpm would complete two revolutions between flashes if these are at the rate of 1200 rpm. The shaft would appear to be
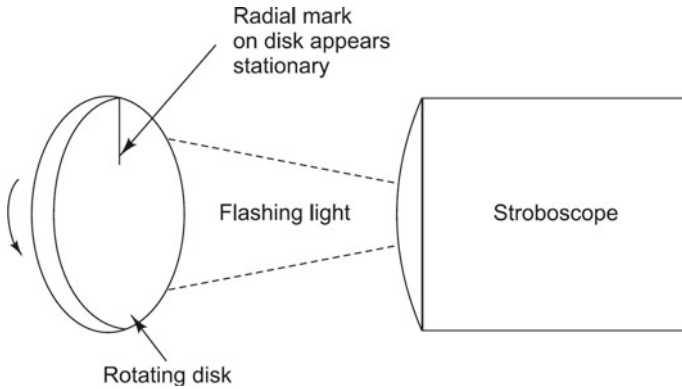
Fig. 7.16 Flash light stroboscope
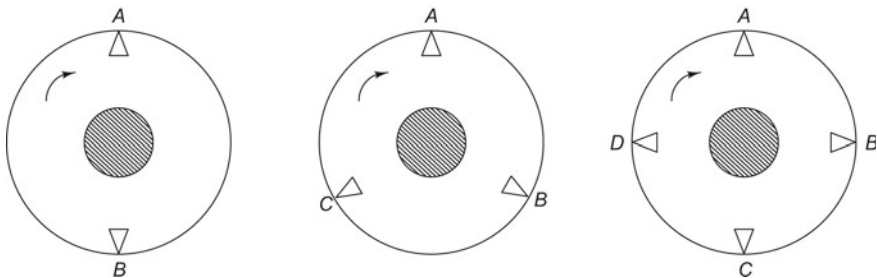
Fig. 7.17 Multiple images for a single mark

stationary at 2400 flashes per minute, and also at 1200, 800, 600, etc., flashes per minute.

However, when the rotational frequency is less than the flashing frequency ($f_r < f_f$) and is a submultiple of f_f, multiple stationary images would be obtained as shown in Fig. 7.17. Let us assume that a flash occurs when the mark is at A. If $f_r = f_f/2$, the next flash will occur after half a revolution, (i.e., when the mark is at B), and again at A when one revolution has been completed. Subsequently, the cycle would be repeated. So the mark will be illuminated at two positions 180° apart causing a double image to appear. With $f_r = f_f/3$, the mark will be illuminated at three positions 120° apart causing three stationary images to appear. This argument can be extended for cases $f_r = f_f/4, f_f/5, \ldots$, etc.

Case (ii) Multiple marks on the shaft: When a number of identical and equally spaced marks are made on the shaft, a stationary pattern is obtained when $f_r = f_f, 2f_f, 3f_f, \ldots$, etc., and also for certain values of f_r which are smaller than f_f, i.e., $f_r = f_f/N, 2f_f/N, 3f_f/N, \ldots$, etc., where N is the number of marks made on the shaft.

7.7.2 Slip Ring Unit

When transducers are required to be mounted on rotating parts of machines, some provision must be made to bring excitation power to the transducer from the stationary power source and to take the output signal away to the stationary signal conditioning amplifier and/or recording device. If the measurements can be completed over a few rotations, the connecting wires can simply be allowed to wind or unwind on the rotating shaft. This is convenient for slow speed machines. When the relative motion involved is small, continuous flexible conductors (often in the form of light coil springs) can be used. However, for measurements during continuous high-speed operation, slip rings, radio telemetry, or some form of magnetic coupling between rotating and stationary parts are required.

The slip ring unit has a wide range of applications, allowing use of inductive and resistive transducers, strain gauges, thermocouples, as well as other transducers. It is an electromechanical device that allows the transmission of power and electrical signals from a stationary to a rotating structure and is similar to the brushes and commutators found in many types of DC motors. It is also called a rotary electrical joint, collector or electric swivel, and can be used in any electromechanical system that requires unrestrained, intermittent or continuous rotation while transmitting power and/or data. It can simplify system operation and eliminate damage-prone wires dangling from moveable joints.

The main parts of the slip ring head are a stator with brushes and a rotor with rings; the arrangement is such that it helps easy mounting. Figure 7.18a shows two pairs of circular slip rings, one pair for providing excitation voltages (either AC or DC) to the bridge on the rotating shaft and the other for retrieving the signal from strain gauges to the stationary signal conditioning and data acquisition systems. The rotor consists typically of 6 or 12 hard slip rings that are conductive circles or rings or bands mounted on the rotor shaft and insulated from it, while the stator has two brush arrays 180° diametrically apart. The rotating slip rings are generally made of or coated with gold, silver, rhodium, or other noble metals and alloys. The stator brushes run in contact with the ring, transferring electrical power or signals to the exterior, non-rotating part of the system. Block-type brushes often are of sintered graphite, while wire-type brushes are alloys of platinum or gold. These are the best-suited materials for the application, ensuring a very small amount of wear, as well as high transmission quality. In order to maintain a high signal-to-noise ratio and good accuracy, the dimensional tolerances, surface hardening of the slip rings and the contact resistance of the brushes are to be carefully considered. The stator is equipped with 6 or 12 solder tags, corresponding to the number of slip rings, for the connection of cables. Each ring is adjacent to the next along the centreline, somewhat like the threads on a bolt; electrical connections from transducers on the rotor or shaft are made to the ring. The rotor part of the slip ring can be fixed either directly or via a mounting plate to the front end of the shaft. The brushes can be engaged even with the shaft running or can be lifted afterwards by just turning a ring, enabling easy replacement of worn brushes. Slip ring brushes, as well as the
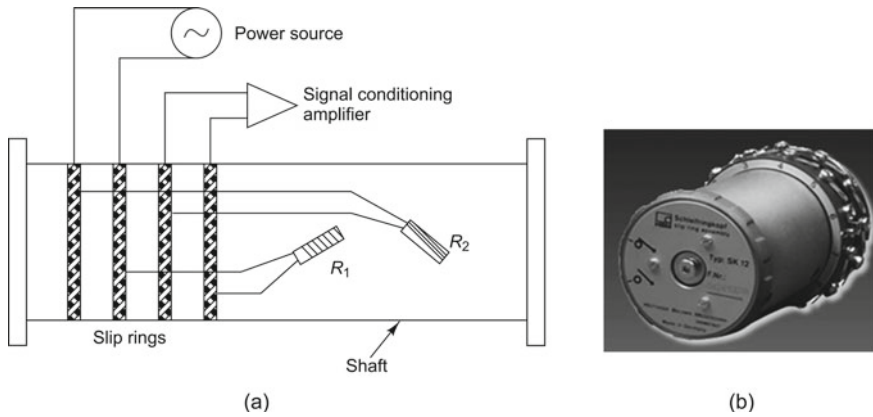
Fig. 7.18 Slip ring unit: **a** slip rings, **b** photograph (Courtesy of HBM, Germany)

support bearings internal to these transducers undergo a lot of wear and tear. The slip ring assemblies have to be fitted such that rough dirt, dust, humidity, oil, solvents and gases which could influence the slip ring transmission or reduce the resistance to ground of a connected measuring point cannot reach the slip rings or get into the case. In a typical industrial application, maintaining an oil-free slip ring is difficult. Figure 7.18b shows a typical slip ring assembly.

An important point to be considered in using slip rings for transmission of low-level instrumentation signals is the electrical noise produced at the sliding contact. A part of the sliding noise is due to thermocouple action if the brushes and rings are of different materials. The other part of the noise is due to random variation of contact resistance caused by surface roughness, vibration, etc. It is possible to reduce this noise by filtering, but this is inconvenient, and not always very effective. The better remedy is to ensure that the thermo-voltage generated between slip rings and brushes is very small, that the contact resistance between slip rings and brushes is very small ($<1 \, \text{m}\Omega$), and that the variation of contact resistance with speed of rotation, temperature, current, etc., is small. A high-quality miniature sliding slip ring may exhibit a contact resistance variation of the order of $0.05 \, \Omega$ peak to peak and $0.005 \, \Omega$ RMS. When the rings are used with strain gauge circuits, particular care must be taken, since the resistance variation of the sliding contact may be of the same order as the small change in resistance of the strain gauge to be measured. In such cases, it is advisable to use the full bridge circuit, as against the half or quarter bridge configurations, so that the sliding contacts can be taken out of the bridge circuit, leading to minimal influence from the slip ring resistance variations. For very precise measurements, the $700 \, \Omega$ strain gauge full bridge circuit is preferred; however, even with 350 or $120 \, \Omega$ full, half, or quarter bridges, the errors will be comparatively small. For the most demanding applications, more complex schemes are available to reduce noise to even lower levels.

Disadvantages

(i) Slip rings have a limitation in terms of the number of simultaneous channels of data that can be collected. The larger the number of channels, the bigger the slip ring becomes, until at around 30 channels, the entire unit becomes very unwieldy. If engineers want to collect dynamic data from a large number of strain gauges, they have to perform two or more test runs. For high-speed data transfer or for data transfer in electromagnetic interference (EMI) sensitive environment, fibre optic rotary joints (FORJs) are available.

(ii) Static charges are another source of problem and disturbances from the same may be got rid of by connecting rotor and stator housing earth with machine earth.

(iii) Another problem is wear and tear of brushes and rings; this can be minimized by engaging the two only during the time of actual measurements.

(iv) Besides, difficulty is faced in applications requiring speeds above 100,000 RPM; extreme care is required because of heating and vibration problems.

Typical Specifications

Number of slip rings: 6 or 12.
Resistance between slip ring and brush <40 mΩ.
Fluctuation of resistance <2 mΩ.
Thermo-electric voltage measured between two brushes when the slip rings are short-circuited (after running until warm) $<10\,\mu$V.
Current limit: 2 A.
Permitted speed for continuous duty: 1–6000 RPM.

7.7.3 Measurements Using Telemetry System

Slip rings have been relatively satisfactory for measurements on rotating machines; however, they have the drawback that the measurement necessitates extension of the rotating shaft and a variety of other modifications and hence their use is not always feasible for in-situ measurements. The use of the short range radio-telemetry (meaning distant measurement) technique, where the data are acquired through wireless transmission of radio waves, is more reliable and convenient for measuring parameters such as vibration, strain, torque, etc. in rotating machines. The sensor (which could be a strain gauge, a thermocouple or an accelerometer) is fixed on the rotating shaft (Fig. 7.19a) and is powered by a battery mounted on the shaft itself. The sensor output modulates a radio frequency (RF carrier) emitted by a miniature shaft-mounted transmitter to a rotating loop antenna. A concentric stationary antenna placed very close (less than 50 mm away) to the rotating antenna induces power in the loop antenna on the rotating shaft and picks up the signal and sends it to a stationary receiver, as shown in Fig. 7.19b.

The receiver receives the modulated signal from the stationary antenna and demodulates it to retrieve the information and display it. The received signals may be in the form of amplified analogue DC/AC voltages and may be analyzed using oscilloscopes or precision voltmeters or Fast Fourier Transform (FFT) analysers/signal analysers or can be recorded on a multichannel instrumentation tape recorder for later analysis. Besides offering an easy solution, this method provides completely noise-free signals with good accuracy and resolution. No additional signal conditioning of the received signals is generally required. The schematic block diagram of a typical telemetry system is shown in Fig. 7.19b.

The telemetry system generally comes with the following subsystems.

- A static strain transmitter for static strain and torque measurements.
- A dynamic strain transmitter for dynamic strain measurement using strain gauges and piezoelectric transducers, as well as vibration accelerometers.
- A piezoelectric module is used along with the dynamic transmitter as charge converter while using vibration accelerometers and other piezoelectric sensors.
- A signal receiver is used for amplification of the received input and for demodulation of the transmitted signal to obtain calibrated strain/ vibration data.
- Batteries for power supply to the transmitter. A power on/off switch is also generally provided and is useful in tests where time interval between assembly and test is quite long.

7.7.3.1   Factors to be Considered in Designing a Telemetry Test

Considerable effort is required in designing a telemetry instrumentation setup. Some of the important points to be considered before carrying out a telemetry test are as spelt out in the following section. With appropriate selection of transmitters, telemetry instrumentation can be utilized to collect signals from various sensors viz. strain gauges, vibration accelerometers, piezoelectric crystals and thermocouples. These sensors are to be wired to the transmitter with appropriate instrumentation cables. The cables are to be routed along the surface of the rotating component and bonded properly to it so as to prevent them from flying out during rotation. Miniature frequency modulated (FM) transmitters, along with battery have to be mounted on the rotating component. The following points are to be considered:

 (i) The sensors mounted on the structure should have sensitivity large enough to give a measurable signal output. They should not create any significant stress or imbalance in the rotating structure. Hence, low weight accelerometers are preferred for vibration measurement and foil type resistance gauges or semiconductor strain gauges are preferred for analyzing strain signals.
 (ii) Neither the sensors nor the transmitters should cause any static or dynamic imbalance in the rotating machine; hence they are generally mounted as diametrically opposite pairs.
(iii) The battery and the transmitter used for these applications typically have limiting static acceleration levels of the order 10,000 or 15,000 g. If large centrifugal
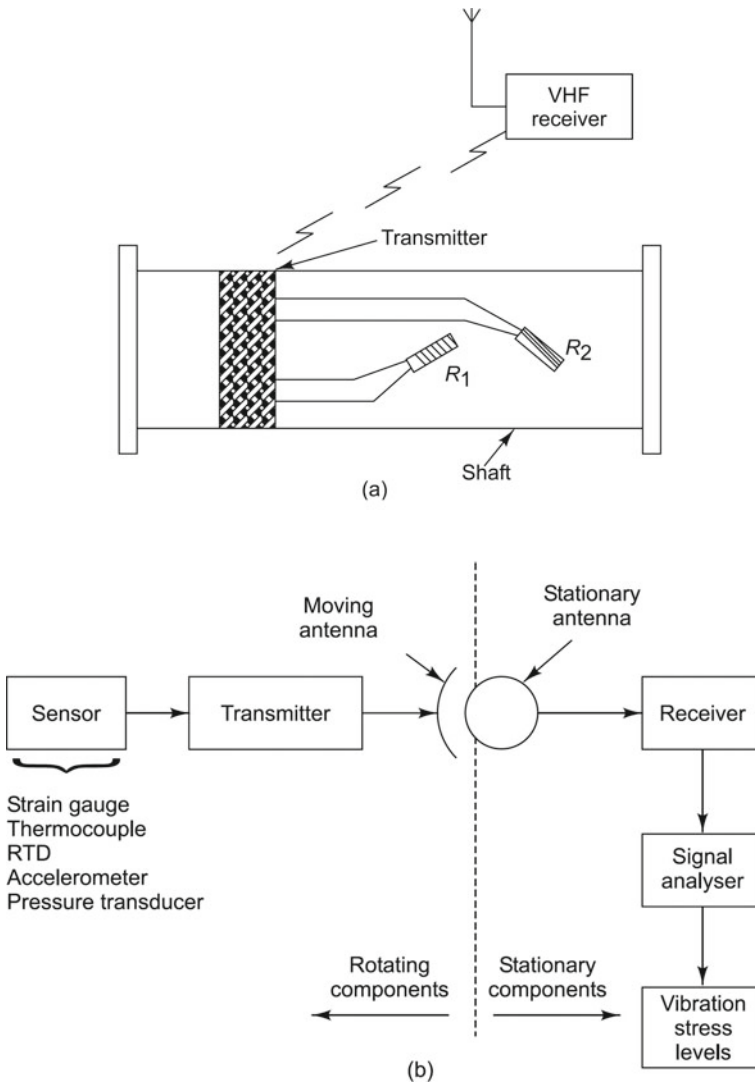
Fig. 7.19 Telemetry set up: **a** sensor and transmitter on rotating shaft, **b** block diagram

forces are expected, as when these devices are located at large radii or when the speed of the machine is high, a very strong transmitter housing is required. The cables/wires also have to withstand large centrifugal forces. Hence it is preferable to locate the transmitter at locations with the minimum radii from the shaft centre. The fixture and the housing should be light in weight, but at the same time strong enough to withstand the centrifugal forces generated during the operation of the machine.

(iv) There should be provision to mount bridge completion strain gauges and balancing circuits for strain gauge measurements.

 (v) The transmitter and the batteries can typically operate up to a maximum temperature of 125 °C. However, when it is required that measurements be carried out at higher temperatures, these units are to be mounted at locations where temperatures are lower, and the sensor wiring has to be routed suitably for connections to the transmitter.

(vi) Both antennas have to be a little flexible for ease of mechanical installation and should be adjusted for maximum coupling so as to obtain maximum induced power and received signal strength.

The telemetry system employs capacitive coupling to transmit the signal from the rotating antenna to the stationary antenna placed at a distance not exceeding 50 mm. In cases where the shafts are of small diameter (less than 200 mm), the antenna design is quite simple with a single loop of wire, properly anchored on the rotating shaft. For shafts of large diameters, more complicated antenna schemes are usually required, such as properly designed dipole antennae, matched to the transmitter output impedance.

7.7.4   Non-contact Rotary Transformer

A rotary transformer (Fig. 7.20) can be employed instead of the slip rings or telemetry system described above. This is especially beneficial when it is required to elimi-
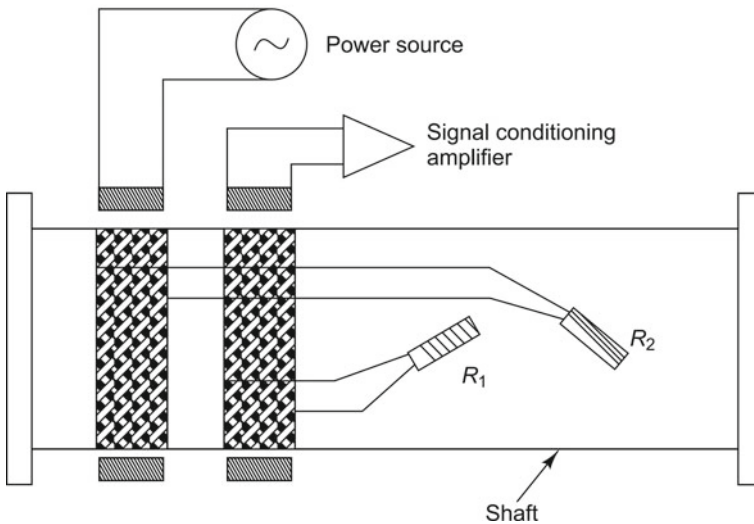
Fig. 7.20   Rotary transformer

nate the noise due to contact vibrations. The system comprises two circular rotary transformers, one to induce the AC bridge supply to the strain gauge bridge and the other to pick up the output signal, as illustrated in Fig. 7.20. The excitation signal is normally a carrier of 20 kHz. The transformers consist of a pair of concentrically wound coils which are suitably aligned for maximum mutual induction, with a high permeability core to improve efficiency.

Many of the transducers and equipment described in this chapter will have to be used for the measurement of strain in the experiments described in Chap. 12.

List of Symbols

Symbol	Meaning	SI unit
A	cross-sectional area	m^2
A_c, A_s	amplitudes of carrier and input (modulating) signals	m
dA	change in area	m^2
dL	change in length	m
dR	change in resistance	Ω
$d\rho$	change in specific resistance or resistivity	$\Omega.m$
D	diameter of conductor, shaft	m
E	Young's modulus of material	Pa
E_i	Wheatstone bridge input voltage	V
E_o	Wheatstone bridge output voltage	V
f_f	flashing frequency	Hz
f_r	rotational frequency	Hz
F	gauge factor	
G	modulus of rigidity	Pa
I_i	input current to Wheatstone bridge	A
K_t	transverse sensitivity	
L	length of conductor	m
L	inductance	H
N	number of marks on shaft	
P	load	N
R	resistance	Ω
R_1, R_2, R_3, R_4	resistances placed in arms 1, 2, 3 and 4 of Wheatstone bridge	Ω
R_c	balance resistor	Ω
R_v	variable resistor	Ω
T	torque	N.m
V_{R_1}, V_{R_2}	voltage drop across resistors R_1, R_2	V
x_m	modulated signal	
$\gamma_{\max}$	maximum shear strain	m/m

Symbol	Meaning	SI unit
ΔL	change in length	m
$\Delta R, \Delta R_1, \Delta R_2,$ $\Delta R_3, \Delta R_4$	change in gauge resistances	Ω
$\varepsilon, \varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$	strains	m/m
ν	Poisson's ratio	
π	coefficient of piezo-resistance	Pa^{-1}
ρ	specific resistance or resistivity of conductor material	$\Omega.m$
ω_c, ω_s	frequencies of carrier and input (modulating) signals	rad/s

Abbreviations

AC	alternating current
DC	direct current
EMI	electromagnetic interference
FFT	fast Fourier transform
FM	frequency modulated
FORJ	fibre optic rotary joint
FRF	frequency response function
ICP	integrated-circuit piezoelectric
LVDT	linear variable differential transformer
RF	radio frequency
RMS	root mean square
RPM	revolutions per minute
SNR	signal-to-noise ratio

Questions

1. What is meant by gauge factor of a resistive strain gauge?
2. What is a bonded wire gauge? What are its drawbacks?
3. What properties affect the resistance of a strain gauge?
4. What are the most important applications of resistance gauges?
5. How is temperature compensation achieved in a strain gauge bridge circuit?
6. What is meant by transverse sensitivity of a resistive strain gauge? How can it be reduced?
7. What does a carrier frequency amplifier do? Where is it used?
8. Why is a strain gauge called a passive transducer?
9. What is meant by the piezoresistive effect of a material?

10. What are the requirements of a good strain gauge?
11. Can strain gauges be peeled off and reused?
12. Where are unbonded strain gauges used?
13. What are strain gauge rosettes?
14. What are the main advantages of foil type gauges?
15. How do you choose the length of a strain gauge for a particular application?
16. What are thin film gauges and where are they used?
17. What signal conditioning device is required for a piezoelectric load cell?
18. What full bridge arrangements are available for the measurement of axial strain in a load cell?
19. What are semiconductor strain gauges?
20. How can you use a piezoelectric load cell for measuring FRFs?
21. How does a piezoelectric load cell work?
22. Give applications where torque measurements are required.
23. What is meant by an ICP type transducer?
24. What is the most common form of bridge arrangement in strain gauge based load cells?
25. What advantage does a strain gauge based load cell have over a piezoelectric crystal-based load cell?
26. What are the most common sources of error in strain gauge measurements?
27. Name two commonly used materials used for manufacturing strain gauges.
28. How can you measure static torques?
29. Why are strain gauges used in Wheatstone bridges and not in stand-alone mode?
30. What is the most commonly available resistance value in strain gauges?
31. What is the principle of operation of an electrical resistance gauge?
32. How does a strain gauge-based pressure transducer work?
33. What is meant by torsional strain?
34. Which strain gauge bridge configurations offer the best linearity?
35. How do absolute and differential pressure transducers work?
36. How can you achieve thermal compensation in a Wheatstone bridge circuit?
37. What is a torque strain gauge?
38. Up to what speeds can you measure using a stroboscope?
39. What is a stroboscope? How is it used for measuring the speed of a rotating component?
40. Can a stroboscope be used for visual monitoring of a rotating component?
41. Where are pressure transducers used?
42. What is a slip ring unit?
43. What are the parts of a slip ring unit?
44. What is meant by persistence of vision?
45. Up to what speeds can slip ring units be used?
46. What is a telemetry unit?
47. What are the main components of a telemetry unit?
48. What is meant by gauge factor of a strain gauge?
49. What do you mean by arms of a Wheatstone bridge?
50. What are the factors to be considered in fixing transducers on rotating machines?

51. Why are two antennas required in a telemetry unit?
52. What is meant by balancing of a Wheatstone bridge?
53. What do you mean by adjacent arms and opposite arms of a Wheatstone bridge?
54. Of what materials are slip rings and brushes made? Why?
55. What sort of sensors can be used in a Wheatstone bridge?
56. When are DC and AC excitation used for Wheatstone bridges?
57. What is the principle of operation of a telemetry unit?
58. What is meant by a quarter bridge?
59. When resistive changes in opposite arms of a Wheatstone bridge are of the same sign, do they have a numerically additive or subtractive effect?
60. What special precautions have to be taken in a radio-frequency telemetry arrangement?
61. What is a rotary electric joint?
62. What precautions should be taken for dummy gauges in a strain bridge circuit?
63. What is a half bridge?
64. What precautions should be taken in the choice of strain gauges for a Wheatstone bridge?
65. What is the difference between a static and a dynamic signal?
66. Can a strain gauge be used for finding natural frequencies and mode shapes? Explain with reasons.
67. Which bridge configuration has the most non-linearity?
68. Why is a carrier frequency of 5 kHz common in most carrier frequency amplifiers?
69. What is a dummy gauge?
70. Which strain gauges have the highest gauge factor? How high?
71. How can you reduce random drift in strain readings?
72. What is meant by amplitude modulation process?
73. How is power got from the stationary supply to the strain gauges bonded on a rotating machine?
74. How does a mechanical disk type stroboscope work?
75. How can you reduce the effect of bridge output variation due to lead wire effects?
76. When resistive changes in adjacent arms of a Wheatstone bridge are of the same sign, they have a numerically subtractive effect. State whether true or false and explain why.
77. Why are balance limit resistor R_c and variable resistor R_V used in strain gauge bridges?
78. How are strain gauges arranged on a torque transducer?
79. When changes in opposite arms are of the opposite sign, they have a numerically subtractive effect. State whether true or false and explain why.
80. Why are full Poisson bridges preferred to other bridge forms in axial load cells?
81. How many channels of data can commercially available slip ring units provide?
82. What is meant by a full bridge circuit?
83. What is meant by strain averaging?

84. Bridge non-linearity is absent only when the resistance changes are such that the currents in the bridge arms remain constant. State whether true or false and explain why.
85. How do you check if a strain gauge has been properly bonded?
86. In order to achieve proper demodulation in a carrier frequency amplifier, what can you state about the modulating and carrier frequencies?
87. Why is a galvanometer used in a Wheatstone bridge?
88. How will a Wheatstone bridge be different for resistive and inductive elements?
89. How is transverse sensitivity minimized in a foil type strain gauge transducer?
90. How do you decide on the specifications of the strain gauge to be used for a particular application?

Fill in the Blanks

91. Sensitivity of a bridge circuit is increased if _____.
92. Two common errors in strain measurement are _____ and _____.

Choose the Correct Answer and Give Justifications

93. Bridge sensitivity can be increased if (a) similar strains are put in adjacent arms, (b) opposite strains are put in opposite arms, (c) similar strains are put in all four arms, (d) opposite strains are put in adjacent arms.

Bibliography

1. Beeby, S. P., Beeby, S., Ensel, G., Kraft, M., & White, N. M. (2004). MEMS mechanical sensors, Artech House Inc., Reprint.
2. De Silva, C. W. (2005). *Mechatronics: An integrated approach*. Boca Raton: CRC Press.
3. Doebelin, E. O. (2004). *Measurement systems: Application and design*. Boston: McGraw-Hill Professional.
4. Dyer, S. A. (2001). *Survey of instrumentation and measurement*. New York: Wiley.
5. Fraden, J. (1996). *Handbook of modern sensors*. New York: Springer.
6. Gatti, P., & Ferrari, V. (1999). *Applied structural and mechanical vibrations theory, methods and measuring instrumentation*, eBook. Taylor & Francis.
7. Khazan, A. D. (1994). *Transducers and their elements*. Englewood Cliffs: PTR Prentice Hall.
8. Measurements Group Inc., Experimental stress analysis, Notebook, Raleigh, NC 27611.
9. Murty, D. V. S. (2004). *Transducers and instrumentation*. New Delhi: Prentice Hall of India Pvt. Ltd.
10. Nakra, B. C., & Chaudhry, K. K. (2004). *Instrumentation measurement and analysis*. New Delhi: Tata McGraw-Hill.
11. Prudenziati, M. (1994). *Handbook of sensors* (Vol. 1, Sec. IV, pp. 189–206). New York: Elsevier Science.
12. Rangan, C. S., Sarma, G. R., & Mani, V. S. V. (1983). *Instrumentation devices and systems*. New Delhi: Tata McGraw-Hill Publishing Company Ltd.
13. Stefanescu, D. M. (2011). *Handbook of force transducers: Principles and components*. Springer.

14. Tse, F. S., & Morse, I. E. (1989). *Measurement and instrumentation in engineering: Principles and basic laboratory experiments*. New York: Marcel Dekker.
15. Window, A. L. (1992). *Strain gauge technology*. Berlin: Springer.
16. Webster, J. G. (2000). *Mechanical variables measurement: Solid, fluid, and thermal*. Boca Raton: CRC Press.
17. http://www.bksv.com.
18. http://www.hbm.com.
19. http://www.omega.com.
20. http://www.pcb.com.
21. http://www.sensorland.com.
22. http://www.senso-metrics.com.
23. http://www.sensorsportal.com.

Chapter 8
Fundamentals of Signal Analysis

8.1 Introduction

Generally, the overall magnitude of a vibration or acoustic signal which is obtained using any one of the transducers we have discussed is alone not sufficient, but often spectral amplitudes or statistical descriptors are required. Historically, people have studied experimental data in the time domain using oscilloscopes, strip chart recorders, etc. The observer then had to rely on his or her powers of observation to make interpretations from the time domain data and this was difficult. For diagnostics, it is important to have information regarding the salient frequency components of the vibration signal which depend on the number of revolutions per minute (RPM) it makes. Simple $1 \times$ RPM information can be got using tachometers or stroboscopes or eddy current probes responding to key phasor information. However, for comprehensive information on the spectral content of the signal being measured, one needs more sophisticated instruments. The advances in electronics and signal processing have led to the development of very efficient real-time analysers/spectrum analysers/signal analysers/fast Fourier transform (FFT) analysers, which give frequency information. Using these analysers, complex vibration waveforms, often referred to as vibration signatures, can be broken down into their spectral components, and defined as a sum of harmonic functions of discrete amplitudes, frequencies, and phases.

This chapter throws light on several aspects of signal analysis which involves representation, transformation, and manipulation of signals: for example, separation of two signals, i.e., signal from a combination of signal plus noise, changing from one domain to another, say from time to frequency domain or vice versa, or obtaining statistical parameters from a signal. Signal processing is extensively used these days in machinery diagnostics and in applications in acoustics including the study of jet noise, traffic noise, machinery noise, sound navigation and ranging (SONAR), underwater acoustics, and instrumentation. This chapter also discusses the issues to be dealt with while carrying out dynamic measurements using a signal analyser, whether it be dedicated or personal computer (PC)-based. The various terms used

in signal processing have been dealt with, without too much mathematical encumbrance.

8.2 Various Steps in Data Acquisition and Processing

INTERESTING FACTS: A data acquisition system (DAS or DAQ) is a device that converts analog electrical signals from sensors measuring physical phenomena into digital format, and then with appropriate software enables data visualization, storage, and analysis. When testing with electronic devices was just evolving, scientists and engineers had to note down readings from metres and later analyse them. Later came data recording devices which were paper-based strip chart recorders which gave hard copies of the time domain signals. Subsequently, when technology further improved, it became possible to record data using magnetic tape recorders. These were quite popular for recording scientific data even in the 1980s and may be said to be the precursors of today's DAQs. IBM came out with the first dedicated data acquisition machine in 1963; it was called IBM 7700 Data Acquisition System. In 1964, they released the 1800 Data Acquisition and Control System which used a disc for storage. In 1981, they marketed the personal computer (PC)-based data acquisition system. Since then, more and more advanced DAQs started proliferating the market. These systems also became smaller, even for a large number of input channels. Today, we have multi-channel miniature, modular, stackable systems with a footprint of the order of 30 mm by 30 mm. These have very high performance and can store a large amount of data.

Most signals in raw form are continuous-time signals, but all subsequent analysis is done on digital signals due to the extensive availability of digital signal processing software/hardware. The first step in any measurement is data collection and for this a transducer which converts the physical/mechanical quantity to a proportional electrical quantity is used. The conversion may happen in one stage as in the case of a thermocouple, where a change in temperature is converted to a proportional change in voltage; or it may be done in many stages, as in the case of an inductive vibrometer, where the change in displacement is converted first to a change in inductance, and subsequently to a change in voltage using a Wheatstone bridge arrangement. The next stage is data recording, which may be done using either a dedicated signal recording device such as an instrumentation tape recorder (still used for dedicated application) or a digital recorder or more commonly by PC-based recording. The third stage involves editing the data, converting it to the required format and pre-processing it before doing any actual processing to extract the required sig-

nal descriptors. This is followed by data qualification since the nature of the signal dictates the analysis techniques to be used. Before analysing a signal, it is required to see if the signal is harmonic or periodic or random. If it is random, checks to see if it obeys stationarity can be done by observing if all data records have the same mean, root mean square (RMS), etc. The last and final step is data analysis, which involves working on individual records and on multiple records. Depending on the requirement, one might want to obtain the spectrum/cepstrum/probability density histogram (PDH)/statistical averages, etc. For many of the present-day applications, real-time operation is often desirable, i.e., the output has to be computed at the same rate at which the continuous-time signal is sampled.

8.2.1   Data Acquisition Systems

Data acquisition systems, as the name implies, are used to collect data, record them, and analyse them. They have a wide range of measurement applications, including collection, characterization, monitoring, and control of data. The data acquisition process has become considerably simplified these days through advances in electronics and has become more accurate, versatile, and reliable. Data acquisition systems typically acquire measured physical parameters such as displacement, velocity, acceleration, temperature, pressure, flow, or sound pressure level or take a specific action based on the data received. A wide range of data acquisition components and solutions are available in the market, ranging from simple recorders to PC plug-in cards to sophisticated mainframe computer systems. The specific requirements of any application will dictate the resolution, accuracy, number of channels, and speed requirements for a data acquisition system. Such a system can also include a multiplexer, which is used for signal switching in multi-channel applications. Generally, in the multiplexed mode, only one channel is connected at a time to the measuring device, with break-before-make switching (i.e., the input is disconnected before a new input is connected). Figure 8.1 shows a typical data acquisition and processing system for a system identification application.

8.2.2   Analogue and Digital Signals

While most of the data obtained using transducers is typically analogue, the processing on these signals is done after digitizing them as mentioned earlier. Hence, it is required to know what these signal forms imply. A signal $x(t)$ is said to be an analogue signal if it is continuous along both x- and y-axes. Here, the x-axis is typically time or space and the y-axis could represent any mechanical or electrical quantity. A signal is said to be a discrete-time signal, if it is discrete along the x-axis only and continuous along the y-axis. In order to convert a signal from continuous time to discrete time, a process called sampling is used; this involves measuring the
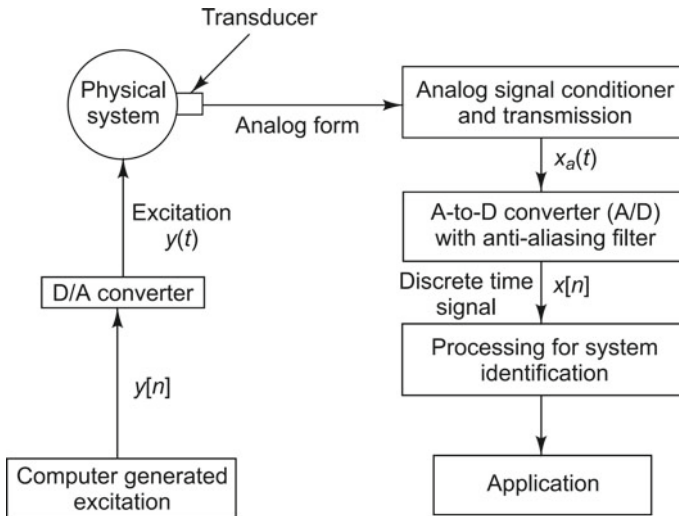
Fig. 8.1 Data acquisition and processing

value of the signal at regular intervals of time and each value is referred to as a sample. The discrete-time signal is thus related to the analogue signal from which it was obtained by sampling. The original analogue signal can be reconstructed from its discrete-time counterpart if the sampling was done in the first place without aliasing (as discussed in Sect. 8.6) and if the sampling frequency is known. A digital signal is one that is discrete along both the x- and y-axes and is generally called a sequence, since it depicts on the x-axis sample number, which is a dimensionless quantity. Figure 8.2 depicts all three kinds of signals. In a digital signal, the variable on the x-axis is made discrete by sampling, and that on the y-axis by a process called quantization (discretization of amplitude). The notation $\{x(n)\}$ is used to represent a complete sequence and $x(n)$ the nth sample of the sequence; however, the latter is often used to denote the sequence $\{x(n)\}$ itself. Typical sequence operations are addition, multiplication, shifting (delaying/advancing), etc..

8.2.3 Analogue-to-Digital Converter (A/D Converter or ADC)

The ADC is the heart of most data acquisition systems and is an electronic device that converts analogue voltages acquired from transducers into their equivalent digital form (in binary or hexadecimal code) representing the quantized amplitude values closest to the actual values of the inputs. There are several types of A/D conversion techniques; they can generally be divided into two types: integrating and non-integrating; the former measures the average input value over a defined time interval, thereby rejecting many types of noise, while the latter samples the instantaneous value
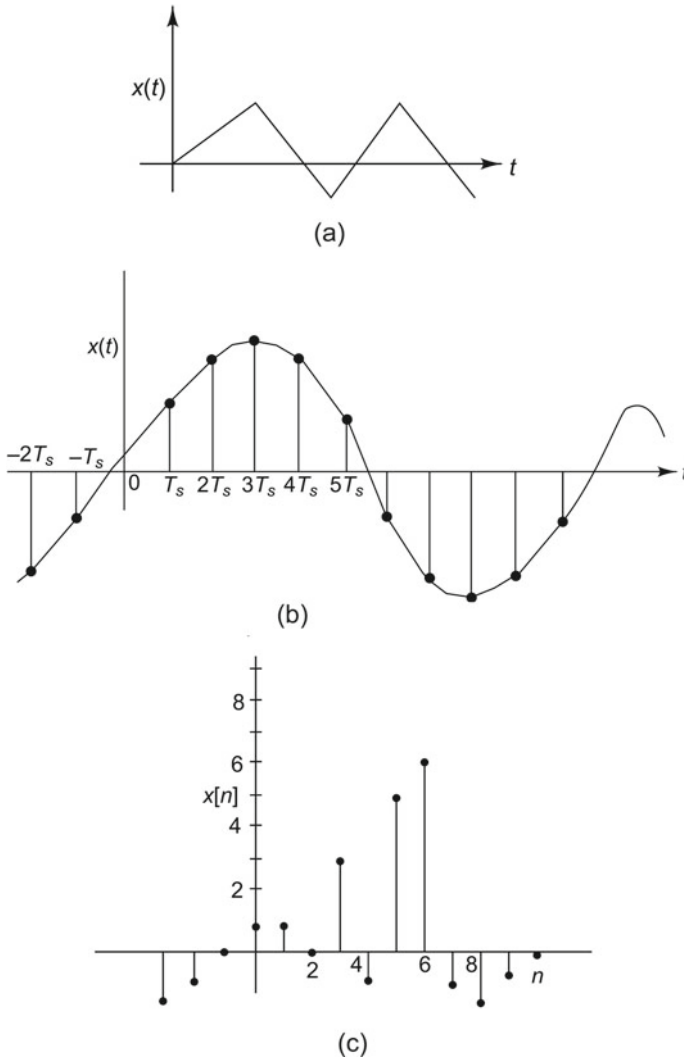
Fig. 8.2 Various types of signals: **a** analogue signal, **b** discrete-time signal, **c** digital signal

of the input (including noise) during a very short time interval and converts it to digital form. One of the simplest forms of ADC is one with uniform sampling, i.e., sampling the analogue signal $x_c(t)$ at regular intervals of time with a clock rate T_s or with sampling frequency $f = 1/T_s$ as shown in Fig. 8.3a. The output of the sample and hold circuit $x_0(t)$ is a staircase waveform and $x[n]$ is a digital signal as shown in Fig. 8.3b. The sample-and-hold circuit is designed to sample $x_c(t)$ as instantaneously as possible and to hold the sample value as nearly constant as possible till the next sample is taken. Typical issues involved in analogue-to-digital conversion are the number
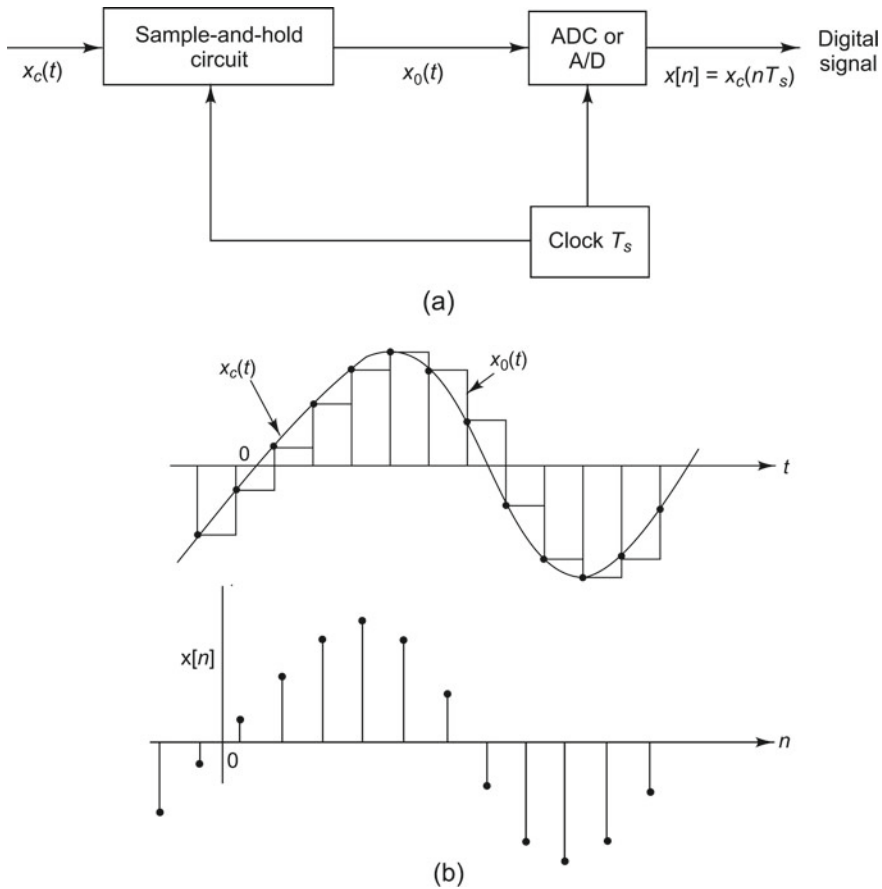
Fig. 8.3 Analogue-to-digital conversion: **a** schematic, **b** typical staircase waveform and digital signal

of bits to be used, which has a direct bearing on the quantization error, the need for sample-and-hold circuits, the required sampling frequency, linearity, etc., and these are discussed in the subsequent sections. The number of bits is decided by the least significant bit (LSB), which is nothing but the percentage of full scale represented by 1 bit; e.g., in a 10-bit ADC or digital-to-analogue converter (DAC) with a full-scale voltage of 10 V, 1 LSB corresponds to $\frac{1}{2^{10}} \times 10\,\text{V} = \frac{1}{1024} \times 10\,\text{V} = 0.009790\,\text{V}$ or 9.76 mV. The sampling frequency is decided by the highest frequency in the signal. Too low a sampling frequency results in what is called aliasing, while too high a sampling frequency involves complicated circuitry. Sampling is generally not invertible, i.e., given the output $x[n]$, it is not always possible to reconstruct $x_c(t)$, the input to the sampler, since many continuous-time signals can produce the same output sequence of samples.

8.2.3.1 Common Terms in ADC or DAC Parlance

Some common terms pertaining to the performance of an ADC or DAC are shown further.

Resolution: This is defined as the fraction of the full-scale range represented by the smallest signal increment that can be detected by a data acquisition system, e.g., 0.0244% of full scale for a 12-bit ADC. It can also be expressed in bits, e.g., as a system having 12-bit resolution or in proportions, i.e., one part in 4,096.

Sampling rate: This represents the speed at which a data acquisition system collects data and is expressed in samples per second. For multi-channel data acquisition systems, the sample rate is typically the speed of the ADC for all channels together. To obtain individual channel sample rate, the speed of the ADC has to be divided by the number of channels being sampled.

Accuracy: This is the difference between the expected and measured output voltages in terms of the change caused by changing the LSB. Most converters have an accuracy of at least ± 1 LSB.

Linearity: This represents how closely the transfer function from input voltage to binary value or its inverse represents a linear function and is usually specified to ± 1 LSB.

Monotonicity: A DAC is said to be non-monotonic if there is a momentary reversal in the expected direction of change.

Zero offset: The zero offset represents the output voltage of a DAC for an input code of zero. It is generally less than 1/2 LSB for unipolar DACs, but bipolar DACs can have considerable zero offset.

Stability: This indicates how constant the full-scale output is with respect to ageing and variations in temperature and power supply.

Settling time: This is an indication of the time taken for the output signal to settle to within 1/2 LSB of its final value after a given change in input scale (usually full scale).

Glitches: These are transients appearing at the output when new digital data are applied to a DAC.

Single-ended (SE) input: Data acquisition devices have either SE inputs or differential inputs or both, and refer to how a signal is wired to a data acquisition device. In SE wiring, each analogue input has a unique high connection, but all channels share a common ground connection.

Differential input: In a differential input configuration, each signal channel has a unique high and unique low connection.

Communication options: One might also want to look into the several different communication options available in ADCs or DACs. Serial communication systems

are preferred when measurements have to be made far away from the computer and long cables are required. RS232 is the most common serial communication option, but has the drawback that it supports transmission distances up to 15 m and SE inputs only; besides it has a very low data rate, up to 20 kbits/s only. For communication over longer distances or at higher bit rates, SE methods are unsuitable. Differential data signals nullify the effect of induced noise signals and ground shifts and thus offer better performance for most applications. RS422 comes under this category of communication devices and is meant for higher bit rates and higher transmission distances than RS232. However, it is RS485 which is supreme for truly multipoint communication networks, catering to a maximum of 32 drivers and receivers on a single (two-wire) bus. RS485 is superior to RS232 and supports transmission distances up to 1.5 km. The universal serial bus (USB) is ideal for data acquisition applications, offering several advantages over conventional series and parallel connections, including higher bandwidth (up to 12 Mbits/s) and the ability to provide power to the peripheral device, making only one cable sufficient to link the data acquisition device to the PC. As computer data acquisition boards plug directly into the computer bus, they are fast and inexpensive. The standard parallel port on a computer can also be used to connect to a data acquisition device. Often parallel port systems support very high sample rates, but the distance between the computer and the data acquisition device is limited to a few metres.

8.2.3.2 Tracking ADC

The tracking ADC (Fig. 8.4) is a simple form of A/D converter. Let $x_c(t)$ be the analogue signal to be converted to digital form. This analogue signal and a reference
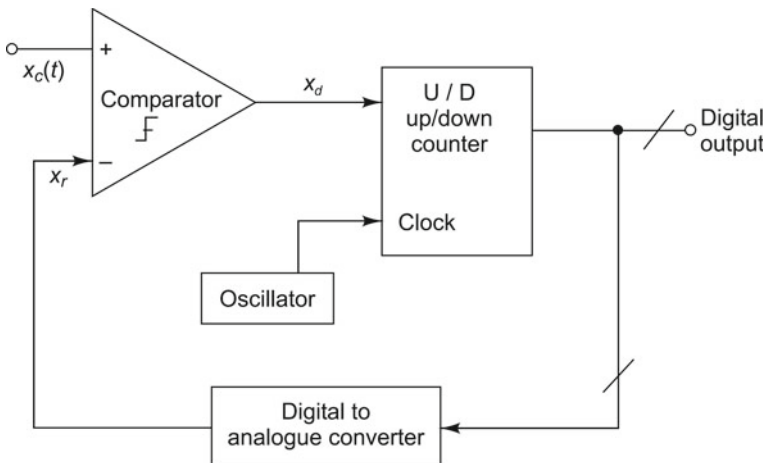
Fig. 8.4 Tracking ADC

signal x_r, form the inputs to a comparator. This reference signal is obtained as the output of a DAC by triggering a binary counter. The output of the counter is converted to analogue form by the D/A converter to produce the reference voltage given by

$$x_r = \left[\left(\frac{a_1}{2} \right) + \left(\frac{a_2}{2^2} \right) + \cdots + \left(\frac{a_n}{2^n} \right) \right] x_{FS} \tag{8.1}$$

where x_{FS} is the full-scale output of the DAC, a_1 is the most significant bit (MSB) corresponding to $(x_{FS}/2)$, and a_n is the LSB corresponding to $(x_{FS}/2^n)$.

The comparator operates such that its output $x_d = 1$ for $x_c > x_r$, and $x_d = 0$ for $x_c < x_r$. The counter counts up as long as it receives 1, i.e., as long as $x_c > x_r$ and the output of the DAC then keeps increasing with increase in digital count of the counter. When the DAC output voltage exceeds the analogue input, i.e., when $x_r > x_c$, the output of the comparator becomes 0, and the counter stops counting, causing the DAC to reveal the digital code corresponding to the analogue input.

As an example, consider a 10-bit ADC with a full scale voltage of 10 V. Let us say we would like to find the digital code corresponding to an analogue input signal of 0.1 V or 100 mV. The voltage corresponding to the LSB for this converter would be $(1/2^{10}) \times 10\,V = 9.76\,mV$. When the counter takes on a count of 1011, the corresponding reference input to the comparator becomes $9.76 \times 11 = 107.426\,mV$, which is greater than the input voltage x_c of 100 mV. The output of the comparator becomes 0, causing the counter to stop counting. The output of the converter for 0.1 V input would thus be 1011. The quantization error is $9.76 \times 11 - 100\,mV = 7.426\,mV$. It may be expected that for a 12-bit ADC, we will get a smaller value of quantization noise. Table 8.1 gives an indication of the resolution obtained with different number of bits in an ADC.

Statistical evaluation of quantization error: The quantization error x may be assumed to have a uniform distribution (for more details on probability distribution, see Sect. 8.9.4.1) with the actual value of the quantized signal lying within ±0.5 LSB of the analogue input. Let its probability density function be denoted as

$$p(x) = 1/LSB, \quad -0.5LSB < x < 0.5LSB$$
$$= 0, \quad \text{otherwise} \tag{8.2}$$

Table 8.1 Number of bits in ADC and resolution

No. of bits (n)	2^n	1 bit with 10 V FS	Resolution (%)
8	256	39 mV	0.390625
10	1024	9.76 mV	0.0976
12	4096	2.4 mV	0.024414
16	65,536	153 μV	0.001526
32	4294967296	0.0002328 μV	0.0000000233

Variance of the error

$$\sigma_x^2 = \int\limits_{-\infty}^{\infty} (x - \mu_x)^2 p(x)\, dx \tag{8.3a}$$

$$= \frac{1}{\text{LSB}} \int\limits_{-0.5\,\text{LSB}}^{0.5\,\text{LSB}} x^2 dx = \frac{\text{LSB}^2}{12} \tag{8.3b}$$

Standard deviation of error

$$\sigma_x = \frac{\text{LSB}}{\sqrt{12}} = 0.29\,\text{LSB} \tag{8.3c}$$

Let us assume that the random vibration of a structure is to be measured with an accelerometer and that the vibration record is to be converted to a digital format for analysis over the frequency range 0–2 kHz with a signal-to-noise ratio (SNR) of at least 80 dB. Let us find the sampling rate and the number of bits/data point required in the A/D conversion. A sampling frequency greater than twice the highest frequency of interest would be required (this is discussed in Sect. 8.6). Hence, a sampling frequency $= 2\frac{1}{2} \times 2 = 5$ kHz may be taken, giving rise to a sampling rate of $(1/5000) = 0.2\,\text{ms}$. It is to be noted here that SNR is specified in terms of dB which is defined as $20 \log_{10}(A/A_{\text{ref}}) = 10 \log_{10}(P/P_{\text{ref}})$, where A and A_{ref} are amplitude-like quantities and P and P_{ref} are power-like quantities.

For a minimum SNR of 80 dB,

$$\frac{S}{N} = \frac{2^2 \times \text{LSB}}{0.29 \times \text{LSB}} = \frac{2^n}{0.29}$$

Hence, $20 \log_{10}\left(\frac{2^n}{0.29}\right) \geq 80\,\text{dB}$ implies that $\frac{2^n}{0.29} \geq 10,000$ or $2^n \geq 2900$. This implies that the number of bits n should be at least 12.

8.2.3.3 Successive Approximation ADC

The successive approximation ADC is much quicker than the tracking ADC because it uses digital logic to converge on the value closest to the input voltage, starting with the most significant bit (MSB) and proceeding to the LSB. As in the tracking ADC, a comparator and a DAC are used for the process. Figure 8.5a shows a flow chart of the conversion process. Figure 8.5b illustrates schematically a 4-bit successive approximation ADC.
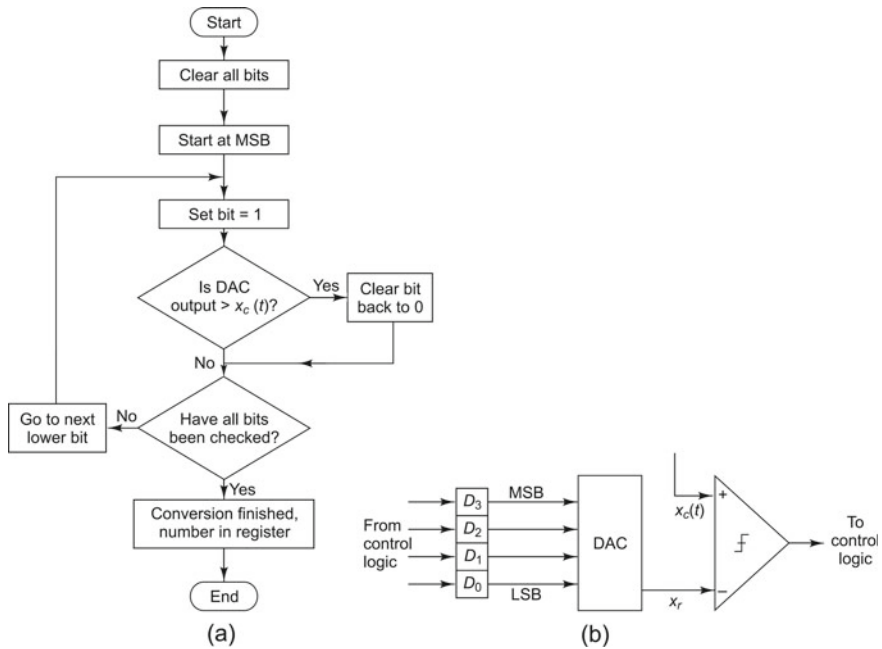
Fig. 8.5 Illustration of successive approximation ADC: **a** flow chart, **b** 4-bit ADC

8.2.4 Digital-to-Analogue Conversion

Some data acquisition systems contain a digital-to-analogue converter or D/A converter which does the opposite job of an ADC. This device converts binary numbers into analogue DC voltages or currents on receipt of command signals from the control hardware. D/A converters are typically used where precise control of external devices is required through an analogue voltage. Thus, the voltage or current from a D/A converter can be used as excitation input to a shaker. In its simplest form, the DAC consists of a network of resistors and switches which provide currents proportional to the weight of each bit, multiplied by its binary value in the parallel digital word. The digital data source is sampled and held in an input register in parallel digital form to serve as a steady input to the DAC. The 0s and 1s are converted to clean digital form using gates which are either on or off depending on the incoming signal. This binary signal can then be converted to an analogue signal using a summing amplifier. The main drawback of this method is that very high precision of the summing resistors is required, and this is especially unsatisfactory for a DAC with a large number of bits. A network with the so-called R-2R ladder of resistances is used to overcome this problem. The DAC with a summing current-to-voltage converter uses an operational amplifier to sum the currents from all the bits and convert them to voltages as shown in Fig. 8.6 for a 4-bit DAC.

Fig. 8.6 R-2R Summing amplifier
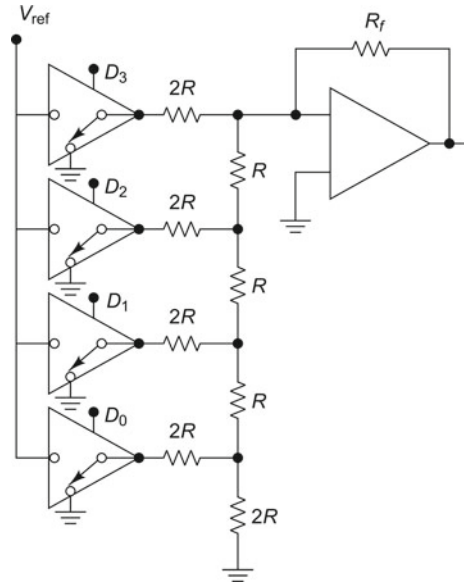
This summing amplifier produces the output

$$V_{\text{out}} = V_{\text{ref}} \left[\frac{D_0}{2^4} + \frac{D_1}{2^3} + \frac{D_2}{2^2} + \frac{D_3}{2^1} \right] \frac{R_f}{R} \tag{8.4}$$

In the above equation, D_0 represents the LSB and D_3 the MSB as before and the D's take on the values 0 or 1. The digital inputs close the switches for a logical 1 and keep it grounded for a logical 0. This can be extended to a DAC with any number of bits using an appropriate R-2R ladder. Offset and gain controls are also generally incorporated in the circuit.

8.3 Some Important Signal Operations

DID YOU KNOW that one of the most common signal operations is conversion from the time domain to the frequency domain and vice versa? It was the French mathematician, Egyptologist, and administrator, Jean-Baptiste-Joseph, Baron Fourier (born 1768) who showed in 1807 how the conduction of heat in solid bodies may be analysed in terms of an infinite mathematical series, the Fourier series, named after him. Fourier considered heat flow in a ring, i.e., a bar bent into a circle and analysed the temperature distribution which is

forced to be spatially periodic. But doubts expressed by Laplace and Lagrange delayed the publication. There were also criticisms made by Biot and Poisson. Nevertheless, a prize was granted to Fourier for his work on the propagation of heat in solid bodies, but with a citation that mentioned lack of generality and rigor! Fourier's idea is, to date, one of the methods of approaching the Fourier integral as a limit, starting from a Fourier series representation. His work stimulated research in mathematical physics and engineering. In spite of his scientific leaning, the French revolution threw him into politics. In 1798, he accompanied Napoleon on an expedition to Egypt and gave advice on engineering and diplomatic undertakings. He served for 3 years as the secretary of the Institut d'Égypte. After his return to France, he published the Description de l'Égypte, a treatise on the ancient civilization of Egypt. He was also appointed administrator for the national government, a position he held from 1802 to 1814.

The signal operations described here are amplitude scaling, translation or time shifting, time scaling, and time inversion/folding. We may need these operations in the course of signal analysis. Though the independent variable considered here and in subsequent sections is time, these operations hold good for signals which are functions of other independent variables such as space or frequency.

8.3.1 Amplitude Scaling

This is typically the kind of operation that takes place in a voltage amplifier. A signal $\phi(t)$ is said to be an amplitude scaled version of a signal $f(t)$ if

$$\phi(t) = kf(t) \tag{8.5}$$

Figure 8.7a shows a signal $f(t)$ and Fig. 8.7b and c show two different cases of amplitude scaling.
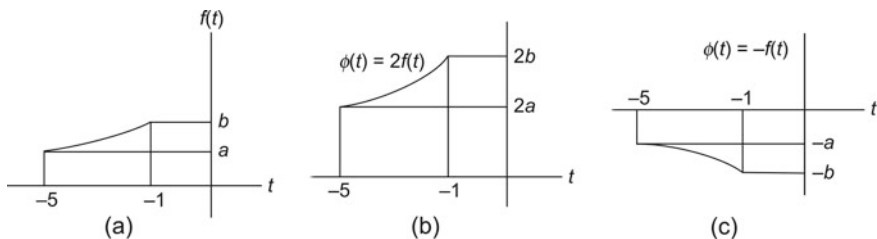
Fig. 8.7 Amplitude scaling: **a** $f(t)$, **b** $\phi(t) = 2f(t)$, **c** $\phi(t) = -f(t)$
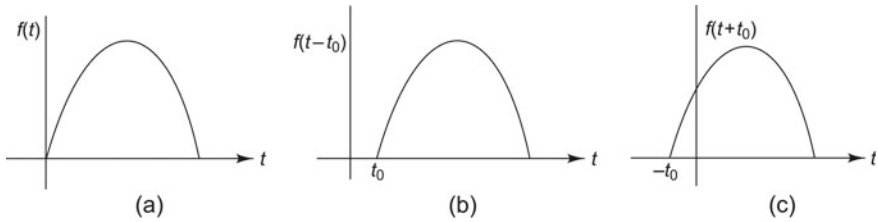
Fig. 8.8   Time shifting: **a** $f(t)$, **b** $\phi(t) = f(t - t_0)$, **c** $\phi(t) = f(t + t_0)$

8.3.2   Translation/Time Shifting

This happens with delaying or advancing operations (Fig. 8.8). Such operations are required while computing auto- or cross-correlations of signals. A signal $\phi(t)$ is said to be a translated version of $f(t)$ if it takes on one of the following forms. We assume that t_0 is positive in these equations.

 (i) $\phi(t)$ is said to be a delayed version of $f(t)$ if

$$\phi(t) = f(t - t_0) \tag{8.6a}$$

The signal $f(t)$ gets shifted to the right by t_0 in this case as seen from Fig. 8.8b.
(ii) $\phi(t)$ is said to be an advanced version of $f(t)$ if

$$\phi(t) = f(t + t_0) \tag{8.6b}$$

In this case, the signal $f(t)$ gets shifted to the left-hand side by t_0 as seen from Fig. 8.8c.

8.3.3   Time Scaling

A signal is said to be time scaled if it undergoes compression or expansion on the time axis. This is the operation that takes place when we record a signal at one speed and replay at another. Frequency scaling is done in filters like 1/1 and 1/3 octave filters and in wavelet transforms.

8.3.3.1   Expansion

When we record a signal at some speed and replay at a much lower speed, as when we view a recorded picture in slow motion, we are, in effect, expanding the signal
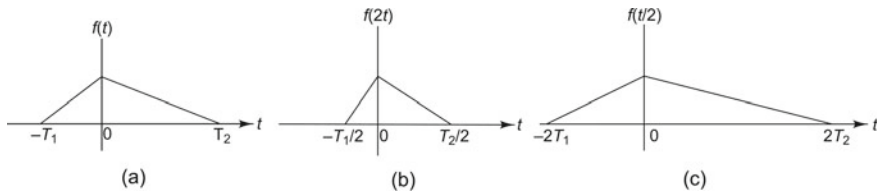
Fig. 8.9 Time scaling: **a** $f(t)$, **b** $f(2t)$, **c** $f\left(\frac{t}{2}\right)$

in the time scale. $\phi(t)$ is said to be an expanded version of $f(t)$ if the following equation is satisfied.

$$\phi(t) = f\left(\frac{t}{a}\right) \tag{8.7a}$$

Here, a is a positive constant.

8.3.3.2 Compression

$\phi(t)$ is said to be a compressed version of $f(t)$ if

$$\phi(t) = f(at) \tag{8.7b}$$

where $a > 1$ as before.

Figure 8.9a–c show $f(t)$, $f(2t)$, and $f\left(\frac{t}{2}\right)$, respectively.

8.3.4 Time Inversion/Reversal/Folding/Flipping

This happens when we replay a tape in the direction opposite to that in which it was recorded. A signal $\phi(t)$ is said to be a time-reversed version of $f(t)$ if

$$\phi(t) = f(-t) \tag{8.8}$$

Figure 8.10a and b show the original and time-reversed versions of $f(t)$. The latter is seen to be a mirror image of the former about the axis $t = 0$ or is said to be a folded or flipped version.

8.3.5 Even and Odd Parts of a Signal

Examples of even and odd signals are the cosine and sine waves, respectively, plotted as a function of time. The former is so-called because it is symmetric about the
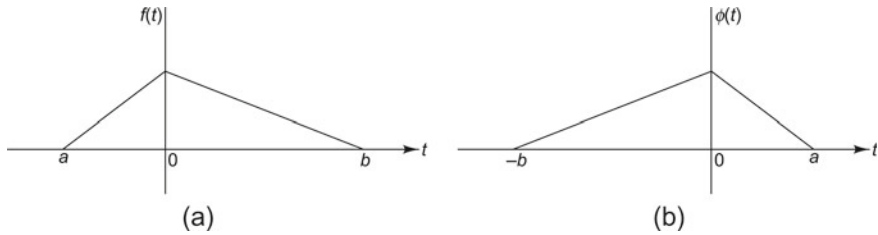
Fig. 8.10 Time inversion: **a** $f(t)$, **b** $\phi(t) = f(-t)$
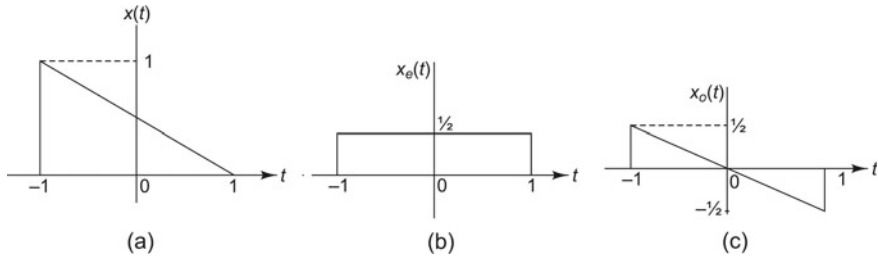
Fig. 8.11 Even and odd parts of a signal: **a** $x(t)$, **b** $x_e(t)$, **c** $x_o(t)$

y-axis, while the latter is anti-symmetric. The even part $x_e(t)$ of a real signal $x(t)$ is such that

$$x_e(t) = x_e(-t) \tag{8.9a}$$

$$x_e(t) = \frac{x(t) + x(-t)}{2} \tag{8.9b}$$

The odd part $x_o(t)$ of a signal $x(t)$ is such that

$$x_o(t) = -x_o(-t) \tag{8.10a}$$

$$x_o(t) = \frac{x(t) - x(-t)}{2} \tag{8.10b}$$

Figure 8.11a–c show $x(t)$, $x_e(t)$ and $x_o(t)$, respectively.

For a complex signal $x(t)$, we have the conjugate symmetric and conjugate anti-symmetric parts which are the even and odd counterparts of a real signal. The conjugate symmetric part of a complex signal is defined as

$$x_e(t) = \frac{x(t) + x^*(-t)}{2} \tag{8.11a}$$

and

$$x_e(t) = x_e^*(-t) \tag{8.11b}$$

The conjugate anti-symmetric part is defined as

$$x_o(t) = \frac{x(t) - x^*(-t)}{2} \tag{8.12a}$$

and

$$x_o(t) = -x_o^*(-t) \tag{8.12b}$$

8.4   Some Important Concepts Related to Signals and Systems

DID YOU KNOW that Duhamel's principle, a method of obtaining solutions to inhomogeneous linear evolution equations, is named after Jean-Marie-Constant Duhamel, a French mathematician and physicist who worked on the transmission of heat in crystal structures? He was primarily a mathematician but also studied the mathematics of heat, mechanics, and acoustics. In 1830, he started teaching analysis at the École Polytechnique, where he was regarded highly as a teacher and was said to have delivered very fine lectures. He remained there until his retirement in 1869. He worked in calculus using infinitesimals; his theorem states that the sum of a series of infinitesimals is unchanged by replacing the infinitesimal with its principal part. While at the École, he worked on the acoustics of vibrating strings and the vibration of air in cylindrical and conical pipes, and also the physics of harmonic overtones. His contributions to the distribution of heat in a solid with a variable boundary temperature led to 'Duhamel's principle' in partial differential equations. He invented a recording instrument consisting of a pen attached to a vibrating string; the pen left a record on a moving plate, but could not play back the etchings it recorded.

8.4.1   Unit Impulse Function

The concept of a unit impulse function is very important in the study of signals and systems. Such a signal can be described as shown in Fig. 8.12a and in the equation below.
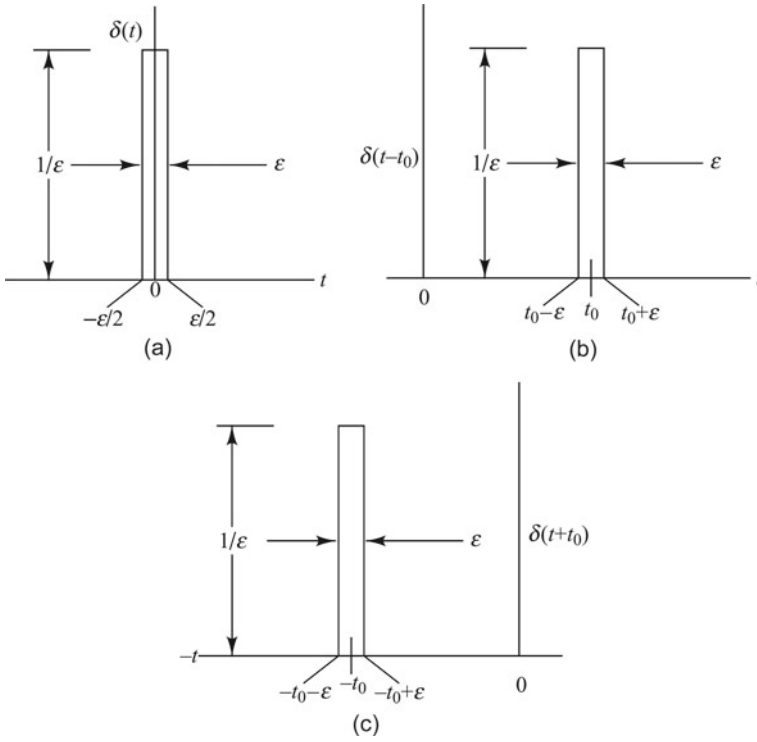
Fig. 8.12 Delta function: **a** $\delta(t)$, **b** delayed version $\delta(t - t_0)$, **c** advanced version $\delta(t + t_0)$

$$\delta(t) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}, \quad -\frac{\varepsilon}{2} \le t \le \frac{\varepsilon}{2} \tag{8.13a}$$
$$= 0, \ |t| > \frac{\varepsilon}{2}$$

Here, ε is a very small, arbitrary quantity. The area under the function is computed by integrating $\delta(t)$ as shown in Eq. (8.13b) and is found to be unity. It is for this reason that this function is called a unit impulse.

$$Area = \int_{-\infty}^{\infty} \delta(t) dt = \lim_{\varepsilon \to 0} \int_{-\varepsilon/2}^{\varepsilon/2} \frac{dt}{\varepsilon} = 1 \tag{8.13b}$$

Delta functions can be translated to any point t_0 in time and can be scaled in magnitude by a constant A to obtain $A\delta(t - t_0)$.

$$A\delta(t) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}, \quad t_0 - \frac{\varepsilon}{2} \le t \le t_0 + \frac{\varepsilon}{2} \tag{8.14a}$$
$$= 0, \text{ elsewhere}$$

It follows that

$$\int\limits_{t_0-\varepsilon/2}^{t_0+\varepsilon/2} A\delta(t - t_0)\mathrm{d}t = A \qquad (8.14\text{b})$$

Figure 8.12a–c show the delta function, its delayed version, and advanced version, respectively.

Besides, when any arbitrary function $x(t)$ is multiplied by the translated delta function $\delta(t - t_0)$ and integrated over time, the value of $x(t)$ at $t = t_0$ is obtained, this being called the sifting property of the delta function. That is

$$\int\limits_{\infty}^{-\infty} x(t)\delta(t - t_0)\mathrm{d}t = x(t_0) \qquad (8.15\text{a})$$

It is also to be noted that convolving any signal $x(t)$ with the time-delayed Dirac delta function $\delta(t - T)$ produces a time delay in $x(t)$ of the same amount as shown below:

$$x(t) * \delta(t - T) = \int\limits_{-\infty}^{\infty} x(\tau)\delta(t - T - \tau)d\tau = x(t - T) \qquad (8.15\text{b})$$

A point of great importance is that any arbitrary signal $x(t)$ can be represented as

$$x(t) = \int\limits_{-\infty}^{\infty} x(\tau)\delta(t - \tau)d\tau \qquad (8.16)$$

Equation (8.16) implies that $x(t)$ may be thought of as the superposition of an infinite number of magnitude scaled and translated (in time) impulse functions. Thus, it is clear that delta functions can be used as basic building blocks for synthesizing arbitrary functions in the time domain, in much the same way that unit sinusoidal functions can be used for synthesizing periodic functions (please refer to descriptions of Fourier series).

8.4.1.1 Unit Sample Sequence

The discrete time counterpart of the unit impulse signal is the unit sample sequence $\delta[n]$.

$$\delta[n] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases} \qquad (8.17)$$
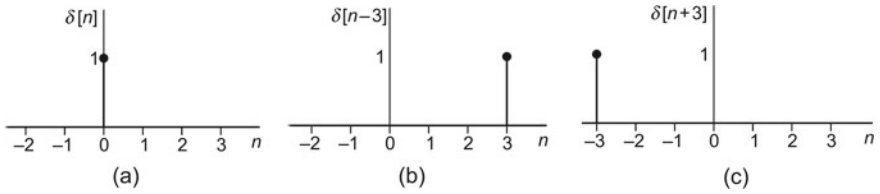
Fig. 8.13 Unit sample sequence: **a** $\delta[n]$, **b** delayed version $\delta[n-3]$, **c** advanced version $\delta[n+3]$

Fig. 8.14 Representation of
a sequence in terms of unit
impulses
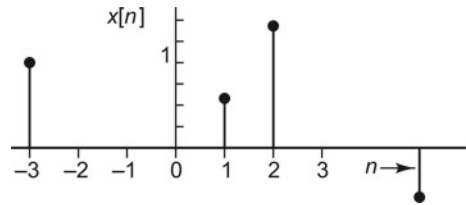
As in the case of the translated delta function for the continuous-time case, we can
have the shifted unit sample functions as shown in Fig. 8.13.

$$\delta[n - n_0] = \begin{cases} 1, & n = n_0 \\ 0, & n \neq n_0 \end{cases} \tag{8.18a}$$

$$\delta[n + n_0] = \begin{cases} 1, & n = -n_0 \\ 0, & n \neq -n_0 \end{cases} \tag{8.18b}$$

Representation of any arbitrary sequence $x[n]$ may be made in terms of $\delta[n]$ as
shown in Fig. 8.14. The signal shown may be expressed mathematically as

$$x[n] = a_1\delta[n - 1] + a_2\delta[n - 2] + a_5\delta[n - 5] + a_{-3}\delta[n + 3] \tag{8.19}$$

In general, any sequence may be represented as

$$x[n] = \sum_{k=-\infty}^{\infty} a_k\delta[n - k] \tag{8.20a}$$

or more generally as

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n - k] \tag{8.20b}$$

This is nothing but the convolution sum representing the superposition of magnitude-scaled and shifted unit impulses. $x[k]$ is the scale factor for the kth sample in the sequence.

8.4.2 Impulse Response Function (IRF)

This is a very important function characterizing the time domain behaviour of a system which (i) is physically realizable, (ii) is stable, (iii) has constant or time-invariant parameters, and (iv) is linear. The impulse response function $h(t)$ is the output of the system (with zero initial conditions) to a unit impulse function. This implies that $h(t) = y(t)$ with $x(t) = \delta(t)$, the unit impulse function. The IRF is purely a function of the system properties and because it gives a complete description of the system's characteristics, it can be used to predict the response to any input. For a single-input single-output (SISO) system, the input $x(t)$ and the output $y(t)$ are related through the convolution integral

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau)\mathrm{d}\tau \tag{8.21}$$

It is seen that Eq. (8.21) is a representation of the output in the form of the superposition of an infinite number of magnitude-scaled and translated IRFs. These equations make the study of the input–output relationship easy to handle. If a physical system can be modelled as a stable, linear time-invariant system within allowable approximations, the opportunity should not be lost.

8.4.3 Linear Time-Invariant System

For such a system, the properties of both linearity and time-invariance hold good. Convolution integral can be used to represent the system as we have just seen. An operator A is said to be linear if for any set of admissible values $x_1, x_2, x_3, \ldots, x_N$ and constants $a_1, a_2, a_3, \ldots, a_N$

$$A\left[\sum_{i=1}^{N} a_i x_i\right] = \sum_{i=1}^{N} a_i A[x_i] \tag{8.22}$$

An operator A is said to be shift-invariant if for any shift t_0 of the input $x(t)$ to $x(t + t_0)$, there results a similar shift of output $y(t)$ to $y(t + t_0)$ for any t_0.

$$\text{i.e., } y(t + t_0) = A[x(t + t_0)] \tag{8.23}$$

$x(t)$ → [Impulse response function $h(t)$] → $y(t) = x(t) * h(t)$
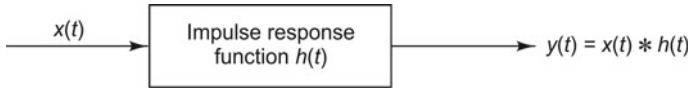
Fig. 8.15 Mathematical representation of an LTI system

Thus, for a linear, time-invariant system, the response $y(t)$ to an arbitrary input $x(t)$ is equal to the convolution of the input with the impulse response of the system $h(t)$ as shown in Fig. 8.15 and in Eq. 8.24

$$y(t) = \begin{cases} x(t) * h(t) \\ \int\limits_{-\infty}^{\infty} x(\tau)h(t - \tau)\mathrm{d}\tau \end{cases} \tag{8.24a}$$

where * represents the convolution operator.

Substituting $t - \tau = t'$ in Eq. (8.24a), we get

$$y(t) = \int\limits_{-\infty}^{\infty} x(t - t')h(t')\mathrm{d}t' \tag{8.24b}$$

Instead of variable t', this equation can be written in terms of variable τ to give

$$y(t) = \begin{cases} \int\limits_{-\infty}^{\infty} h(\tau)x(t - \tau)\mathrm{d}\tau \\ h(t) * x(t) \end{cases} \tag{8.24c}$$

From Eqs. (8.24a)–(8.24c) it is seen that

$$y(t) = h(t) * x(t) = x(t) * h(t) \tag{8.25}$$

This implies that we can change the order of the IRF and system input in the computation of the system output in a series combination of the system. This is a very important relationship, serving as the foundation for engineering analysis of linear systems. An equivalent description of the IRF in the frequency domain is the Fourier transform of the impulse response, and this is called the frequency response function (FRF) of the system. According to the convolution theorem which will be seen later, the convolution of the input with the IRF is equivalent to multiplication of the input spectrum with the FRF of the system. Since multiplication is an operation which is easier to perform than convolution, frequency domain analysis is often preferred to time domain analysis with the final result being transformed back into the time domain if required.

8.4.3.1 Discrete-Time Convolution or Convolution Sum

For the discrete-time case, we have the concept of a linear shift-invariant (LSI) system, akin to the linear time-invariant (LTI) system and the convolution integral is replaced by the convolution sum. Although the convolution sum is analogous to the convolution integral, it is to be remembered that it is not an approximation to the integral. While the integral plays mainly a theoretical role for continuous systems, the sum often serves as an explicit realization of a discrete-time linear system.

For any input $x[n]$, the output $y[n]$ is

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] = \sum_{k=-\infty}^{\infty} h[k]x[n-k] \qquad (8.26)$$

where $h[n]$ is the IRF.

8.4.4 Causal Systems and Signals

This relates to the cause and the effect produced in a system. A causal system is one, the output of which is a function of the present and past inputs only, not future inputs. Causal signals may be defined in a similar fashion. Causal signals are those which are zero for $t < 0$. Since the IRF $h(t)$ is the output of a system to the unit impulse function $\delta(t)$ which is zero for $t < 0$, $h(t) = 0$ for $t < 0$ for a causal system, or in other words the IRF is a causal signal. Non-causal signals are those that exist for $t < 0$ as well as for $t \geq 0$. Anti-causal signals are those which are zero for all $t \geq 0$. Examples of such signals are shown in Fig. 8.16.

8.4.5 Stability

This is a property of a system, not a signal. There are several definitions for stability available in literature. We shall discuss one that is appropriate for causal, linear, time-invariant systems. A system is stable in the *bounded-input, bounded-output* (BIBO) sense, if and only if output $y(t)$ is bounded for all possible bounded inputs $x(t)$. If even one bounded input gives rise to an unbounded output, then the system is said to be BIBO unstable.

An input $x(t)$ is said to be bounded if

$$|x(t)| < B_x < \infty \qquad (8.27)$$
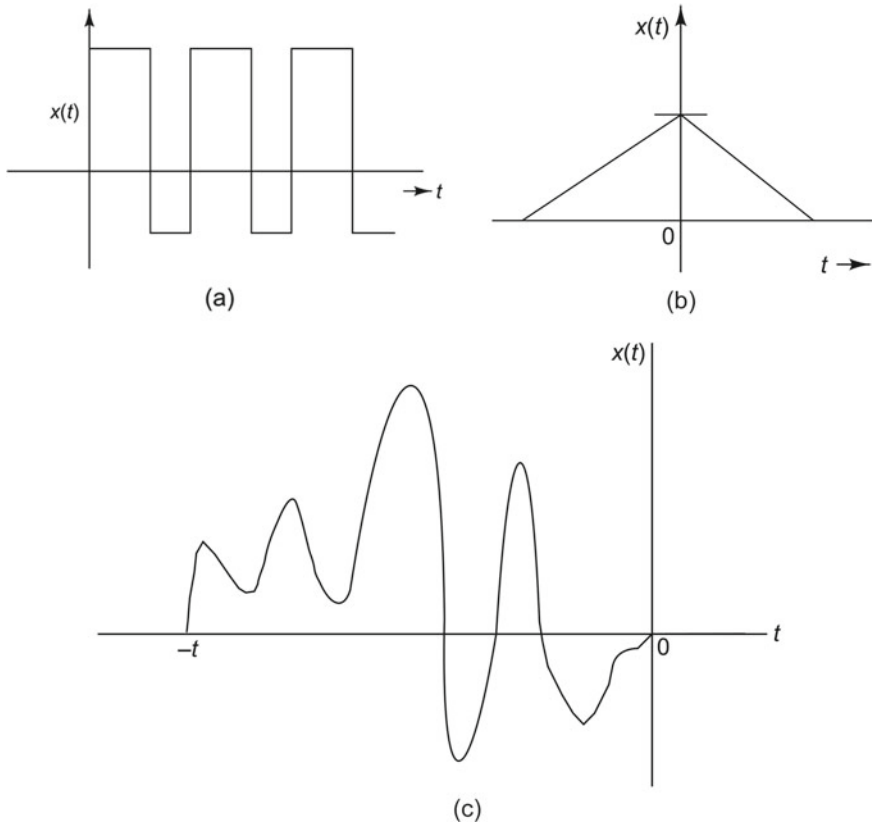
Fig. 8.16 Causality of signals: **a** causal signal, **b** non-causal signal, **c** anti-causal signal

For such a bounded input to an LTI system, we can expect that

$$
|y(t)| \leq
\begin{cases}
\displaystyle\int_{-\infty}^{\infty} |x(\tau)||h(t-\tau)|d\tau \\[2ex]
\displaystyle B_x \int_{-\infty}^{\infty} |h(t-\tau)|d\tau
\end{cases}
\tag{8.28}
$$

Since any output is obtained as the convolution of the input with the IRF for an LTI system, this implies that if the IRF is absolutely integrable, then the output to every bounded input is bounded. From this, it follows that for an asymptotically stable system (for which $h(t)$ decays to zero as time goes to infinity),

$$
\int_{-\infty}^{\infty} |h(t)|dt < B_h < \infty
\tag{8.29}
$$

and

$$|y(t)| \leq B_x B_h < \infty \qquad (8.30)$$

It is to be noted that the IRF tells us all we want to know about a system. For instance, if we want to know about the stability and causality of a system, we just have to inspect $h(t)$ of the system.

For stability of a discrete-time system, the following necessary and sufficient condition has to be satisfied.

$$S = \sum_{k=-\infty}^{\infty} |h(k)| < \infty \qquad (8.31)$$

8.4.6   *Frequency Response Function*

The dynamic properties of a system may be described in the frequency domain by what is called the FRF. For a physical understanding of the FRF, it is worth mentioning that a stable LTI system, in response to a harmonic excitation at a particular frequency, produces a harmonic output at the same frequency. The FRF at a particular frequency is the response of such a system to unit harmonic input at that frequency and is generally plotted at all possible frequencies in the frequency range of interest. The FRF and hence magnitude and phase lag of the output are a function of the system parameters alone, being dependent neither on the input, nor on the output. The system gain is defined as

$$|H(\omega)| = \left| \frac{Y(\omega)}{X(\omega)} \right| \qquad (8.32)$$

and the associated phase lag as

$$\phi(\omega) = \tan^{-1} \frac{\text{Im}\{H(\omega)\}}{\text{Re}\{H(\omega)\}} \qquad (8.33)$$

Here, $Y(\omega)$ and $X(\omega)$ constitute the Fourier transforms of the output and input, respectively. The FRF is also the Fourier transform of the IRF. Thus,

$$H(\omega) = \int_{-\infty}^{\infty} h(t) e^{-i\omega t} \, dt \qquad (8.34a)$$

It follows that the IRF is the inverse Fourier transform of the FRF.

$$h(t) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} H(\omega) e^{i\omega t} d\omega \qquad (8.34b)$$

It may be noted that the lower limit of integration can be taken as 0 in the transform Eqs. (8.34a) and (8.34b) since $h(t) = 0$ for $t < 0$ for a physically realizable or causal system and since negative frequencies ω do not have any practical meaning. The FRF is a complex, frequency-dependent quantity. It is generally described graphically in three different forms:

(i) $|H(\omega)|$ versus frequency and $\angle H(\omega)$ versus frequency, constituting a set of two plots. Conventionally, phase is defined such that phase lags are considered positive. If log–log scales are used for the axes in the magnitude plots, these are called Bodé plots.

(ii) $\text{Re}\{H(\omega)\}$ versus frequency and $\text{Im}\{(\omega)\}$ versus frequency, constituting a set of two plots.

(iii) Polar plot where $\text{Re}\{H(\omega)\}$ is plotted versus $\text{Im}\{H(\omega)\}$ as ω varies from zero to infinity. It may be observed that the frequency information is not explicit in this notation and has to be marked manually. The various plots are discussed in detail in Chap. 9.

8.4.6.1 Obtaining FRF of a System

The FRF of a system may be obtained in the following two ways:

(i) Excite the system with $\delta(t)$ to obtain $h(t)$ as output; then $H(\omega)$ is obtained by the Fourier transformation of $h(t)$ as described in Eq. (8.34a).

(ii) Alternatively, $H(\omega)$ may be obtained by imparting swept sine or pure random excitation to obtain $H(\omega)$ straightaway from a frequency domain operation.

In the first method, the system is excited with a unit impulse to obtain $h(t)$. The Fourier transform of the unit impulse is given by

$$\Im\{\delta(t)\} = \int\limits_{-\infty}^{\infty} \delta(t) e^{-i\omega t} dt = 1 \qquad (8.35)$$

This implies that the unit impulse has a spectrum with ω or f extending from $-\infty$ to $+\infty$ with spectral height unity. Thus, in the context of vibration testing, when we excite a system with a unit impulse in the time domain, we are in effect imparting harmonic excitation at all frequencies simultaneously to the system, causing it to respond simultaneously at all its natural frequencies. This is the basis of a free vibration test, which is a time domain method used for bringing out the system parameters in the form of the IRF. The FRF $H(\omega)$ may then be got by doing a Fourier transformation on the IRF.

In the second method, we obtain the FRF by imparting harmonic excitation. With the function $e^{i\omega t}$ as input, the output of an LTI system is obtained from the convolution operation as

$$y(t) = \begin{cases} \int\limits_{-\infty}^{\infty} h(\tau)x(t-\tau)d\tau = \int\limits_{-\infty}^{\infty} h(\tau)e^{i\omega(t-\tau)}d\tau \\ e^{i\omega t} \int\limits_{-\infty}^{\infty} h(\tau)e^{-i\omega\tau}d\tau = e^{i\omega t} H(\omega) \end{cases} \tag{8.36}$$

This equation is a statement of the fact that complex exponentials are eigenfunctions of LTI systems and $H(\omega)$ are the eigenvalues. When we subject a system to swept sine excitation, we impart sequentially harmonic excitations of different frequencies. The output in this case is directly brought out as the FRF since this is a frequency domain method. To avoid the large amount of time taken for testing in the frequency domain using swept sine excitation, we may resort to pure random excitation, where instead of exciting the system at increasing or decreasing frequencies sequentially, we excite the system at all frequencies simultaneously, thus obtaining the FRF in one shot and thereby reducing testing time considerably. The FRF that we get, whether it is from swept sine excitation or random excitation, is nevertheless the same, being defined only by the system. This forms the basis for frequency domain methods.

8.5 Frequency Domain Analysis

INTERESTING FACTS: The word frequency immediately brings to our mind its unit Hz, named after the brilliant German physicist and experimentalist Heinrich Rudolf Hertz. He excelled at languages and enjoyed studying the sciences and mathematics even when young. He enjoyed working with his hands. He learned to operate a lathe, built models and even constructed a sophisticated spectroscope. He used this device to conduct his own experiments in physics and chemistry. He became a student of Hermann von Helmholtz at the age of 21. When he was professor at the Karlsruhe Polytechnic, he produced electromagnetic waves in the laboratory

and established that James Clerk Maxwell's theory of electromagnetism is correct. He measured the length and velocity of the waves and proved beyond doubt that the nature of these waves was the same as those of light and heat waves, establishing that light and heat are electromagnetic radiations. Electromagnetic waves were therefore named after him and called Hertzian waves. He wrote many scientific papers which were translated into English. They were published in three volumes: Electric Waves, Miscellaneous Papers, and Principles of Mechanics. Hertz is also credited with discovering the photoelectric effect.

Most of the early test methods for system identification were time domain methods. Later these methods were replaced by frequency domain methods. Narrow-band frequency analysis was originally done using constant-bandwidth narrow-band analysers. A good description of frequency analysers is given in Sect. 8.8. These methods were exceedingly slow, especially for large signal bandwidths and were limited to a 1.5–2.0 Hz resolution. A statistically averaged analysis of data from 0 to 500 Hz with 2 Hz resolution would typically take 2 h to complete. With the progress in the electronics industry in the late 1960s and in signal processing techniques in the early 1970s, the FFT-based spectrum analysers were born. The FFT behaves like a bank of parallel filters. The FFT-based machines inherently provide phase information which is required for the high-speed FRF computation. The popularity of the two-channel machines led to the demand for multi-channel FFT machines. Frequency analysis using these machines utilizes the Fourier transform pair of equations to convert a signal from the time domain to the frequency domain and vice versa. Fourier's approach assumes that any waveform can be represented as the superposition of a series of sine and cosine waves at different frequencies. A Fourier series is used for representation of continuous-time periodic signals. For continuous-time aperiodic transient signals, the Fourier transform or analysis equation is used. This is defined as

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} \, dt \tag{8.37a}$$

The inverse Fourier transform or synthesis equation is given by

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} \, d\omega \tag{8.37b}$$

All signals are not Fourier transformable. The question that arises is which class of signals can be represented by Fourier and inverse Fourier transforms. The existence of the transform is assured for any $x(t)$ that satisfies the Dirichlet conditions (readers are suggested to refer to any book on signal analysis). The first of these conditions is that

$$\int_{-\infty}^{\infty} |f(t)| \mathrm{d}t < \infty \tag{8.38a}$$

Since

$$|X(\omega)| = \left| \int_{-\infty}^{\infty} x(t) \mathrm{e}^{-\mathrm{i}\omega t} \mathrm{d}t \right| < \int_{-\infty}^{\infty} |x(t)||\mathrm{e}^{-\mathrm{i}\omega t}| \mathrm{d}t \leq \int_{-\infty}^{\infty} |x(t)| \mathrm{d}t < \infty \tag{8.38b}$$

it follows that for the Fourier transform to exist $|X(\omega) < \infty|$ for all ω. Thus, the sufficient condition for convergence is

$$|X(\omega)| \leq \int_{-\infty}^{\infty} |x(t)| \mathrm{d}t < \infty \tag{8.39}$$

Therefore, if $x(t)$ is absolutely integrable, then $X(\omega)$ exists, or in other words any stable function has a Fourier transform. Any finite-length function thus has a Fourier transform because it is absolutely integrable. We know that the output $y(t)$ of an LTI system with input $x(t)$, and finite-duration impulse response $h(t)$ can be obtained from the convolution integral. Since Eq. (8.40) has to be satisfied for BIBO stability, it can be said that any system with a finite-duration impulse response (FIR) will be stable and will have a Fourier transform.

$$S = \int_{-\infty}^{\infty} |h(t)| \mathrm{d}t < \infty \tag{8.40}$$

Some functions $x(t)$ are not absolutely integrable, but are square integrable as shown further.

$$\int_{-\infty}^{\infty} |x(t)|^2 \mathrm{d}t < \infty \tag{8.41}$$

If we relax the condition of uniform convergence and have mean square convergence instead, then those functions $x(t)$ which are not absolutely integrable may also be assumed to have Fourier transforms.

8.5.1 Symmetry Properties of the Fourier Transform

Table 8.2 gives the symmetry properties of the Fourier transform. The conjugate symmetric and conjugate anti-symmetric parts of a complex function, as described

Table 8.2 Symmetry properties of the Fourier transform

Signal $x(t)$	Fourier transform $X(\omega)$
$x(t) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} X(\omega)e^{i\omega t}\,d\omega$	$X(\omega) = \int\limits_{-\infty}^{\infty} x(t)e^{-i\omega t}\,dt$
$x^*(t)$	$X^*(-\omega)$
$x^*(-t)$	$X^*(\omega)$
$\text{Re}\{x(t)\}$	$X_e(\omega)$, conjugate symmetric part of $X(\omega)$
$i\,\text{Im}\{x(t)\}$	$X_o(\omega)$, conjugate anti-symmetric part of $X(\omega)$
$x_e(t)$, conjugate symmetric part of $x(t)$	$\text{Re}\{X(\omega)\}$, i.e., $X_R(\omega)$
$x_o(t)$, conjugate anti-symmetric part of $x(t)$	$i\,\text{Im}\{X(\omega)\}$, i.e., $iX_I(\omega)$
For $x(t)$ real, the following properties apply	
$X(\omega) = X^*(-\omega)$	
$X_R(\omega) = X_R(-\omega)$, real part is even	
$X_I(\omega) = -X_I(-\omega)$, imaginary part is odd	
$\|X(\omega)\| = \|X(-\omega)\|$, magnitude is even	
$\angle X(\omega) = -\angle X(-\omega)$, phase is odd	
$x_e(t)$, even part of $x(t)$	$X_R(\omega)$
$x_o(t)$, odd part of $x(t)$	$iX_I(\omega)$

in Eqs. (8.11) and (8.12) are used here. The corresponding even and odd parts of a real function are as described in Eqs. (8.9) and (8.10).

8.5.2   Fourier Transform Theorems

The Fourier transform is very powerful mainly due to the many theorems describing the properties of the transformation; these theorems provide insight into the nature of physical systems. The forward and inverse Fourier transform equations offer complementary views of the same signal. Hence, if some operation is performed on one equation of the pair, then it follows intuitively and can be proved mathematically that an equivalent operation can be performed on the other equation. In all the equations below, we will use the notation $x(t) \leftrightarrow X(\omega)$ which means that $x(t)$ has a Fourier transform $X(\omega)$.

8.5.2.1   Linearity

This theorem states that if a function is scaled in the time domain, its transform also gets scaled by the same amount. Thus adding two functions in the time domain corresponds to adding the two frequency spectra. If $x_1(t) \leftrightarrow X_1(\omega)$ and $x_2(t) \leftrightarrow X_2(\omega)$, then

$$ax_1(t) + bx_2(t) \leftrightarrow aX_1(\omega) + bX_2(\omega) \tag{8.42}$$

8.5.2.2   Scaling

This theorem states that if a signal is compressed in the time domain, it gets expanded in the frequency domain and vice versa. A change in the scale of the time axis by a positive real constant a changes the scale of the frequency axis in the spectrum by $1/a$. Thus,

$$x(at) \leftrightarrow \frac{1}{|a|} X\left(\frac{\omega}{a}\right) \tag{8.43a}$$

$$x(t/a) \leftrightarrow |a|X(\omega a) \tag{8.43b}$$

An implication of this theorem is that for any real $x(t)$,

$$x(-t) \leftrightarrow X(-\omega) \tag{8.43c}$$

This is got by letting $a = -1$ in Eq. (8.43a). In other words, flipping the time function about the origin corresponds to flipping its spectrum about the origin. If $x(t)$ is complex, then

$$x^*(-t) \leftrightarrow X^*(\omega) \tag{8.43d}$$

8.5.2.3   Time Shifting and Frequency Shifting

Time shift: Shifting a function in time (delaying or advancing) induces a phase shift proportional to the frequency and to the amount of time shift.

$$\text{Thus } x(t - t_0) \leftrightarrow X(\omega)e^{-i\omega t_0} \tag{8.44a}$$

Frequency shift: Conversely, shifting in frequency causes the time function to be multiplied by a unit phasor with angle proportional to time and to the amount of shift.

$$\text{i.e., } x(t)e^{i\omega t_0} \leftrightarrow X(\omega - \omega_0) \tag{8.44b}$$

8.5.2.4   Duality Between Time and Frequency Domains (Transform of a Transform)

In order to convert from the frequency domain back to the time domain, we normally use the inverse Fourier transform. However, if instead, we take a forward Fourier transform of the spectrum, the result is a time function which has been reversed about the y-axis. This implies that

$$X(\omega) \leftrightarrow x(-t) \tag{8.45a}$$

and

$$X(t) \leftrightarrow x(-\omega) \tag{8.45b}$$

For example, $\mathrm{rect}(t) \leftrightarrow \mathrm{sinc}(f)$ implies that $\mathrm{sinc}(t) \leftrightarrow \mathrm{rect}(-f)$, where the rectangular and sinc functions are as shown below.

$$\mathrm{rect}\left(\frac{t}{\tau}\right) = \begin{cases} 1, & |t| < \frac{\tau}{2} \\ 0, & |t| > \frac{\tau}{2} \end{cases} \tag{8.46}$$

$$\mathrm{sinc}(t) = \frac{\sin(\pi t)}{\pi t} \tag{8.47}$$

8.5.2.5   Convolution

This theorem states that if two functions are convolved in the time domain, then their Fourier transforms are multiplied in the frequency domain. Conversely, if they are multiplied in the time domain, then their Fourier transforms are convolved in the frequency domain. If $x_1(t) \leftrightarrow X_1(\omega)$ and $x_2(t) \leftrightarrow X_2(\omega)$ and if $y(t) = x_1(t) * x_2(t) = \int\limits_{-\infty}^{\infty} x_1(\tau)x_2(t - \tau)\mathrm{d}\tau$, then

$$x_1(t) * x_2(t) \leftrightarrow X_1(\omega).X_2(\omega) \tag{8.48}$$

Typically for a practical LTI system, the output $y(t)$ is the convolution of the input $x(t)$ and the IRF $h(t)$. The convolution theorem implies that the Fourier transform of the output of such a system is the product of the Fourier transform of the input and the FRF, which is the Fourier transform of the IRF. Convolution obeys the commutative, associative, and distributive laws of algebra as shown below.

Commutative law : $x_1(t) * x_2(t) = x_2(t) * x_1(t)$ (8.49a)

Associative law : $x_1(t) * ((x_2(t) * (x_3(t)) = (x_1(t) * x_2(t)) * x_3(t)$ (8.49b)

Distributive law

$$x_1(t) * (x_2(t) + x_3(t)) = (x_1(t) * x_2(t)) + (x_1(t) * x_3(t)) \qquad (8.49c)$$

8.5.2.6 Modulation or Windowing

This is similar to convolution; here, multiplication in the time domain is equivalent to convolution in the frequency domain. If $x(t) \leftrightarrow X(\omega)$ and $w(t) \leftrightarrow W(\omega)$, then

$$y(t) = x(t).w(t) \leftrightarrow Y(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\theta).W(\omega - \theta)d\theta \qquad (8.50)$$

From Euler's relationship $\cos\theta = (e^{i\theta} + e^{-i\theta})/2$, we infer that multiplication of a time function by a cosine signal splits the frequency spectrum, causing half of the spectrum to shift to a negative frequency to the left-hand side and half to shift to the right-hand side to the corresponding positive frequency on the frequency axis. The frequency shift theorem states that $x(t)e^{i\omega t_0} \leftrightarrow X(\omega - \omega_0)$ and $x(t)e^{-i\omega t_0} \leftrightarrow X(\omega + \omega_0)$ (from Eq. 8.44b). From the linearity theorem, it follows that

$$x(t)\cos\omega_0 t \leftrightarrow \frac{1}{2}[X(\omega + \omega_0) + X(\omega - \omega_0)] \qquad (8.51a)$$

and

$$x(t)\sin\omega_0 t \leftrightarrow \frac{1}{2i}[X(\omega - \omega_0) - X(\omega + \omega_0)] \qquad (8.51b)$$

The modulation theorem is the basis of amplitude modulation.

8.5.2.7 Differentiation in Time

Differentiating a function in time causes a 90° phase shift in the spectrum and scales the magnitude of the spectrum in proportion to the frequency. Repeated differentiation leads to the result

$$\frac{d^n x(t)}{dt^n} \leftrightarrow (i\omega)^n X(\omega) \qquad (8.52)$$

This theorem helps us understand why differentiation of a signal has the reputation of being a noisy operation. If there is high-frequency noise in a signal, it will get greatly amplified by differentiation.

8.5.2.8   Differentiation in Frequency

This theorem states that multiplying a function $x(t)$ by t amounts to differentiating the signal in the frequency domain.

$$tx(t) \leftrightarrow i\frac{dX(\omega)}{d\omega} \tag{8.53a}$$

Repeated multiplications lead to the result

$$(-it)^n x(t) \leftrightarrow \frac{d^n X(\omega)}{d\omega^n} \tag{8.53b}$$

8.5.2.9   Integration

Integrating a function in the time domain introduces a $-90°$ phase shift in the spectrum and scales the magnitude of the spectrum inversely with frequency. Thus,

$$\int_{-\infty}^{t} x(u)du \leftrightarrow \frac{X(\omega)}{i\omega} + C \tag{8.54}$$

where C is a constant.

From this theorem, we see that the integration operation is akin to passing the signal through a low-pass filter which blurs the signal. Practically speaking, integrators are more popular than differentiators for the reason that they do not amplify high-frequency noise.

8.5.2.10   Parseval/Rayleigh Theorem

This theorem states that the total power in any physical system can be computed, either by integrating over the time domain or over the frequency domain. What is referred to as Parseval's energy conservation theorem in the context of Fourier series representation is often referred to as Rayleigh's theorem in the context of Fourier transforms.

$$\int_{-\infty}^{\infty} |x(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega \tag{8.55a}$$

The quantity $|X(\omega)|^2$ is called the energy density spectrum or power spectral density (PSD) and is a quantity which is extensively used for the description of random signals for which the Fourier transform does not exist. The integral on the left-hand side gives the total amount of energy in the signal as computed in the time domain,

Table 8.3 Fourier transform theorems

Time signal	Corresponding transform
$x(t) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} X(\omega)\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}\omega$	$X(\omega) = \int\limits_{-\infty}^{\infty} x(t)\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}t$
$ax_1(t) + bx_2(t)$	$aX_1(\omega) + bX_2(\omega)$
$x(at)$	$\frac{1}{|a|} X\left(\frac{\omega}{a}\right)$
$x^*(-t)$	$X^*(\omega)$
$x(t - t_0)$	$X(\omega)\mathrm{e}^{-\mathrm{i}\omega t_0}$
$x(t)\mathrm{e}^{\mathrm{i}\omega t_0}$	$X(\omega - \omega_0)$
$x(-t)$	$X(\omega)$
$X(t)$	$x(-\omega)$
$\int\limits_{-\infty}^{\infty} x_1(\tau)x_2(t - \tau)\mathrm{d}\tau$	$X_1(\omega)X_2(\omega)$
$x(t)\cos\omega_0 t$	$\frac{1}{2}[X(\omega + \omega_0) + X(\omega - \omega_0)]$
$x(t)\sin\omega_0 t$	$\frac{1}{2}[X(\omega - \omega_0) - X(\omega + \omega_0)]$
$\int\limits_{-\infty}^{t} x(u)\mathrm{d}u$	$\frac{X(\omega)}{\mathrm{i}\omega} + C$, where C is a constant
$\dfrac{\mathrm{d}^n x(t)}{\mathrm{d}t^n}$	$(\mathrm{i}\omega)^n\, X(\omega)$
$(-\mathrm{i}t)^n x(t)$	$\dfrac{\mathrm{d}^n X(\omega)}{\mathrm{d}\omega^n}$
$\int\limits_{-\infty}^{\infty} x(t + \tau)x^*(\tau)\mathrm{d}t$	$|X(\omega)|^2$
$\int\limits_{-\infty}^{\infty} |x(t)|^2\mathrm{d}t = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} |X(\omega)|^2\mathrm{d}\omega$	

whereas the integral on the right-hand side gives the total amount of energy computed in the frequency domain. The modulus is found since the integrand is, in general, complex and therefore it is the magnitude of this quantity which is integrated. A more general formulation of Parseval's theorem for two signals is as follows. If $x_1(t) \leftrightarrow X_1(\omega)$ and $x_2(t) \leftrightarrow X_2(\omega)$, then

$$\int\limits_{-\infty}^{\infty} x_1(t)x_2^*(t)\mathrm{d}t = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} X_1(\omega)X_2^*(\omega)\mathrm{d}\omega \qquad (8.55\mathrm{b})$$

Table 8.3 summarizes all the Fourier transform theorems seen and Table 8.4 shows the Fourier transforms of some common signals. Figure 8.17 illustrates the scaling property of the Fourier transform and Fig. 8.18 shows the Fourier transform of some common functions. In the above equations, rect and sinc functions are as defined in Eqs. (8.46) and (8.47).

Table 8.4 Fourier transforms of common signals

Function	Time waveform $g(t)$	Spectrum $G(f)$
Rectangular	$\text{rect}\left(\frac{T}{\tau}\right)$	$\tau\,\text{sinc}(\tau f)$
Constant	1	$\delta(f)$
Impulse at $t = t_0$	$\delta(t - t_0)$	$\exp(-i2\pi t_0 f)$
Sinc	$\text{sinc}(2Wt)$	$\frac{1}{2W}\text{rect}\left(\frac{f}{2W}\right)$
Phasor	$\exp[i(2\pi f_0 t + \phi)]$	$\exp(i\phi)\delta(f - f_0)$
Harmonic	$\cos(2\pi f_c t + \phi)$	$\frac{1}{2}[\exp(i\phi)\delta(f - f_c) + \exp(-i\phi)\delta(f + f_c)]$
Impulse train	$\displaystyle\sum_{k=-\infty}^{\infty} \delta(t - kT)$	$\displaystyle f_0 \sum_{n=-\infty}^{\infty} \delta(f - nf_0), \quad f_0 = \frac{1}{T}$
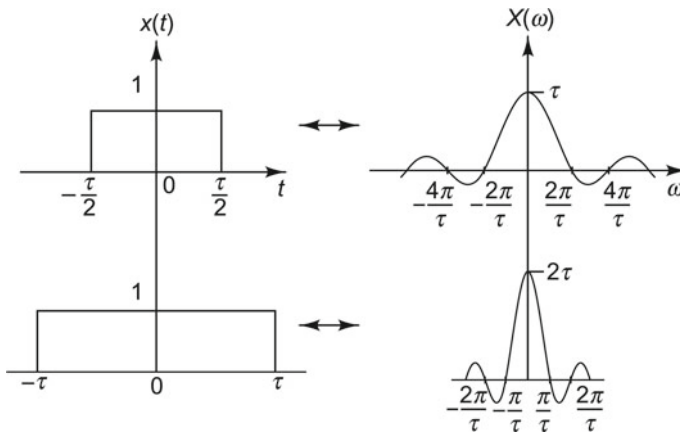
Fig. 8.17 Fourier transform-scaling property

8.5.3 Fourier Transform of Sequences

The Fourier transform of a sequence $x[n]$ can be represented as

$$X(e^{i\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-i\omega n} \tag{8.56a}$$

and the inverse Fourier transform is defined as

$$x[n] = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(e^{i\omega})(e^{i\omega n})d\omega \tag{8.56b}$$
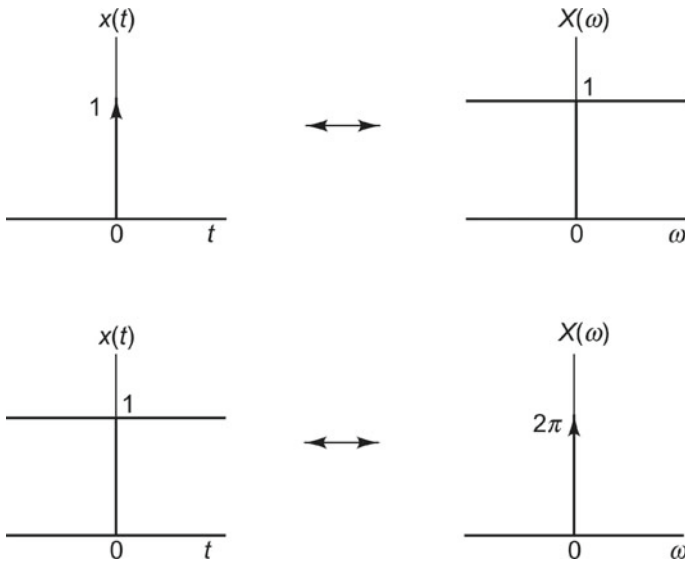
Fig. 8.18 Common Fourier transforms

$X(e^{i\omega})$ is a continuous function of ω and is periodic in ω with period 2π. $x[n]$ is seen to be a superposition of infinitesimally small complex sinusoids of the form $\frac{1}{2\pi} X(e^{i\omega}) e^{i\omega n} d\omega$ with frequency ω ranging over a length 2π. Thus, there is a clear equivalence between the Fourier series representation of a continuous-time periodic signal and the Fourier transform of a discrete-time signal. Hence, all properties of the Fourier series can be applied to the Fourier transform representation of a sequence with appropriate interpretation of variables. All the Fourier transform theorems explained earlier hold good for sequences also, with some minor variations.

8.6   Sampling of Continuous-Time Signals

DID YOU KNOW that Harry Nyquist was the first scientist to hint at the sampling process in 1928? In his paper, related to telegraph transmission theory, he proved that as many as 2B independent pulse samples could be transmitted through a system of bandwidth B. However, he did not explicitly talk about sampling and reconstruction of continuous-time signals. Around the same time, Karl Küpfmüller came up with the theory that the impulse response of a band-limited filter is a sinc function; this filter is used as a reconstruction filter after sampling and is sometimes called Küpfmüller filter. In 1949, American mathematician, electrical engineer, and cryptographer Claude E.

Shannon proved the sampling theorem, which is essentially a continuation of Nyquist's result. Other scientists who had published similar results earlier were British mathematician Edmund Taylor Whittaker in 1915 and V.A. Kotelnikov from Russia in 1933, British mathematician J. M. Whittaker in 1935, and Hungarian-British electrical engineer Dennis Gabor in 1946.

Often any signal to be analysed is analogue in nature as mentioned earlier, typically being a continuous function of time (or space in a digitized image), or any other independent variable. It is generally captured as a series of values recorded on a tape, data logger, or computer. It must then be converted into digital form by an analogue-to-digital converter for further processing. For this conversion, it is necessary to sample the continuous-time signal in some manner. Any analogue signal $x_c(t)$ can be converted to a discrete-time form by sensing values of $x_c(t)$ at regular intervals of time nT_s (where n is an integer), to obtain sequence $x[n]$. T_s is called the sampling period or sampling interval and this type of sampling done at regular intervals is called uniform sampling. It is customary to specify a sampling rate or frequency $f_s = (1/T_s)$ in hertz, rather than the sampling period in seconds. Each sample value resulting from sampling is associated with the instant of time when it was captured. The samples can be used to reconstruct the original signal by simply joining the points to form the envelope of the original analogue signal $x_c(t)$ and this process is called reconstruction; for this, the sampling frequency should be known. If there is a rapid variation of the signal with respect to time, it implies that there are high-frequency components in the signal. Intuitively, one could say that the higher the frequencies of the components present in the signal, the higher the sampling frequency should be. If the sampling frequency is not high enough, or if we perform 'undersampling', some of the information that is present in the signal may be lost due to a phenomenon called aliasing. In cinematography, aliasing manifests as the wagon-wheel effect, in which a spoked wheel appears to rotate too slowly or even backwards due to an insufficient frame rate. The reconstructed signals are better with higher sampling frequencies or with 'oversampling'; however, this calls for faster ADCs and more storage, increasing the cost of signal processing. Therefore, the consideration of sampling frequency is very important and one should weigh the pros and cons of each application, and be aware of the trade-offs involved. In practice, the sampling interval is typically quite small, of the order of milliseconds or microseconds for vibration and acoustic signals. Figure 8.19 shows the effect of sampling at a sufficiently high frequency.

8.6.1 Undersampling and Aliasing

If the sampling rate (frequency) is not high enough, it is impossible to rebuild the original signal from the samples. In fact, the actual high frequency components in the original signal take on the identity or alias of lower frequencies and this phenomenon is called aliasing. Specifically, undersampling causes signal components with a fre-

quency higher than half the sampling frequency to appear as low frequencies which are not actually present in the original signal, or which are indistinguishable from genuine low-frequency components and get superimposed over them, causing distortion of the signal. Besides, some of the frequencies in the original signal may be irreparably lost in the reconstructed signal. Aliasing occurs because signal frequencies can overlap if the sampling frequency is too low and frequencies "fold" about half the sampling frequency, which is also called the folding frequency as will be evident from the theoretical explanation in Sect. 8.6.5. If for instance, we listen to music with high-frequency components, which has been reconstructed after sampling at too low a frequency, the low-frequency aliases of the undersampled high frequencies will be audible. Figure 8.20 shows two cases of undersampling and aliasing: (i) $f_s = 2f_0$ and (ii) $f_s < 2f_0$, where f_0 is the frequency of the sine wave and f_s is the sampling frequency. The reconstructed signals built from the data samples will not represent the original signals at all in both cases since the sampled values do not have enough information to do so. To prevent or reduce aliasing, two things can be done: (i) increase the sampling rate, to above twice the highest frequency of interest and (ii) introduce an anti-aliasing (low-pass) filter to limit the frequencies in the analogue signal to less than half the sampling frequency.

Fig. 8.19 Sampling at a sufficiently high rate
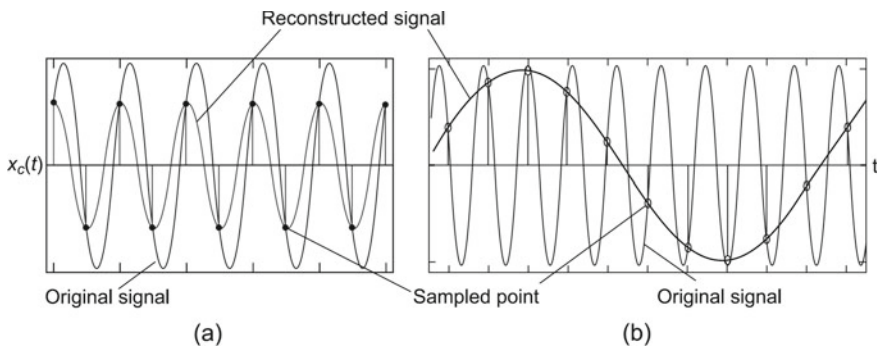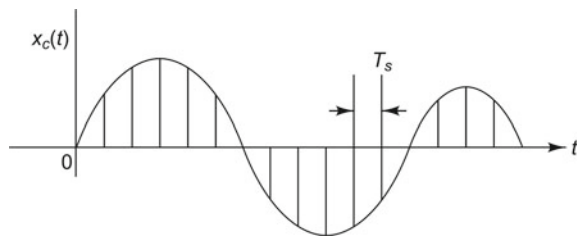
Fig. 8.20 Depiction of undersampling: **a** $f_s = 2f_0$, **b** $f_s < 2f_0$
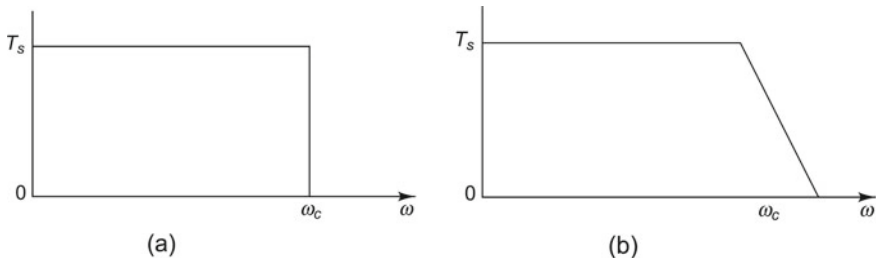
Fig. 8.21 Anti-aliasing filter: **a** ideal, **b** actual

8.6.2 Anti-aliasing Filter

Sometimes the highest frequency components of a signal may not be of interest for the analysis or may simply be noise. The amount of aliasing due to such signals may be reduced by restricting the bandwidth and filtering out these higher frequency components before sampling the signal. This is done using an anti-aliasing filter, which is nothing but an ideal low-pass filter (which filters out high-frequency components and lets lower frequency components through) with very sharp cut-off characteristics. Theoretically, a perfect boxcar low-pass filter with a cut-off at exactly the Nyquist frequency (highest frequency in the signal) should eliminate all the frequencies above the Nyquist frequency. However, ideal filters with infinite attenuation at the cut-off frequencies cannot be realized. All practical filters only cause considerable attenuation of frequencies beyond the cut-off frequency, and do not remove them completely. However, the low-frequency aliases thus generated have very low-amplitude levels, so as not to cause a problem. Due to the sharp roll off requirements, anti-aliasing filters are expensive. Through combinations of digital and analogue low-pass filter technology, industry has been able to provide anti-aliasing filters having an attenuation of 60–120 dB/octave beyond the cut-off frequency. The cut-off frequency is generally set at 60–80% of the Nyquist frequency, so as to drive the aliasing signals into the analyser noise. Figure 8.21a shows the characteristics of an ideal anti-aliasing filter and Fig. 8.21b that of an actual anti-aliasing filter.

For example, if the highest frequency of interest in a signal is 500 Hz, the filter is typically set at 300–400 Hz depending on the sharpness of the anti-aliasing filter. This implies that one will lose the frequency components between 300 and 500 Hz, since their magnitude and phase information have got modified. This is the price that one pays for anti-aliasing protection.

8.6.3 Sampling Theory

In the case of periodic or regular sampling, a sequence of samples $x[n]$ is obtained from a continuous time signal $x_c(t)$ according to the relation
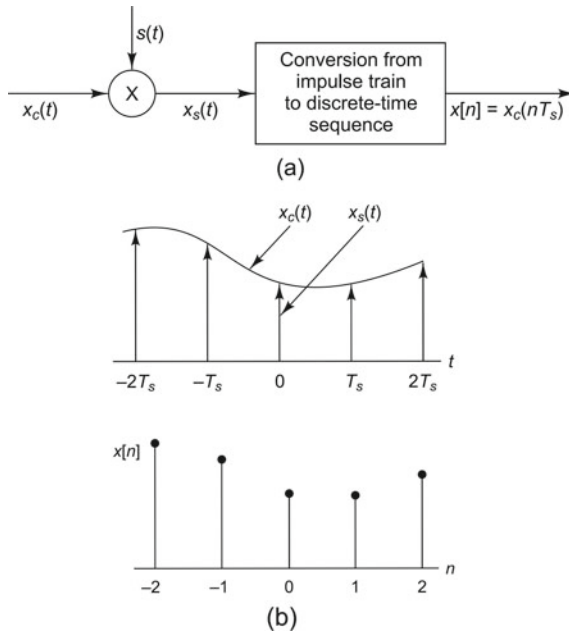
Fig. 8.22 Sampling operation: **a** conversion from analogue signal to sequence, **b** sampling of continuous-time signal and resulting output sequence

$$x[n] = x_c(nT_s), \quad -\infty < n < \infty \tag{8.57}$$

where T_s is the sampling period and $f_s = (1/T_s)$ is the sampling frequency. It is convenient to mathematically represent the sampling process in the two stages depicted in Fig. 8.22a. It consists of an impulse train modulator followed by conversion of the impulse train to a sequence. Let $x_c(t)$ represent the analogue signal to be sampled. Sampling in the time domain is done mathematically by multiplying $x_c(t)$ with a periodic impulse train of period T_s (train of delta functions T_s apart). This corresponds to a point by point multiplication of the sampling impulse function with $x_c(t)$. By doing so, only the values of $x_c(t)$ corresponding to the instants at which the impulse train occurs are sifted, the product being zero for all other times. This is analogous to sampling $x_c(t)$ with a frequency $f_s = (1/T_s)$, resulting in the new sampled signal denoted as $x_s(t)$. Figure 8.22b illustrates a continuous-time signal, the result of impulse train sampling and the corresponding output sequence. The essential difference between $x_s(t)$ and $x[n]$ is that $x_s(t)$ is, in a sense a continuous-time signal (specifically an impulse train) which is 0 except at integral multiples of T_s. Whereas, the sequence $x[n]$ is indexed on the integer variable n (which is just a constant) and does not contain any explicit information about sampling rate. Furthermore, samples of $x_c(t)$ are represented as finite numbers in $x[n]$, rather than as the areas of impulses as in $x_s(t)$.

8.6.3.1   Frequency Domain Representation of Sampling

To derive the relation between input and output of an ideal continuous-to-discrete (C/D) converter in the frequency domain, let us first consider conversion of $x_c(t)$ to $x_s(t)$ through an impulse train modulator. The modulating signal or sampling signal $s(t)$ is a periodic impulse train.

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \tag{8.58}$$

where $\delta(t)$ is the unit impulse function or Dirac delta function.

With time interval T_s between consecutive samples of the continuous-time signal $x_c(t)$, the sampled time signal $x_s(t)$ can be represented as

$$x_s(t) = \begin{cases} x_c(t).s(t) \\ x_c(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \\ \sum_{n=-\infty}^{\infty} x_c(nT_s)\delta(t - nT_s) \end{cases} \tag{8.59}$$

Through sifting property of the impulse function, $x_s(t)$ can be expressed as

$$x_s(t) = x_c(nT_s), \quad n = \pm 1, \pm 2, \pm 3, \ldots \tag{8.59b}$$

The spectrum of the sampled signal, $x_s(t)$ may be obtained by taking its Fourier transform.

$$X_s(f) = \Im\{x_s(t)\} = \int_{-\infty}^{\infty} x_s(t)e^{-i2\pi ft}\, dt \tag{8.60}$$

It may be noted that $x_s(t)$ is obtained as the product of $x_c(t)$ and $s(t)$, which is a train of impulses. The multiplication of two functions in the time domain corresponds to the convolution of the Fourier transforms of the two functions in the frequency domain. Thus, the Fourier transform of $x_s(t)$ is the convolution of the Fourier transform of $X_c(\omega)$ and $S(\omega)$ which can be expressed as

$$X_s(\Omega) = \frac{1}{2\pi}X_c(\Omega) * \Im\left\{\sum_{n=-\infty}^{\infty} \sigma(t - nT_s)\right\} \tag{8.61}$$

Since we know the spectrum of the original signal, $x_c(t)$, we need find only the Fourier transform of the train of impulses. Since $s(t)$ is an impulse train which is a periodic function in the time domain, it can be represented by a Fourier series. Thus, we may write

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) = \sum_{n=-\infty}^{\infty} c_n e^{i2\pi nt/T_s} \qquad (8.62a)$$

where the Fourier series coefficients are

$$c_k = \frac{1}{T_s} \int_{-T_s/2}^{T_s/2} \sum_{n=-\infty}^{\infty} \delta(t - nT_s) e^{-i2\pi nt/T_s} dt = \frac{1}{T_s} \qquad (8.62b)$$

The Fourier series of an impulse function with spacing T_s in the time domain is seen to result in an impulse train with spacing $1/T_s$ in the frequency domain. In evaluating this, the limits of integration are specified for one period only (avoiding placement of impulses at the integration limits). Thus, $s(t)$ assumes the following simplified expression: which is easily Fourier transformable

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} e^{i2\pi nt/T_s} \qquad (8.62c)$$

Since any signal can be synthesized from its inverse Fourier transform and since

$$\Im^{-1}\{\delta(\omega - \omega_0)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \delta(\omega - \omega_0) e^{i\omega t} d\omega = e^{i\omega_0 t} \qquad (8.63)$$

we can express $e^{i2\pi nt/T_s}$ in Eq. (8.62c) as

$$e^{i2\pi nt/T_s} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \delta\left(\Omega - \frac{2\pi}{T_s}\right) e^{i\Omega t} d\Omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \delta(\Omega - \Omega_s) e^{i\Omega t} d\Omega \qquad (8.64)$$

Therefore, Fourier transform of Eq. (8.62c) is

$$S(\Omega) = \Im\left\{ \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \right\} = \frac{2\pi}{T_s} \sum_{k=-\infty}^{\infty} \delta\left(\Omega - k\frac{2\pi}{T_s}\right)$$

$$= \frac{2\pi}{T_s} \sum_{k=-\infty}^{\infty} \delta(\Omega - k\Omega_s) \qquad (8.65)$$

Thus, the Fourier transform of a periodic impulse train in the time domain is a periodic impulse train in the frequency domain.

We can now express the Fourier transform of the sampled signal as follows:

$$X_s(\Omega) = \begin{cases} \frac{1}{2\pi} X_c(\Omega) * \frac{2\pi}{T_s} \sum_{k=-\infty}^{\infty} \delta(\Omega - k\Omega_s) \\ \frac{1}{T_s} \sum_{k=-\infty}^{\infty} X_c(\Omega - k\Omega_s) \end{cases} \tag{8.66}$$

Equation (8.66) provides the relationship between Fourier transforms of the input and output of the impulse train modulator. It implies that the Fourier transform of $x_s(t)$ consists of periodically repeated copies of $X_c(\Omega)$, i.e., the Fourier transform of $x_c(t)$. The copies of $X_c(\Omega)$ are shifted by integral multiples of the sampling frequency and then superimposed to produce the periodic Fourier transform.

It is also to be remembered that $x_s(t)$ is the sampled signal and $x[n]$ the resulting sequence. Since $x[n] = x_c(nT_s)$, to get the Fourier transform of $x[n]$, i.e., $\Im\{x[n]\} = X(e^{i\omega})$, we have to do some normalization of the frequency axis so that frequency $\Omega = \Omega_s$ in $X_s(\Omega)$ is normalized to $\omega = 2\pi$ for $X(e^{i\omega})$. This is in view of the fact that $X(e^{i\omega})$ is periodic in ω with period 2π, while $X_s(\Omega)$ is periodic in Ω with period Ω_s. Thus, $X(e^{i\omega})$ is simply a frequency-scaled version of $X_s(\Omega)$ with the scaling specified by $\omega = \Omega T_s$. Therefore,

$$X_s(\Omega) = X(e^{i\omega}) = X(e^{i\Omega T_s}) \tag{8.67}$$

Hence,

$$X(e^{i\omega}) = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} X_c\left(\frac{\omega}{T_s} - k\frac{2\pi}{T_s}\right) \tag{8.68}$$

This is the Fourier transform of the sequence $x[n]$.

8.6.4 Nyquist–Shannon Sampling Theorem

The sampling theorem was first proposed by Harry Nyquist in 1928, but he did not explicitly consider the problem of sampling and reconstruction of continuous signals. Later in 1949, Claude E. Shannon came up with the proof of the sampling theorem. The theorem describes two important processes: the sampling process, in which an analogue signal is converted to a discrete-time signal and the reconstruction process, in which the original analogue signal is recovered from the discrete-time signal. This theorem can be stated as follows.

Let $x_c(t)$ be a band-limited signal with $X_c(\Omega) = 0$ for $|\Omega| > \Omega_N$. Then $x_c(t)$ is uniquely determined by its samples $x[n] = x_c[nT_s]$, $n = 0, \pm1, \pm2, \ldots$ if $\Omega_s = \frac{2\pi}{T_s} > 2\Omega_N$. The highest frequency in the signal Ω_N is commonly referred to as the Nyquist frequency. The Nyquist sampling theorem states that the sampling frequency Ω_s should be greater than twice the highest frequency Ω_N present in the signal in order to prevent aliasing from occurring and in order to be able to reproduce the original signal exactly. Given the highest frequency Ω_N in the original signal, the theorem gives the lower bound on the sampling frequency for which perfect reconstruction can

be achieved. This lower bound $2\Omega_N$ is called the Nyquist rate. If instead, the sampling frequency is given, the theorem gives an upper bound for frequency components $\Omega_N < \Omega_s/2$. This upper bound is called the Nyquist frequency and is denoted as Ω_N. Another name for the Nyquist frequency is folding frequency. When the condition $\Omega_s > 2\Omega_N$ is met for the highest frequency component in the original signal, it is said that the Nyquist criterion is satisfied. The signal to be sampled must be band-limited in the sense that the spectral amplitudes of frequency components with frequencies above the cut-off frequency should be zero or close to zero so that their influence on the reconstructed signal will be minimal. However, in practice, a signal is never perfectly band-limited and one may be forced to use an anti-aliasing filter.

8.6.5 *Mathematical Sampling and Aliasing: Harmonic Signal*

Understanding what aliasing does to individual sinusoids is a big help in understanding what happens when they are superimposed over each other. Envisaging the sampling process with sinusoidal signals is relatively easy since they have only one frequency component. Figure 8.23 shows the effect of sampling a cosine signal in the frequency domain. Figure 8.23a shows the Fourier transform of the signal $x_c(t) = \cos(\Omega_0 t)$. Figure 8.23b shows the spectrum of the impulse train which does the sampling. Figure 8.23c shows the Fourier transform of the sampled signal $x_s(t)$ for the case of no aliasing, i.e., with $\Omega_0 < (\Omega_s/2)$. It can be seen that we have copies or replicas of the original spectrum shown in Fig. 8.23a about $0, \Omega_s, 2\Omega_s, 3\Omega_s$, etc. It is seen that these replicas do not overlap, making it possible to extract the original spectrum from that of the sampled signal by passing the sampled signal through a reconstruction (low-pass) filter with gain T_s and cut-off frequency $\Omega_0 < \Omega_c < (\Omega_s/2)$ as shown in Fig. 8.23d. Thus, for cases where sampling has been done with a sufficiently high sampling frequency as to not cause aliasing, the choice of Ω_c is straightforward. Figure 8.23e corresponds to the Fourier transform of the low-pass filtered output for $\Omega_0 < (\Omega_s/2) = (\pi/T_s)$ with $\Omega_c = (\Omega_s/2)$. Figure 8.23f shows the Fourier transform of $x_s(t)$ for $\Omega_0 > (\Omega_s/2)$, i.e., with aliasing arising from undersampling. It can be seen that the replicas overlap in this case and hence the original signal spectrum cannot be retrieved using a low-pass filter as shown in Fig. 8.23g which corresponds to the Fourier transform of the low-pass filtered output for $\Omega_0 > (\pi/T_s)$ with $\Omega_c = (\Omega_s/2)$. The reconstructed output is $x_r(t) = \cos(\Omega_0 t)$ without aliasing (Fig. 8.23d) and with aliasing the reconstructed output is $x_r(t) = \cos(\Omega_s - \Omega_0)t$ as seen in Fig. 8.23g, i.e., the frequency Ω_0 has taken on the identity or alias of the low-frequency signal $(\Omega_s - \Omega_0)$ as a consequence of the sampling and reconstruction processes.

$X_c(\Omega)$

π

$-\Omega_0$ 0 Ω_0

Ω

(a)

$S(\Omega)$

$2\pi/T_s$

$-2\Omega_s$ $-\Omega_s$ 0 Ω_s $2\Omega_s$

Ω

(b)

$X_s(\Omega)$

π/T_s

Ω_s

$-2\Omega_s$ $-\Omega_s$ $-\Omega_0$ 0 Ω_0 $\Omega_s-\Omega_0$ $\Omega_s+\Omega_0$ $2\Omega_s$

Ω

(c)

$H_r(\Omega)$

T_s

$-\Omega_s/2$ $\Omega_s/2$

$-\Omega_0$ 0 Ω_0
$-\Omega_c$ Ω_c

Ω

(d)
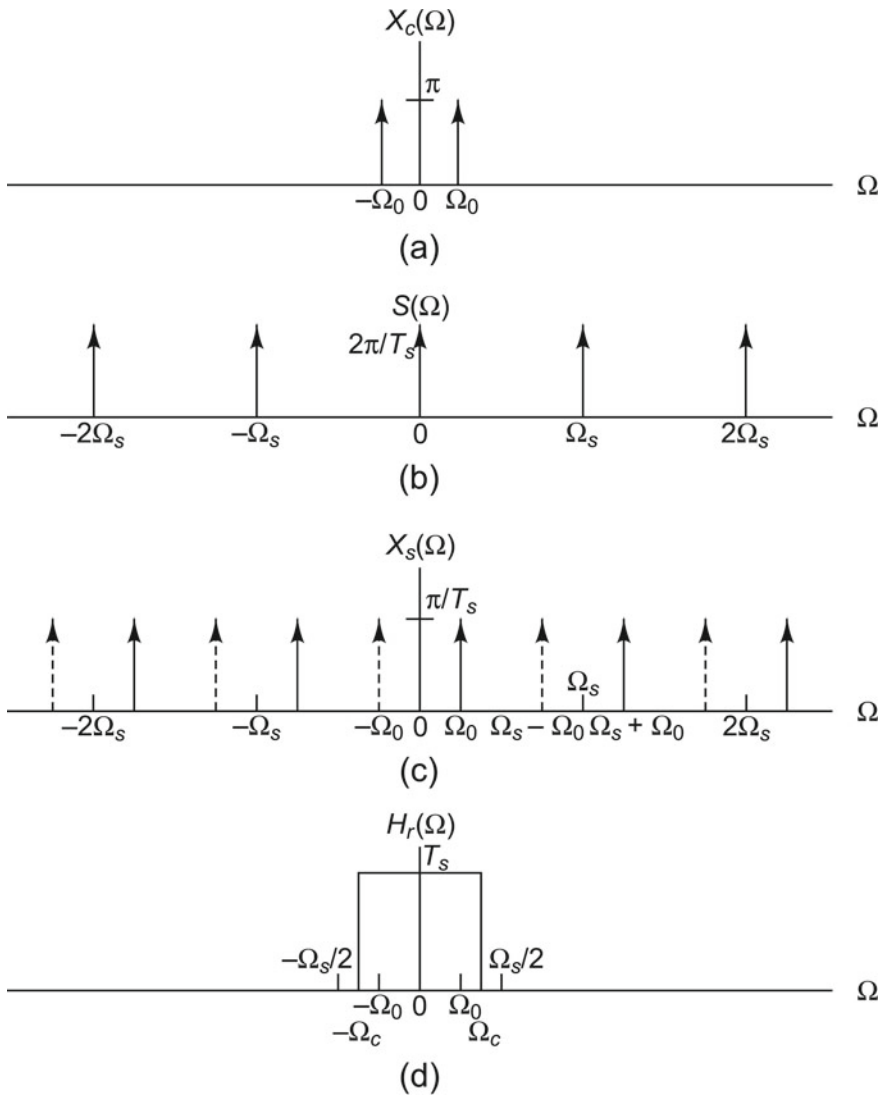
Fig. 8.23 Sampling and aliasing for a cosine signal:**a** $X_c(\Omega)$, **b** $S(\Omega)$ with $\Omega_0 < (\Omega_s/2)$, i.e., no aliasing, **c** $X_s(\Omega)$ with $\Omega_0 < (\Omega_s/2)$, i.e., no aliasing, **d** $H_r(\Omega)$ with $\Omega_0 < \Omega_c < (\Omega_s/2)$, **e** $X_r(\Omega)$ for case without aliasing, **f** $X_s(\Omega)$ with $\Omega_0 > (\Omega_s/2)$, i.e., with aliasing, **g** $X_r(\Omega)$ for the case with aliasing

8.6.6 Band-Limited Sampling and Aliasing

In practice, signals are quite often band-limited signals with a definite cut-off frequency. To avoid aliasing in the sampled band-limited signal, we must ensure that the

$$X_r(\Omega)$$

(e)

$$X_s(\Omega)$$

(f)

$$X_r(\Omega)$$

(g)
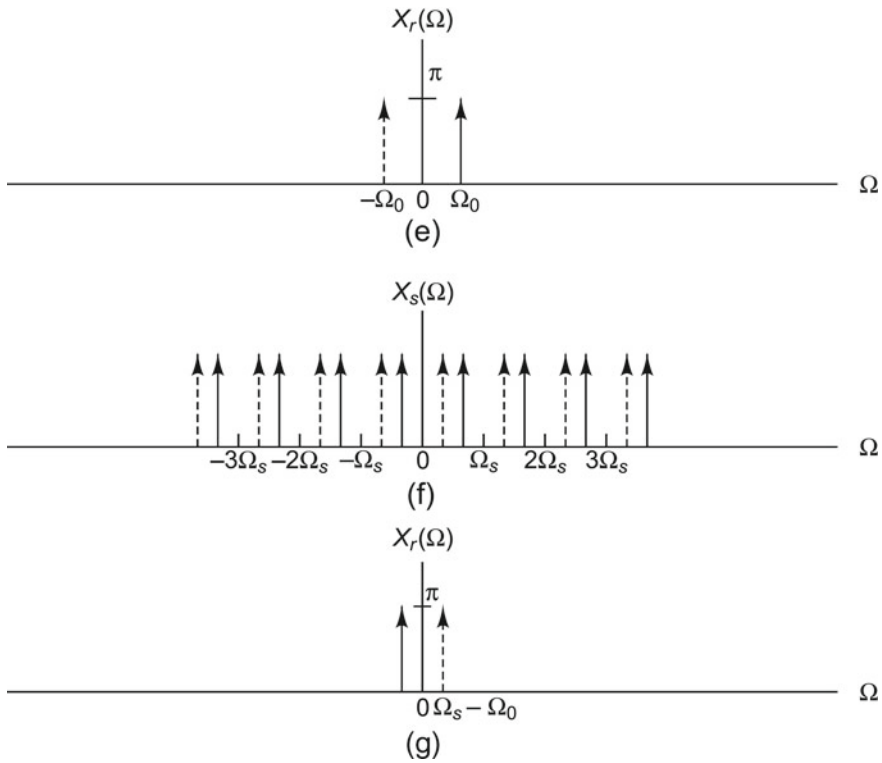
Fig. 8.23 (continued)

spectrum replicas do not overlap. Figure 8.24 shows the sampling and reconstruction processes for a band-limited signal $x_c(t)$. Figure 8.24a shows the Fourier transform of $x_c(t)$, where the highest non-zero frequency component in $X_c(\Omega)$ is Ω_N. Figure 8.24b represents the periodic impulse train $S(\Omega)$. Figure 8.24c shows $X_s(\Omega)$ obtained as a result of convolving $X_c(\Omega)$ with $S(\Omega)$. From this figure, it is evident that when $\Omega_s - \Omega_N > \Omega_N$ or $\Omega_s > 2\Omega_N$, the replicas of $X_c(\Omega)$ do not overlap and therefore when they are added together in Eq. (8.65), there is formed (to within a scale factor of $1/T_s$), a replica of $X_c(\Omega)$ at each integral multiple of Ω_s. Therefore, $X_c(\Omega)$ can be recovered from $X_s(\Omega)$ through the process $X_r(\Omega) = H_r(\Omega)X_s(\Omega)$, where $H_r(\Omega)$ is an ideal low-pass filter having frequency response as shown in Fig. 8.24d and with cut-off frequency Ω_c such that $\Omega_N < \Omega_c < (\Omega_s - \Omega_N)$. The recovered signal $X_r(\Omega)$ would be as shown in Fig. 8.24e, where it is assumed that $\Omega_s > 2\Omega_N$. If this inequality does not hold good, i.e., if $\Omega_s \leq 2\Omega_N$, the copies of $X_c(\Omega)$ overlap so that when they are added together, $X_c(\Omega)$ is no longer recoverable by low-pass filtering. This is brought out in Fig. 8.24f. In this case, the reconstructed output $x_r(t)$ has got aliased.

8.6.7 Reconstruction of Original Signal Using Reconstruction Filters

Reconstruction of the original signal is possible if sampling has been done at a sufficiently high frequency in the first place without leading to aliasing and provided that we know the sampling period. Reconstruction is an interpolation process that mathematically defines a continuous-time signal $x_s(t)$ from the discrete samples of sequence $x[n]$ and at instants of time between the sampling instants nT_s. Given the sequence $x[n]$, it is possible to form the impulse train $x_s(t)$ in which the impulses have areas equal to the successive sequence values.

$$x_s(t) = \sum_{n=-\infty}^{\infty} x[n]\delta(t - nT_s) \tag{8.69}$$

The nth sample is associated with the impulse at $t = nT_s$, where T_s is the sampling period and $\Omega_s = 2\pi/T_s$ is the sampling frequency. If the impulse train is fed to an ideal low-pass continuous-time filter with impulse response $h_r(t)$ and frequency response $H_r(\Omega)$, the output is

$$x_r(t) = \sum_{n=-\infty}^{\infty} x[n]\, h_r(t - nT_s) \tag{8.70}$$

Reconstruction of a band-limited signal from its samples $x[n]$ is done as shown in Fig. 8.25.

This ideal reconstruction (low-pass) filter has a gain T_s to compensate for the factor $1/T_s$ incurred during the sampling process. It is a rectangular function in the frequency domain with a cut-off frequency $\Omega_N < \Omega_c < \Omega_s - \Omega_N$. This corresponds to a sinc function in the time domain as shown below.

$$h_r(t) = \frac{\sin(\pi t/T_s)}{\pi t/T_s} \tag{8.71}$$

Figure 8.26 shows the IRF and FRF of the reconstruction filter.

This function $h_r(t)$ is of infinite duration in the time domain, which means that it is impossible to have a perfect reconstruction filter in this domain. Substituting for $h_r(t)$ from Eq. (8.71) into Eq. (8.70), we get

$$x_r(t) = \sum_{n=-\infty}^{\infty} x[n]\frac{\sin[\pi(t - nT_s)/T_s]}{\pi(t - nT_s)/T_s} \tag{8.72}$$

Thus, the ideal reconstruction process is obtained by multiplication in the frequency domain of $X_s(\Omega)$ and $H_r(\Omega)$ or the convolution in the time domain as shown by Eq. (8.72) and involves superposition of scaled and delayed sinc functions ($h_r(t)$)
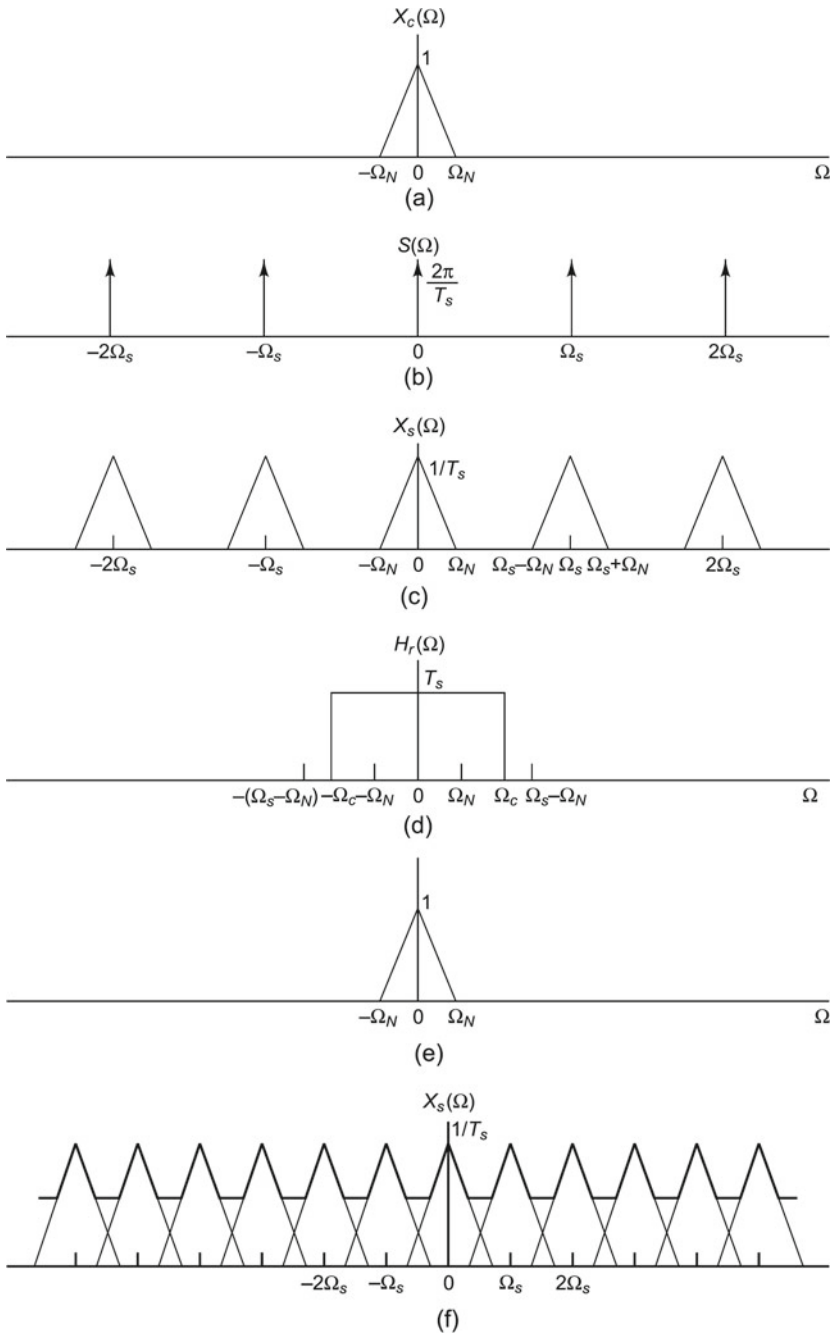
Fig. 8.24 Depiction of sampling and reconstruction for a band-limited signal: **a** $X_c(\Omega)$, **b** $S(\Omega)$ with $\Omega_N < \frac{\Omega_s}{2}$, i.e., no aliasing, **c** $X_s(\Omega)$ with $\Omega_N < \frac{\Omega_s}{2}$, i.e., no aliasing, **d** $H_r(\Omega)$ with $\Omega_N < \Omega_c < (\Omega_s - \Omega_N)$, **e** $X_r(\Omega)$ for case without aliasing, **f** $X_s(\Omega)$ with $\Omega_s \leq 2\Omega_N$, i.e., with aliasing
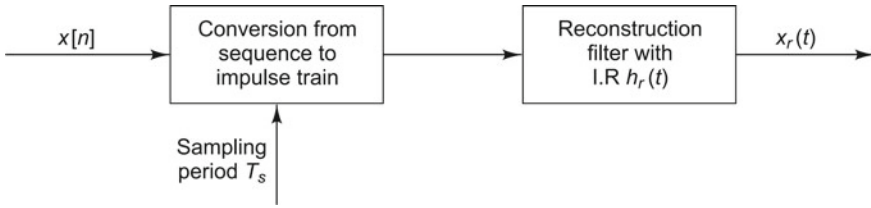
Fig. 8.25 Reconstruction process
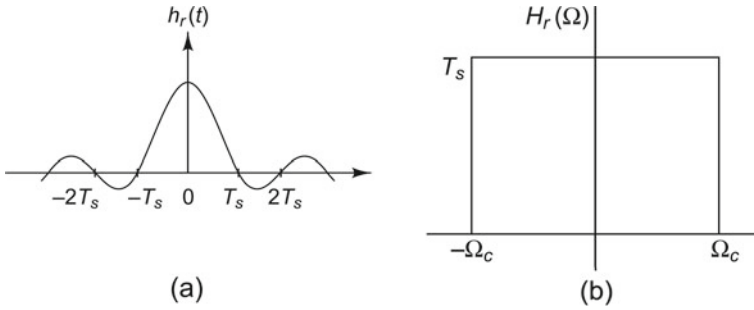
Fig. 8.26 Ideal reconstruction filter: **a** $h_r(t)$, **b** $H_r(\Omega)$
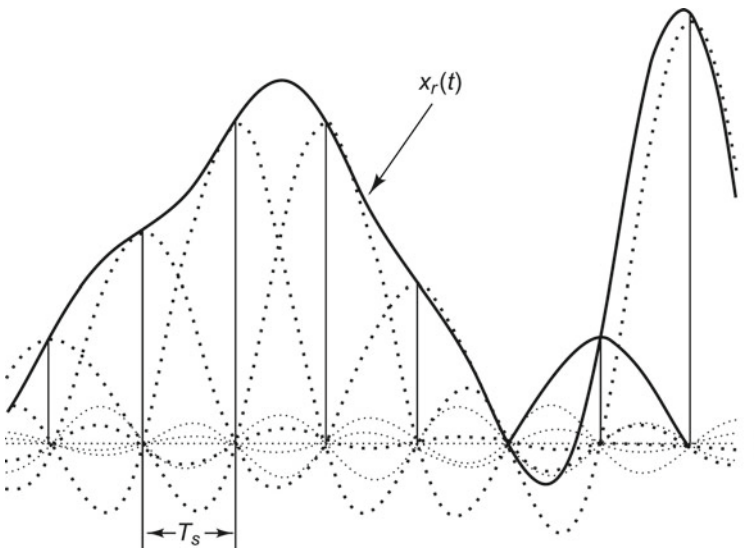
Fig. 8.27 Reconstruction of band-limited signals

as depicted in Fig. 8.27. The scaled and time-shifted sinc functions are continuous, making the superposition of these also continuous. Ideal reconstruction cannot be realized in practice since it implies that each sample contributes to the reconstructed signal at almost all instants of time, requiring summing an infinite number of terms. Therefore, some type of approximation of the sinc function, which is finite in length, has to be used and leads to an interpolation error.

8.7 The Fast Fourier Transform

DID YOU KNOW that the German scientist Carl Friedrich Gauss in 1805 developed a form of the Discrete Fourier Transform (DFT) even before Fourier published his results in 1822? Gauss' algorithm was similar to the one of Cooley and Tukey. Unfortunately, he did not publish his findings in his lifetime, since he found other methods that seemed to be more useful to solve his problem of determining the orbit of certain asteroids from sample locations. Scientists did not realize this when Gauss' collection of research publications was printed in 1866. It was much later, in fact 160 years later, that in 1965, Cooley and Tukey reinvented the fast Fourier transform (FFT). Between 1805 and 1965, several scientists in Great Britain and in other countries developed efficient algorithms to calculate the DFT. All of these were unrelated to Gauss' work and were not as general or as well formulated as Gauss' or the one of Cooley and Tukey. In 1965, the US military was interested in analysing seismological time series data. Dr. Richard L. Garwin at IBM Watson Research Centre thought that with a very fast Fourier transform, sensors planted at several locations in the ground could locate nuclear explosions. He suggested the idea of how such a transform could be programmed to be much faster and presented it to both John Wilder Tukey, who was in President Kennedy's Science Advisory Committee and James W. Cooley at IBM Watson Research Centre. They implemented his idea, leading to the successful location of nuclear explosions to within 15 km of where they were occurring. This efficient algorithm proposed by Cooley and Tukey for the calculation of the DFT revolutionized the field of digital signal processing. In the years that followed, many extensions and modifications of the original algorithm were made and the form that is used today had been frozen by the early 1970s.

The continuous Fourier transform converts a continuous-time signal of infinite duration into a continuous-frequency spectrum with an infinite number of sinusoidal components as we have seen before. A discrete Fourier transform (DFT) on the other hand, deals with sequences which are of finite duration and periodic and which are obtained by sampling a continuous-time signal at regular intervals. This is similar to the description given in Sect. 8.5.3, except that the sequence in this case is assumed

to be periodic and of finite duration. It is assumed that the sequence is obtained from a continuous-time signal $x(t)$ of duration T and that the data are taken with a sampling period of T_s and that sampling has been done without aliasing. Thus, $T = NT_s$ where N is the total number of samples taken in the time duration $[0, T]$. Due to this process is implicit, the assumption that the data being analysed have a fundamental frequency of $1/T$ Hz; i.e., the data are periodic in the time window of length T seconds. It is to be noted that this is not usually true with real data. The signal $x(t)$ can be represented as

$$x(t) = \sum_{n=0}^{N-1} x(nT_s)\delta(t - nT_s) \tag{8.73a}$$

The Fourier transform for this signal may be written as

$$X(f) = \begin{cases} \int\limits_{-\infty}^{\infty} x(t)e^{-i2\pi ft}\,dt \\ \sum\limits_{n=0}^{N-1} x(nT_s)e^{-i2\pi fnT_s} \end{cases} \tag{8.73b}$$

The computation is restricted to N frequencies: $0, \frac{1}{T_s}, \frac{2}{T_s}, \frac{3}{T_s}, \ldots, \frac{N-1}{T_s}$. Setting $f = \frac{k}{T} = \frac{k}{NT_s}$ in the above equation, we obtain the kth element in the DFT equation as

$$X[k] = \begin{cases} \sum\limits_{n=0}^{N-1} x[n]e^{-i(2\pi/N)kn}, \;\; k = 0, 1, 2, 3, \ldots, N-1 \\ \sum\limits_{n=0}^{N-1} x[n]W_N^{kn} \end{cases} \tag{8.74a}$$

where

$$W_N = e^{-i(2\pi/N)} \tag{8.74b}$$

Thus, N samples of the sequence $x[n]$ are taken and transformed into N complex Fourier coefficients (at integral multiples of the base frequency $1/T$ Hz). Though $X[k]$ itself is approximate, the original sequence $x[n]$ can be got from it exactly, allowing the reconstruction of the original waveform.

The nth element in the inverse DFT (IDFT) is computed as

$$x[n] = \begin{cases} \frac{1}{N} \sum\limits_{k=0}^{N-1} X[k]e^{i(2\pi/N)kn}, n = 0, 1, 2, 3, \ldots, N-1 \\ \frac{1}{N} \sum\limits_{k=0}^{N-1} X[k]W_N^{-kn} \end{cases} \tag{8.75}$$

Equation (8.74) may be simplified as

$$X[k] = \begin{cases} \sum_{n=0}^{N-1} (\text{Re}\{x[n]\} + i\,\text{Im}\{x[n]\})(\text{Re}\{W_N^{kn}\} + i\,\text{Im}\{W_N^{kn}\}), \\ \qquad\qquad k = 0, 1, 2, 3, \ldots, N-1 \\ \sum_{n=0}^{N-1} (\text{Re}\{x[n]\}\text{Re}\{W_N^{kn}\} - \text{Im}\{x[n]\}\text{Im}\{W_N^{kn}\} \\ \qquad + i\,(\text{Im}\{x[n]\}\text{Re}\{W_N^{kn}\} + \text{Re}\{x[n]\}\text{Im}\{W_N^{kn}\})) \end{cases} \tag{8.76}$$

It is to be noted here that both $x[n]$ and $X[k]$ are periodic with period N. The 0–$N-1$ values of the sequence in the time domain are called points and in the frequency domain they are known as lines. Inspection of the equation reveals that even for the computation of one line in the frequency domain, a very large number of complex additions and multiplications are involved; the larger the value of N, the larger are the two.

Cooley and Tukey in 1965 proposed an algorithm using the symmetry and periodicity properties of W_N^{kn}. This algorithm called the FFT led to a considerable reduction in computational effort in doing a DFT.

The complex conjugate symmetry of W_N^{kn} gives

$$W_N^{k(N-n)} = W_N^{-kn} = (W_N^{kn})^* \tag{8.77a}$$

The property of periodicity in n and k results in the equation

$$W_N^{kn} = W_N^{k(n+N)} = W_N^{n(k+N)} \tag{8.77b}$$

Using the above properties and what is called a decimation-in-time procedure, the number of complex multiplications and additions required for an N point DFT computation gets reduced to $N \log_2 N$ as compared to N^2 multiplications and additions by brute force. This has made real-time analysis highly feasible.

It should be borne in mind that the original time series consisting of N points, when Fourier transformed, generates $N/2$ unique frequency coefficients only, from 0 to $(N/2) - 1$. Curiously, one started with N points; now there are only $N/2$ points. Since each Fourier component is described by the equation

$$X[k] = A_k + i\,B_k \tag{8.78}$$

the total number of descriptive frequency components is N and thus no information has been lost, only transformed into a new domain.

8.8 FFT Analyser Setup

INTERESTING FACTS: Two very common ways of analysing a signal are as a function of time, called time domain analysis, or as a function of frequency, called frequency domain analysis. It is good to know the general concepts associated with time and frequency, the relations between them and the constraints resulting from their combined use. Spectral representations of a signal, though attractive, may be insufficient from a physical standpoint, since time cannot be done away with while describing a signal. A very basic problem in signal processing is handling signals that are compact both in time and frequency. Spectrum analysis using classical Fourier analysis assumes that signals are infinitely long in time or are periodic. But in practice, many signals are of short duration, and many signals also change to a large extent over their duration. Such signals are hence poorly represented by Fourier methods. German physicist Werner Heisenberg proposed the Heisenberg uncertainty principle or indeterminacy principle which states that the position and the velocity of an object cannot both be measured exactly, at the same time. In the context of signals, Heisenberg's principle tells us that any signal cannot be compact or localized both in time and frequency and this gives rise to an "uncertainty" lower bound. Besides, a signal cannot confine its entire energy to finite intervals in the time and frequency domains, even if these intervals are large. The time–frequency localization is then measured in the mean squares sense. Hence, time–frequency analysis which studies a signal in both the time and frequency domains simultaneously, is required. This is better suited for non-stationary signals. In this context, the instantaneous frequency evolving from temporal variations becomes meaningful.

8.8.1 Setup for Spectrum Collection

The spectrum option available in FFT analysers is such as to cater to a relatively wide range of frequencies; hence, the analyser is forced to acquire a short time record. While deciding on the analyser settings, it has to be understood that the time record length depends only on the line spacing of the FFT spectrum (Example 1) and the sampling rate depends only on the frequency range of the FFT spectrum (Example 2), and they are independently adjustable. The sampling frequency of the time record for many analysers is 2.56 times the highest frequency in the spectrum. Therefore, a frequency span of 1000 Hz requires a sampling rate of 2560 samples per second. As the size of the transform N is generally fixed for a given analyser, the frequency

range covered and the resolution are determined solely by the time length of each sample. Let us consider two examples.

Example 1: A 400-line spectrum extending from DC to 1000 Hz will have a line spacing of 1000/400, or 2.5 Hz. The length of the time record used to calculate this spectrum is 1/2.5, or 0.4 s. The frequency resolution or the line spacing of a spectrum is therefore the reciprocal of the total time duration of the time domain record.

Example 2: In many analysers, the size of the transform N is a power of 2 such as 512, 1024, etc. Let us consider the sampling of a continuous-time function with a sampling frequency of 1 kHz or with $T_s = 0.001$ s, and number of data points in the analyser $= 1024$. The total duration of the time record is $T = NT_s = (1024)(0.001 \text{ s}) = 1.024$ s. The spectral resolution is equal to the base frequency of the approximate Fourier series $1/T = 1/1.024 = 0.9766$ Hz.

8.8.2 Setup for Analysing Time Waveforms

While setting up an analyser to store time records or waveforms, an important point to be considered is that the frequency range normally convenient for viewing a spectrum is usually not suitable for looking at the waveform. Most FFT analysers, with the exception of PC-based analysers, do not allow one to set up specific sampling rates or time domain record lengths; the setup is typically in terms of the frequency span and the frequency resolution.

Considering Example 1 above, the time record will show details that happen in a 0.4 s time span, but in practice, when looking at a machine vibration waveform, we often have to look at events occurring over a much longer duration. For instance, if we have to look at beat phenomena in the vibration signature of a machine, or of the combined vibration of two machines running at slightly different speeds, the duration of the waveform required would be at least several seconds long and hence it has to be ensured that there is adequate memory available in the data collector to store the waveform. It is best to use the lowest sampling rate and the shortest time record length that will provide the needed data. To acquire a waveform lasting 5 s, a setup with a line spacing of 1/5 Hz is required, and this can be done by adjusting the number of lines of resolution and the frequency span to convenient values.

If we want to observe beats that occur only once in several seconds in a waveform, the sampling rate needed will not be very high, something like 50 samples per second probably being sufficient. This corresponds to a frequency range of 50/2.56 (assuming a sampling frequency 2.56 times the highest frequency), or 19.53 Hz; thus, the 20 Hz range in the frequency analyser could be chosen.

On the other hand, if it is required to capture a waveform that might have glitches at 50 times per second, sampling has to be done fast enough to resolve each glitch. Sampling could be done at 1000 samples per second, necessitating a frequency span of 1000/2.56, or about 390 Hz. Suppose, as another example, a pinion is rotating at 1800 RPM (30 Hz) and it is required to capture one revolution (which will take

1/30th of a second) for purposes of time synchronous averaging. To set up an FFT analyser to collect a record of length 1/30th second, its frequency resolution, or line spacing, must be 30 Hz. In order to acquire a little more than one revolution, 1/25th second of time domain data might be acquired. There are multiple combinations of frequency range and resolution that could be used to do this: a span of 5000 Hz with 200 lines resolution or a span of 2500 Hz and 100 lines of resolution, or a span of 250 Hz at 10 lines resolution.

8.8.3 Leakage and Windowing

One of the main limitations in the computation of the DFT is that the data record using which computation is done is of finite length in the time domain. Besides, the assumption behind the computation of the DFT is that the signal is periodic in the finite length record. However, in actual practice, we perform FFT on signals that are not repetitive or periodic in the window considered, violating the basis behind the DFT computation. Only in artificially contrived situations is it possible to have the data processed periodic in the data window. Due to the violation of this assumption, there is leakage of energy out of one resolution line of the spectrum to other lines, leading to significant broadening of the 'theoretical' line spectra. Figure 8.28a shows a sinusoidal wave that is exactly periodic in the window and one that is not periodic in the time record. The corresponding spectrum takes the form of a line for the former as seen in Fig. 8.28b, while for the latter, it is seen that the discontinuities near the ends of the window result in a high-frequency content which is not present in the original signal, leading to smearing or spreading which can clearly be seen in the spectrum. This smearing of energy through the frequency domain is known as leakage.

Let us recollect that leakage is due to the fact that a finite time record with a non-integral number of cycles has been taken for computation of the FFT. A sine wave will have a single-line spectrum only if it exists for all time, from minus infinity to plus infinity, or if the time record considered for computation of the spectrum has an integral number of cycles in the window so that it would exactly match the actual input waveform. However, if we are measuring a continuous signal like a sine wave from a finite time record, leakage is caused if the input is not periodic in this record. The problem of leakage is severe enough to entirely mask small signals close to the sine wave under consideration. Obviously, leakage is not a problem with transient signals that die down before the end of the time window as shown in Fig. 8.29 and which are therefore known as self-windowing functions. Any function like a transient which does not require a window as it occurs completely within the time record is called a self-windowing function. There are many examples of self-windowing functions, such as impacts, shock signals, damped vibration response, sine bursts, noise bursts, chirp bursts, and pseudo-random noise. In vibration measurement, signals which are much more complex than a sinusoid of a single frequency occur and the effect of
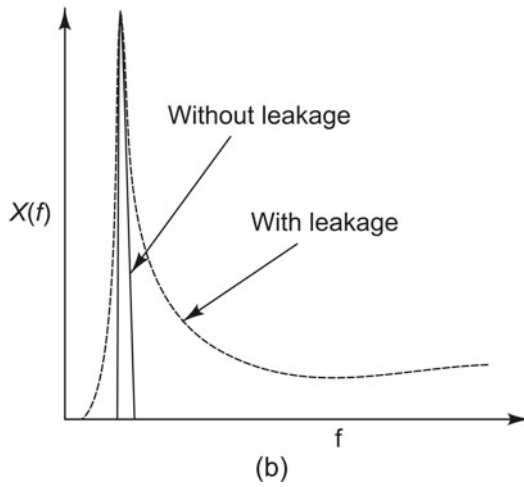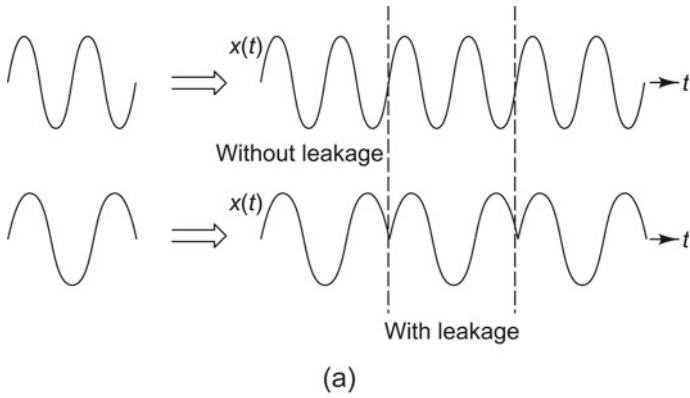
Fig. 8.28 Phenomenon of leakage: **a** time domain data, **b** spectra
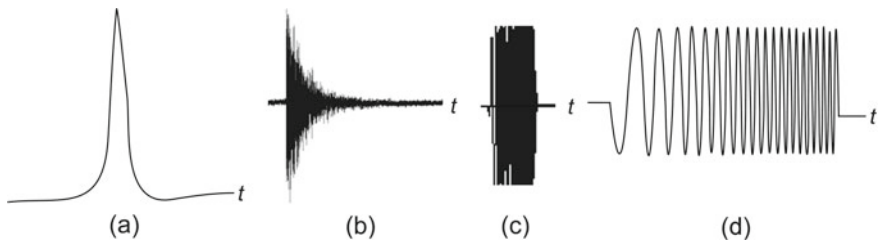
Fig. 8.29 Self-windowing signals: **a** impact, **b** shock response, **c** noise burst, **d** chirp sine

leakage can be quite severe. One would be tempted to conclude that the FFT is not a very useful spectrum analysis technique due to the associated problem of leakage. However, there is a solution to this problem and this is known as windowing.

Fig. 8.30 Windowed signal
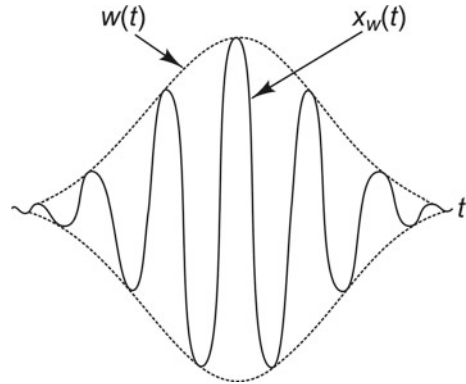
Windowing involves the imposition of a prescribed profile on the time record before computing the Fourier transform and this is done by multiplying the time record by a window function of known shape and which is typically zero at the ends of the record. If the time record, the FFT of which is to be obtained, is denoted as $x(t)$ and the profile of the window function is depicted as $w(t)$, then the windowed or analysed signal is $x_w(t) = x(t)w(t)$. Due to windowing, the signal is forced to be periodic in the analysis window as shown in Fig. 8.30. The envelope depicts the window function $w(t)$ and $x_w(t)$ is the windowed function.

Though the signal has been forced to be periodic as a result of windowing, the amplitudes, especially near the ends of the record, have been modified due to the modulation or multiplication process involved and the resultant FFT will yield improper spectral amplitudes. Hence, the FFT analyser has to rescale the FFT output to correct for the smaller spectral amplitudes obtained. Analysers are generally provided with a number of time domain windows such as Hanning, Hamming, tapered cosine, rectangular, flat-top, force-response windows, etc. One should select the window best suited to his/her application. Figure 8.31 a-f shows the different types of windows available. The Hanning or Cosine taper windows are typically used for steady periodic or random signals, while the exponential window is used for transient vibration records which contain quite a bit of information concentrated in the initial part of the time record and would thus lose information by a choice of other windows.

8.8.3.1 Hanning Window

There are a large number of functions available to window data, but the most commonly used one is called the Hanning window (Fig. 8.31a). Generally, the Hanning window does an exceptionally good job on random signals, and a reasonably good job on sinusoidal signals, irrespective of whether they are periodic or non-periodic in the time record. The problem with the Hanning window is that spectral amplitudes get changed. The window makes the transient look more like a sine wave than a
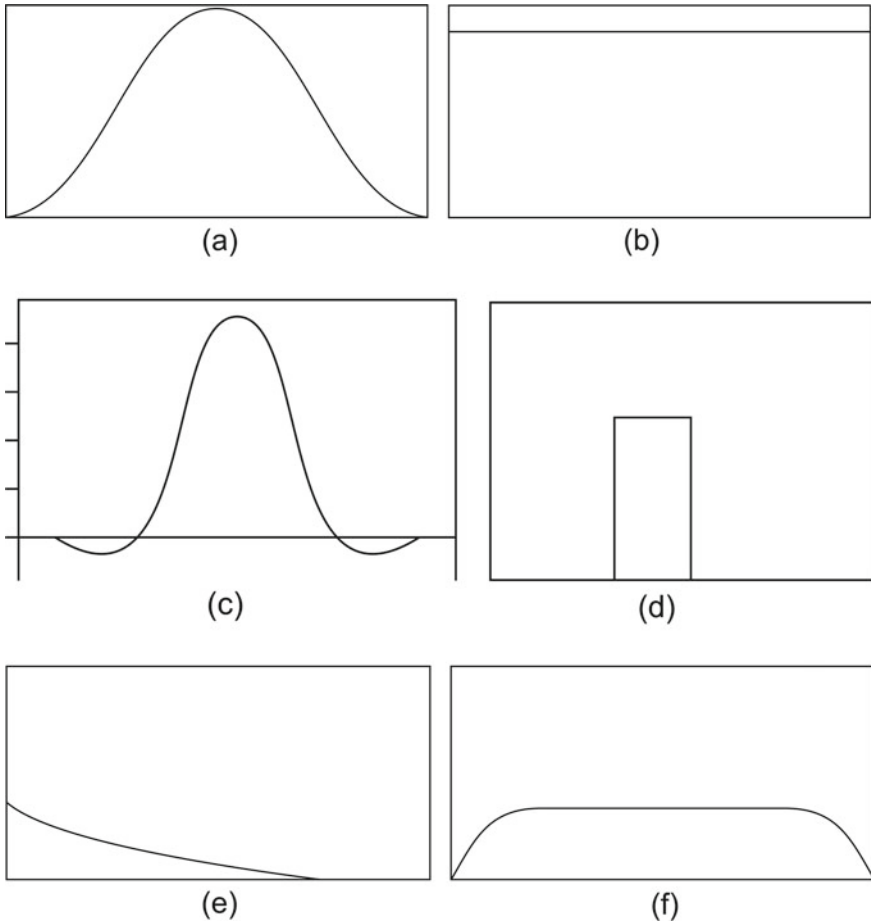
Fig. 8.31 Different types of windows: **a** Hanning, **b** rectangular, **c** flat top, **d** force, **e** response (exponential), **f** tapered cosine

transient. Therefore, for transients, the Hanning window is not a good choice. For such signals, it is preferable to give equal weightage to all the data in the time record. Hence, it is preferable to use the uniform or rectangular window which weights the entire time record uniformly.

8.8.3.2 Rectangular/Uniform Window

The problem in using a Hanning window with a transient signal has just been discussed. Hence, for transients, we might prefer to use a uniform or rectangular window which does not attenuate the signal at its ends. The rectangular window (Fig. 8.31b)

is typically used for self-windowing functions, the values of which are zero near the ends of the window. Self-windowing functions are often used as excitation signals in measuring the FRFs of systems, especially if the system is lightly damped with a sharp peak at resonance, the reason being that such functions generate no leakage in the FFT. Without a self-windowing excitation, energy could leak from a lightly damped resonance into adjacent lines in the spectrum, with the result that the FRF obtained would show greater damping than actual. Typically, a hammer with a built-in force transducer is used to excite a structure for the measurement of FRFs. The time record of the exciting force should just correspond to the transient impact imparted to the structure and should have the form of a half-sine wave, which is self-windowing and dies out within the time window of analysis. However, in practice, it is seen that the movement of the hammer before and after the impact causes stray noise signals in the time record and one way of eliminating this noise is through the use of a rectangular force window shown in Fig. 8.31d. This window is unity where the impact data is valid and zero everywhere else so that the analyser does not measure any stray noise.

8.8.3.3 Flat-top Window

We have seen that the Hanning function gives the filter a very rounded top. In the signal being analysed, if most of the signal is in the middle of the time record, then it will be measured accurately using the Hanning window; if not, the modulating effect of the window will attenuate the signal by up to 1.5 dB (16%). This error is unacceptably large if we are trying to measure the amplitude of a signal accurately. This deficiency is overcome by choosing a window function which has a flatter pass band (in the frequency domain). The flat-top window has exactly this nature and helps avoid the distortion of spectral amplitudes caused by the Hanning window. The error in amplitude resulting from this window does not exceed 0.1 dB (1%), a reasonably good improvement over the Hanning window. The Hanning and flat-top window functions enable the analyst to choose between increased accuracy or improved frequency resolution. Figure 8.31c shows the flat-top window.

8.8.3.4 Force Window

A force window (Fig. 8.31d) is essentially a rectangular window of adjustable width which is used to window the impact force signal in experimental modal analysis. This window does not alter the actual force pulse, but helps minimize the noise in the rest of the window and thus improves the SNR. It also removes the pulse due to a double hit. Data capture may be effected by appropriate triggering.

8.8.3.5   Exponential/Response Window

The exponential window (Fig. 8.31e) is especially suited for processing free vibration responses. During impact excitation, the output of the force transducer is connected to one channel of the signal analyser and the output of the response transducer to the second channel. It is known that the response of a structure is zero at the beginning of the time window before the impact excitation is imparted. Hence, the response window that is applied need not force the function there to zero. Besides, most of the structural response is obtained at the beginning of the time record; hence, we have to ensure that the signal is heavily weighted by the response window function here. However, the response has to be forced to zero near the end of the time record and an exponentially weighted window called the response window is used. Generally, three or four decay rates are available and may be chosen depending on the inherent damping in the system. Unlike the Hanning window, the response window is not zero at both ends of the time record. A response window imposed on the response of a lightly damped structure which has not fully decayed by the end of the time record gives an impression that the system has more damping than it really does. The effect of this additional damping due to the exponential window should be corrected to get the true estimate of modal damping.

8.8.3.6   Tapered Cosine Window

The cosine tapered (Tukey) window (Fig. 8.31f) sets the data at the ends of the window smoothly to zero without significantly reducing the gain of the transform of the windowed signal.

8.8.4   Averaging

A very important feature of digital spectral analysis concerning the specific requirements for processing random signals is averaging. A pure random signal never repeats in the measurement window and therefore excites the shaker differently during each measurement period. When analysing random vibration signals, the Fourier transforms cannot be computed, since they do not exist for a random signal and we must instead obtain the power spectral densities and correlation functions which are used to characterize such signals. In order to reduce the statistical variance of a measured power spectral density with a random excitation function such as pure random excitation and also to reduce the effects of non-linearities, it is necessary to employ an averaging process. The distortion, which shows up as random noise in each power

spectrum, can then be averaged out of the measurement, just like any other type of extraneous noise. When a pseudo-random excitation is used, this distortion noise cannot be averaged out since all the power spectra are exactly alike, i.e., they contain the same distortion components at the same frequencies. The choice of the number of averages required is decided by the statistical reliability required and the degree to which spurious random noise is to be removed from the signals. For valid random signal processing, the following condition should be satisfied.

$$2BT \geq K \tag{8.79}$$

where B is the frequency bandwidth, T the total time encompassing the data, and K a constant. If K is at least 10, then there is an 80% probability that the estimated spectrum lies between 0.5 and 1.5 times the true value. If K approaches 100, then reasonably reliable estimates are got, in the sense that there is an 80% probability that the measured value is within 18% of the true value. The larger the averaging time, the smaller are the variations in the mean and the better the spectral estimates, but the longer is the processing time.

There are two types of averaging, sequential and overlap averaging. In the former, m samples, each of duration T, and mutually exclusive are averaged as shown in Fig. 8.32a. However, since modern analysers compute the DFT in an extremely short duration of time, a new transform could be computed faster than a new complete sample record has been collected. Under such conditions, it is convenient to perform a second FFT computation as soon as possible, using the most recent N data points, even though some of them may have been used for computation of the previous transform. This overlap averaging process is depicted in Fig. 8.32b. It is evident that 100 averages performed using overlap averaging will not have the same statistical properties as would 100 completely independent samples.

8.8.5 Zoom

So far, we have discussed the basic DFT; this however is often found to have limitations of inadequate frequency resolution, especially at the lower end of the frequency spectrum and especially, for lightly damped systems. This difficulty arises on account of fewer discrete points available (N) for FFT computation, the maximum frequency range to be covered and/or the length of time record available. One solution to this problem would be to use a larger N in the computation of the transform. Though this gives finer frequency resolution around the regions of interest, it has the drawback that it is forced to compute more information than is required. One solution to this problem is to use the zoom option available in spectrum analysers. This allows the user to utilize all the spectral lines available (1024 or 2048) in a narrow frequency

Fig. 8.32 Averaging
process: **a** sequential
averaging, **b** overlap
averaging
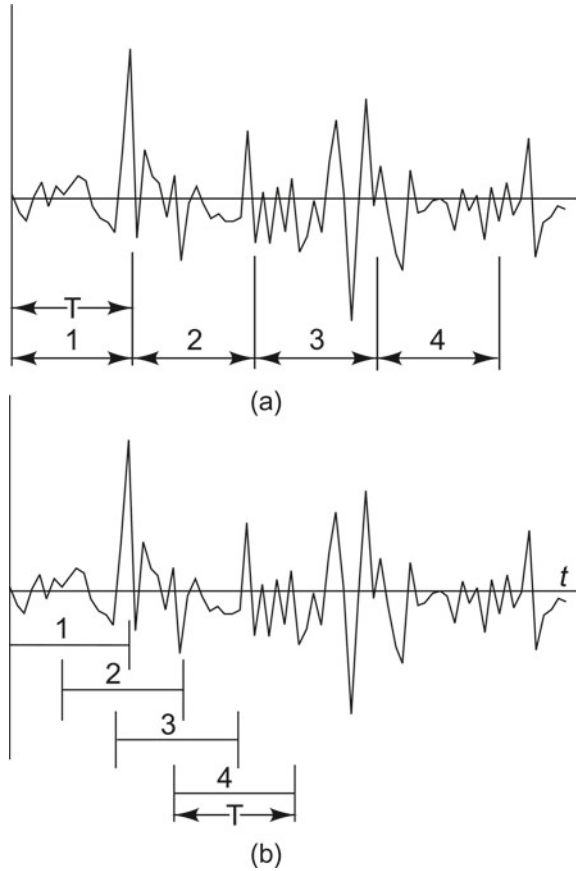
range between $f_{\min}$ and $f_{\max}$ (instead of between 0 and $f_{\max}$ without zooming).
Thus a zoom $\times 2$ will give a resolution double that obtained without zooming and a
zoom $\times 16$, 16 times the resolution without zooming. Under such circumstances, it
is advisable to have both the response and excitation signals subjected to the same
amount of zooming to compute FRFs. When using the zoom feature to measure
the FRF in a narrow frequency range, it is to be ensured that there is practically no
vibration energy outside the required frequency range. Hence, the excitation energy
supplied to the structure should be band limited to the analysis range. This feature
is not available on all analysers.

8.9 Dealing with Random Signals

INTERESTING FACTS: When we talk of random signals, the phrase that often come to the mind is "Gaussian probability distribution". This has been named after Johann Friedrich Carl Gauss, German mathematician, considered as one of the most influential and greatest mathematicians of all time. Even when he was seven, his mathematics teachers noticed his immense potential. He combined scientific theory and practice like no other mathematician before or after him. He has made immense contributions to the fields of number theory, differential geometry, non-Euclidean geometry, probability theory, geodesy, astronomy, the theory of functions, and electromagnetism, having published proliferously in these fields. Some of his major contributions are construction of a regular 17-gon by ruler and compass, a major advance in this field, dissertation on the fundamental theorem of algebra, prediction of the position of asteroid Ceres by developing basics of pattern recognition, studies on the motion of celestial bodies, studies on non-Euclidean geometry, discussion of probability distribution and statistical estimators, development of the method of least squares, work on differential geometry and Gaussian curvature, surveying, the first electric telegraph along with Wilhelm Weber, as well as the theory of electromagnetism. The international unit of magnetic induction, the Gauss, is named after him, as also the Gaussian distribution and Gaussian least squares method.

8.9.1 Introduction

For deterministic processes, there are mathematical models establishing a functional relationship between different physical variables, such that for specific values of the relevant variables of the physical phenomenon, a particular state of the process is obtained. In the case of a random process, however, the different states cannot be predicted exactly, but can only be predicted with certain measures of likelihood. A deterministic signal is one which has a well-defined waveform and which can be exactly resynthesized from frequency spectra, i.e., the magnitudes and relative phases of the synthesising sinusoidal waves. The value of a random signal or waveform, however, is not specified at every instant of time and it is not possible to predict its future values with certainty on the basis of the past values. Moreover, in some cases, the physical laws governing a phenomenon may be so complex that a detailed analysis becomes intractable, or there may be some inherent uncertainties associated with the phenomenon. A probabilistic model is therefore used to incorporate these uncertainties in a rational and systematic manner. In fact, it can be said that a proba-

bilistic model is generally more realistic as compared to a deterministic model. While a deterministic model can be used to predict the definite state of a phenomenon, the probabilistic model helps investigate the inherent unavoidable and unexplained variations seen in almost all physical phenomena. It quantifies these variations and may also provide clues to the causes of the variations. This model may be validated by statistical studies on available observations of the random phenomenon.

8.9.2 Modelling of Random Processes

By definition, a random process is a family of random variables. The random processes occurring in most engineering applications are either functions of time or space or both. The theory of random processes has been applied to many problems in physics and engineering involving uncertainty like kinetic theory of gases, statistical mechanics, Brownian motion, statistical communication theory, turbulent fluid flow, and earthquake engineering. Some of the typical random variables that we come across are noise of a jet engine, height of waves in a choppy sea, road undulations, vibration at the driver's seat of a vehicle, pressure gusts encountered by an aeroplane in flight, ground motion during an earthquake, electrical output of a noise generator, acoustic pressures generated by air rushing through a pipe, micro-roughness of a machined part, etc.

There are essentially three approaches to stochastic modelling of random processes. The first approach is by modelling them as probabilistic processes defined by Gaussian or Poisson or any of the other numerous distributions available and estimating the parameters of the distributions from the actual records available. In the second method, the sample functions of the random process are modelled as the outputs of appropriately designed filters to white noise excitation. The filter weights, which are the parameters of the model, are estimated by making use of the actual observations of the random phenomena. The third approach is non-parametric in nature, and here the process is described by higher order statistics. In some cases, the stochastic model is simply assumed to be a mathematically convenient process like the Gaussian process, with little regard to the information available about the actual process.

Random signals may be expressed as a function of any variable. Typically for vibration or acoustic applications, they are expressed as a function of time t or space s, e.g., road roughness may be expressed as $x(t)$ in temporal form, or $x(s)$ in the spatial domain. The output of a linear time-invariant system to a random input is also random. Thus, for the road roughness described above, vehicle response may be expressed as $y(t)$ or $y(s)$. Figure 8.33 shows a typical 'sample record' or 'sample function' $x(t)$, which is nothing but a single time history representing a random phenomenon (observed over a finite time interval).

Fig. 8.33 Sample function
of a random process
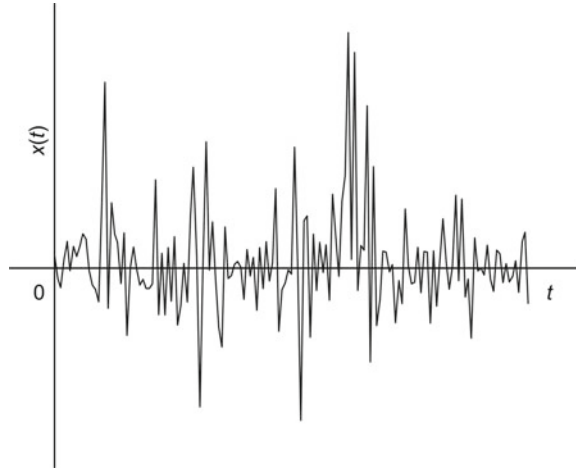
8.9.3 Probabilistic Model

For random signals, it is not possible to predict what the value of a signal is at a
particular instant of time t_1, but these signals can very conveniently be described in
probabilistic terms. There are various ways to characterize a probabilistic model. The
most common is the probability density function (PDF), in which the probability of
occurrence of a particular value from among the range of probable values is computed.
Equivalent ways are the cumulative distribution function (CDF), the moments, the
characteristic function, the moment-generating function, etc.

8.9.3.1 Probability Density Function (PDF) and Cumulative Distribution Function (CDF)

Random quantities can be described in a probabilistic sense. The probability of an
event which is absolutely certain to occur may be taken as unity, while that of an
event which is not at all likely to occur is taken as zero, with the probability of any
other event lying in between the two. We all know that in the tossing of a coin, the
probability that the outcome will be a head is the same as the probability that the
outcome will be a tail and this is 0.5. Mathematically, this may be written as

$$\text{Prob}[H] = \text{Prob}[T] = 0.5 \tag{8.80}$$

This has been arrived at intuitively, based on the knowledge of the geometric sym-
metry of the coin and the resulting outcomes. We also know that the total probability,
i.e., $\text{Prob}[H] + \text{Prob}[T] = 1$. This argument can be extended to the tossing of a sym-
metrical six-sided unloaded die. We can guess that the probability that the outcome

of a toss would be 1, 2, 3, 4, 5, or 6 would all be the same, i.e., 1/6, again based on symmetry. Let the outcome of a particular toss be denoted by N. This outcome will be known only after the trial. Let n denote all possible previously known outcomes, i.e., 1, 2, 3, 4, 5, or 6. Now we can write

$$p(5) = \text{Prob}[N = 5] = 1/6 \qquad (8.81)$$

and the total probability is

$$P(n) = \sum_{r=1}^{n} p(r) = 1 \qquad (8.82)$$

The quantity $p(n)$ is known as probability density, while the quantity $P(n)$ is known as probability distribution or cumulative distribution.

Let us extend this argument to the case where the random variable takes on a large number of discrete values. Assuming that in this case there is no symmetry possible, a very large number of trials have to be conducted and the probability of the occurrence of an event has to be ascertained from the same. If a trial is repeated M times and if an event A occurs m times, the probability of occurrence of event A is defined as

$$p(A) = \lim_{M \to \infty} \left(\frac{m}{M} \right) \qquad (8.83)$$

In general, a discrete random variable can take on values $x_1, x_2, x_3, \ldots$, etc. Then the probability that it takes on a value x_i is given by

$$\text{Prob}(x_i) = p_i \geq 0 \qquad (8.84)$$

and total probability is

$$P(x) = \sum_{i} p_i = 1 \qquad (8.85)$$

For a continuous random variable X, probability density $p(x)$ is defined as the probability that $X = x$ and is a non-negative function, i.e., $p(x) \geq 0$ for all x. Here, x is any previously specified or known possible value of the function and X is the unknown result of a trial, which will be known only after the trial.

If a large number of samples of a continuous random variable are obtained to produce a histogram, or frequency diagram, depicting relative frequencies of occurrences of various values of the random signal, then this histogram will correspond to the PDF of the random variable, provided that a very large number of possible values are considered (i.e., they are sampled with very fine resolution Δx). A PDF is thus nothing but a smoothened out version of a histogram and is a convenient probabilistic quantity when there are only a finite number of possible occurrences, since this will give rise to probabilities which are small fractional values of the total number of occurrences.

However, when there is an infinite number of occurrences possible, as in the case of a continuous variable $x(t)$, the probability of occurrence of one value out of the infinite possible values would tend to zero. Hence, it would be convenient to use probability distribution or cumulative distribution instead of probability density in the case of continuous random variables, in contrast to the description of discrete random variables. For example, the probability that the height of an individual is exactly 165.249735 cm is practically zero, but the probability that the height is within the interval 165.1 and 165.3 cm is a measurable quantity. In such cases, it would be more realistic to find out the (total) probability $P(x)$ that the height of an individual would be less than or equal to a certain specified value, since this is a finite quantity, and can be obtained based on a sufficiently large number of trials. The probability that the value of x is less than or equal to some specified value X is the probability distribution or cumulative distribution and is

$$P(x) = \text{Prob}[X \leq x] = \int_{-\infty}^{x} p(x)\mathrm{d}x \qquad (8.86)$$

For the discrete variable case, the probability distribution function is given by

$$P(x) = \text{Prob}(X \leq x) = \sum_{x_i \leq x} p_i \qquad (8.87)$$

The total area under any probability density curve corresponds to the total probability of the occurrences of all possible values of a variable. This implies that the total integral of $p(x)$ must be 1, i.e., the total area under the probability density graph is 1.

$$\int_{-\infty}^{+\infty} p(x)\mathrm{d}x = 1 \qquad (8.88)$$

The probability density $p(x)$ of a random variable $x(t)$ can also be computed from the time domain and is simply equal to the fraction of the time the variable takes on values between x and $(x + \mathrm{d}x)$ to the total length of the time record T. The computation has to be done over a sufficiently long record of length T in the manner depicted in Fig. 8.34a. Alternately, $P(x_1)$ can be determined by finding out the total time that $x(t)$ is less than a value x_1, by summing all time intervals such as $\Delta t_1, \Delta t_2, \Delta t_3, \ldots,$ etc., over a sufficiently long record of length T as shown in Fig. 8.34b. This duration divided by the total time duration of the sample function represents the probability that $x(t)$ is less than x_1.

If $p(x)$ is known, then the (total) probability $P(x)$ that the value of the variable $x(t)$ is within an interval $[x_1, x_2]$ is described by the integral of the function $p(x)$ over the integration interval of the variable.
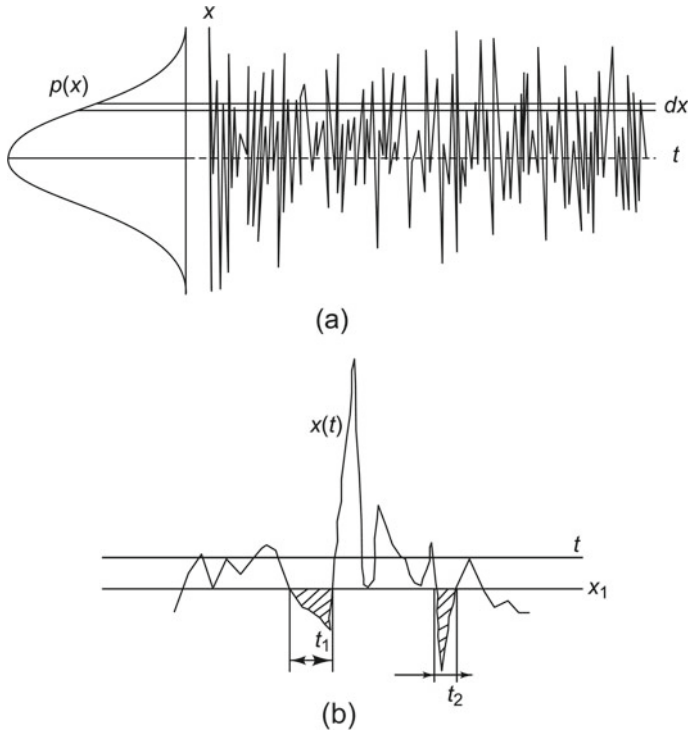
Fig. 8.34 Computation of probability functions: **a** density function, **b** distribution function

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} p(x)\mathrm{d}x \tag{8.89}$$

This total probability that variable X will lie between two specific values x_1 and x_2 can be expressed mathematically as the difference of probabilities as shown on the right-hand side of the equation below.

$$\mathrm{Prob}[x_1 \leq X \leq x_2] = \begin{cases} \mathrm{Prob}[X \leq x_2] - \mathrm{Prob}[X \leq x_1] \\ P(x_2) - P(x_1) \end{cases} \tag{8.90}$$

Hence, the probability that X lies in the very small interval $(x, x + \mathrm{d}x)$ can be written as

$$\mathrm{d}P(x) = \begin{cases} \mathrm{Prob}[X \leq x + dx] - \mathrm{Prob}[X \leq x] \\ P(x + \mathrm{d}x) - P(x) \end{cases} \tag{8.91}$$

For very small dx, this probability will be proportional to dx and maybe denoted as $p(x)dx$.

Therefore,

$$dP(x) = p(x)dx \tag{8.92}$$

It is possible to say from intuition also that if a probability distribution has density $p(x)$, then in the infinitesimal interval $(x, x + dx)$, it has probability $p(x)dx$. In the limit $dP(x)$ and dx becoming infinitesimally small, probability density $p(x)$ can be obtained from probability distribution as shown below:

$$p(x) = \frac{dP(x)}{dx} \tag{8.93}$$

where $p(x)$ and $P(x)$ are the probability density and distribution, respectively, as stated earlier. Therefore,

$$P(x) = \int_{-\infty}^{x} p(z)dz \tag{8.94}$$

For practical continuous variables like the height or weight of a person, the lower limit of integration is set to 0.

Figure 8.35 shows typical probability density and distribution curves. The shapes of these curves give a qualitative indication of the nature of distribution. If a variable has values closely surrounding a mean value, the $p(x)$ curve would be a steep narrow curve, centred around the mean, while the $P(x)$ curve would show a steep rise in this area, increasing monotonically from 0 at $t = -\infty$ to 1 at $t = +\infty$.

The word 'confidence' is often used when describing a statistical distribution. When we say that the declared life of a bulb is 180 days based on a 96% confidence level, it means that 96% of a very large number of bulbs tested has a life greater than 180 days. Or in other words, any sample taken would have at least 0.96 probability of exceeding the declared life of 180 days. Thus, a confidence level is simply a measure of probability and can be obtained from the cumulative distribution curve.

Figures 8.36a–c show the time records, probability density, and distribution curves, respectively. The first row corresponds to a narrow-band random signal, the second row to a broadband random signal, and the third row to a sinusoidal signal.

8.9.3.2 Probability of Joint Events

When considering probability of joint events, we define what are called independent events and dependent events.

Independent events or variables Two discrete events, A and B are said to be independent or uncorrelated if occurrence of event A does not affect the occurrence of event B, and vice versa. The outcomes of two tosses of a die constitute a typical example of two independent events. The joint PDF, i.e., the probability of both events
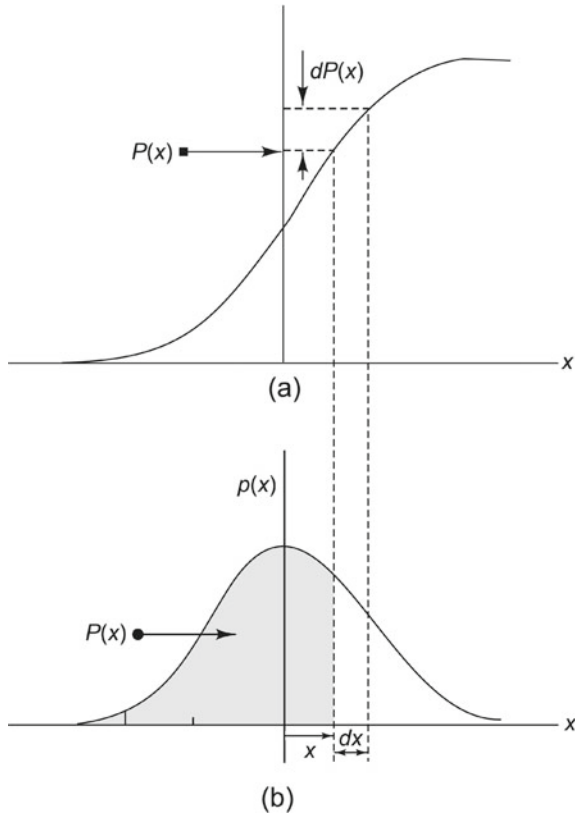
Fig. 8.35 Probability functions: **a** probability distribution, **b** probability density

occurring is then

$$p(a, b) = p(a)p(b) \tag{8.95a}$$

When dealing with two continuous random variables, X and Y, the joint PDF is given by

$$p(x, y) = p(x)p(y) \tag{8.95b}$$

The joint CDF for this pair of random variables X and Y is given by

$$P(x, y) = \begin{cases} \text{Prob}(X \leq x, Y \leq y) \\ P(x)P(y) \end{cases} \tag{8.96}$$

where $P(x)$ and $P(y)$ are the marginal distributions of X and Y. The right-hand side of the above equation represents the probability that the random variable X takes on a value less than or equal to x and that Y takes on a value less than or equal to y. This
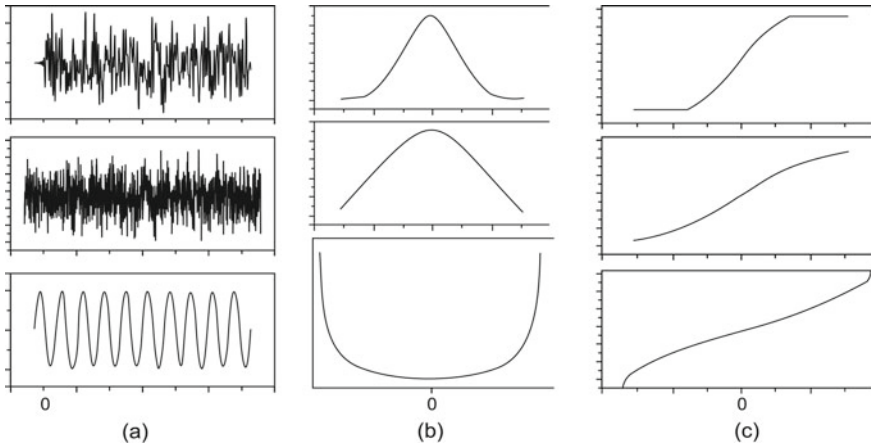
Fig. 8.36 Probability for three typical signals: **a** $x(t)$ versus t, **b** $p(x)$ versus x, **c** $P(x)$ versus x

can be extended to the case of N continuous random variables $X_1, X_2, X_3, \ldots, X_N$. They are all independent of each other or uncorrelated if and only if

$$p(x_1, x_2, x_3, \ldots, x_N) = p(x_1)p(x_2)p(x_3) \ldots p(x_N) \qquad (8.97)$$

If the joint PDF can thus be factored into a product of N functions of one variable each as shown above, the CDF can be defined as

$$P(x_1, x_2, x_3, \ldots, x_N) = P(x_1)P(x_2)P(x_3) \ldots P(x_N) \qquad (8.98)$$

Dependent events or variables Two variables are said to be dependent if they are related to each other, such as the load and deflection of a cantilever beam. Another example could be the humidity and rainfall at a particular location. Let us suppose that an experiment is conducted with two sets of possible outcomes. Suppose that out of N experiments performed, n produce outcome A; of these N experiments, let us suppose that m experiments also produce outcome B. The joint probability is the probability that the outcome A from one set occurs together with outcome B from the other set.

$$p(a, b) = \lim_{N \to \infty} \left(\frac{n}{N} \right) \left(\frac{m}{n} \right) = p(a)p(b/a) \qquad (8.99)$$

where $p(b/a)$ is the probability of B given that A has occurred.

For the case of two continuous random variables, the probability distribution and density functions are given by

$$P(x, y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} p(\xi, \eta) \mathrm{d}\xi \, \mathrm{d}\eta \tag{8.100a}$$

$$p(x, y) = \frac{\partial^2 P(x, y)}{\partial x \partial y} \tag{8.100b}$$

For the case of N variables,

$$P(x_1, x_2, x_3, \ldots) = \int\limits_{-\infty}^{x_1} \int\limits_{-\infty}^{x_2} \int\limits_{-\infty}^{x_3} p(\xi, \eta, \zeta, \ldots) \mathrm{d}\xi \, \mathrm{d}\eta \, \mathrm{d}\zeta \ldots \tag{8.101a}$$

$$p(x_1, x_2, x_3, \ldots) = \frac{\partial^N P(x_1, x_2, x_3, \ldots)}{\partial x_1 \partial x_2 \partial x_3 \ldots} \tag{8.101b}$$

Here, $p(x_i)$ is the PDF associated with the variable X_i alone and is called the marginal density function. It can be deduced from the joint PDF of the random variables $X_1, X_2, X_3, \ldots, X_N$ by integrating with respect to all the other $N - 1$ variables as shown below

$$p(x_i) = \int \cdots \int \int \cdots \int p(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_N) \mathrm{d}x_1 \ldots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \ldots \mathrm{d}x_N \tag{8.102}$$

8.9.4 Some Common Distributions

Though there are a very large number of distributions that are in use, such as rectangular or uniform, Gaussian, binomial, Poisson, etc., we shall discuss a few distributions that come in handy while dealing with random variables in the context of random vibration.

8.9.4.1 Rectangular Distribution

A rectangular distribution is shown in Fig. 8.37a. It is assumed that the probability density $p(x)$, of the variable, x, is constant between the limits a and b. The PDF for this distribution is given by

$$p(x) = \frac{1}{b - a} \tag{8.103}$$

8.9.4.2 Binomial Distribution

This distribution may be used to determine the probability of failure of a component during its life. It is used to find the probability of occurrence of a single event from

a very large number of trials when the event has only two possible results, say success and failure, which are mutually exclusive. Other assumptions made behind this distribution are that the results of repetitive trials are independent of each other and that the probabilities of occurrence are time-invariant. Let us consider a series of trials in which the probability of occurrence of a particular event is a and that of its non-occurrence is $1 - a = b$. The probability that the event will not occur in any one trial is $1 - a$. The probability that it is not at all present in n trials is $(1 - a)^n$. Hence, the probability that the event is not completely absent in n trials or in the other words that it will occur in n trials is

$$P = 1 - (1 - a)^n \tag{8.104}$$

The above equation gives the probability of occurrence in a number of trials n, given the probability of occurrence in one trial.

Equation (8.104) can be expanded using the binomial theorem since $a < 1$,

$$\text{i.e., } P = \left[1 - na + \frac{n(n-1)}{2} p^2 \ldots \right] \tag{8.105}$$

When n is very large

$$P \sim na \tag{8.106}$$

The mean value is na where n is the number of trials and a the probability of success. Let us consider a more general case of a sequence of n trials having k successes with probability a and $n - k$ failures with probability $b = 1 - a$. If we neglect the order in which these events occur, then the ways in which these events occur is given by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{8.107}$$

Hence, the probability of k successes in n trials is equal to

$$p(k \text{ successes, } n \text{ trials}) = \binom{n}{k} a^k b^{n-k} \tag{8.108}$$

This equation represents the binomial distribution with parameters a and n. A random variable X is said to have a binomial distribution if

$$p(X = k) = \binom{n}{k} a^k b^{n-k}, \quad k = 0, 1, \ldots \ldots n \tag{8.109}$$

The probability of success in all the other $n - k$ trials is 0.

The sum of the probabilities is given by the binomial theorem as

Fig. 8.37 Probability density function: **a** rectangular distribution, **b** binomial distribution

(a)

(b)
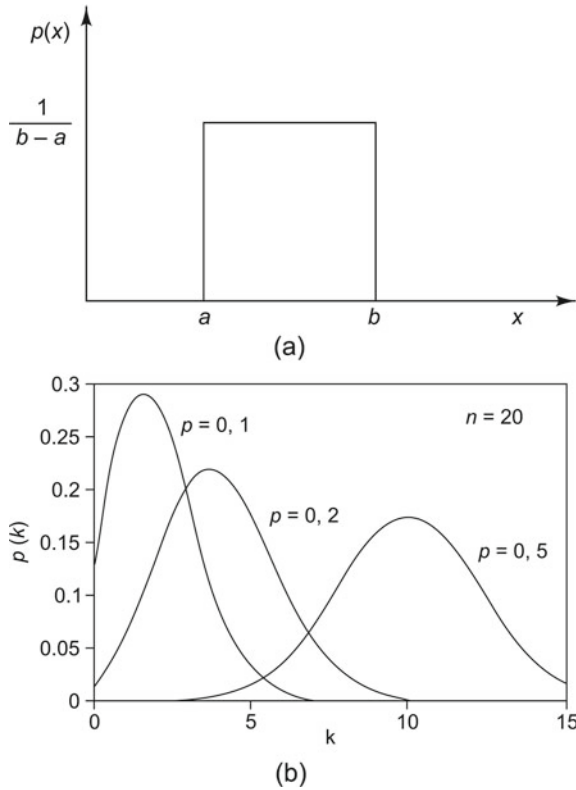
$$\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = (a + b)^n = 1 \qquad (8.110)$$

Figure 8.37b shows the binomial distribution.

8.9.4.3 Gaussian Distribution/Normal Distribution

The normal distribution, also called the Gaussian distribution, is a very important form of continuous probability distribution widely used in science and engineering, for the reason that many real-life signals can be idealized as Gaussian. It is also called the bell curve because its probability density is a bell shaped curve. Many measurements of physical phenomena, such as radio noise, jet engine pressure fluctuation, atmospheric turbulence, thermal noise, etc., can be approximated by the normal distribution. This distribution is the most extensively used family of distributions in statistics to find significance levels in many hypothesis tests and confidence intervals. In modelling applications, such as linear and non-linear regression, the error term is
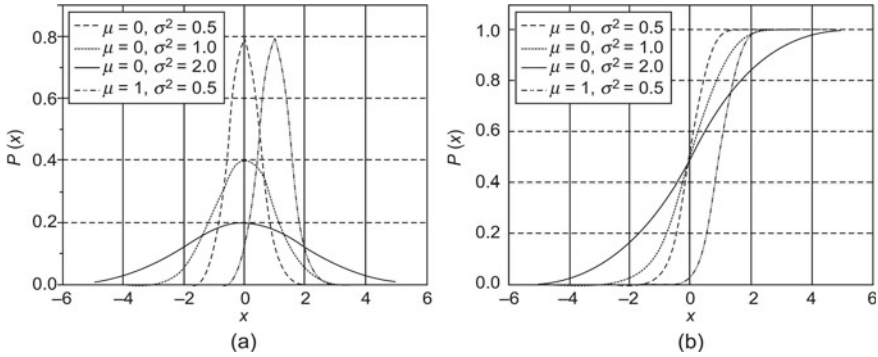
Fig. 8.38 Normal distribution for different μ and σ: **a** probability density function, **b** probability distribution function

often assumed to follow a normal distribution. The advantage of being able to model a random process as a Gaussian distribution is that the output of a linear time-invariant system to Gaussian random input is also Gaussian. Under certain conditions, such as for independent and identically distributed variables with finite variance, the sum of a large number of random variables is approximately normally distributed and this is the essence of the central limit theorem. The practical importance of this is that the normal CDF can be used as an approximation to some other CDFs. For example, a binomial distribution with parameters n and a is nearly normal for large n and a not too close to 1 or 0. Thus, for most distributions, if the number of events or trials is very large, then the Gaussian distribution function may be used to describe them.

The probably density function for a Gaussian distribution shown in Fig. 8.38 is given by

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty \qquad (8.111)$$

where μ and σ totally characterize the Gaussian signal, μ being the mean, and σ being the standard deviation. To indicate that a real-valued random variable X is normally distributed with mean μ and variance $\sigma^2 \geq 0$, we write $X \sim N(\mu, \sigma^2)$. The mean, or the expected value corresponds to the centroid of the PDF; it is also the point at which the PDF is maximum. The variance σ^2 is a measure of the dispersion of the random variable around the mean. Thus, this distribution is highly mathematically tractable, being characterized by only two parameters, the mean and the variance (square of standard deviation) and if these are known, the PDF is defined. Figure 8.38a and b show the probability density and distribution functions for various values of μ and σ.

If the mean value μ is zero, then

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x^2/2\sigma^2)} \qquad (8.112)$$

The formula for the CDF of the normal distribution does not exist in the form of a simple closed-form expression. It may be expressed in terms of the density function as follows:

$$\text{Prob}(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(y-\mu)^2/2\sigma^2} dy = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$= \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \tag{8.113}$$

where

$$\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_{0}^{z} \exp^{-y^2} dy \tag{8.114}$$

The Gaussian distribution can be normalized so that $\mu = 0$ and $\sigma = 1$ and this is called the standard normal distribution. This is done by a change of variables such that $z = [(x-\mu)/\sigma]$. It may be noted that the case corresponding to $\mu = 0$ and $\sigma = 1$ in Fig. 8.38 represents the standard normal distribution. The equation for the PDF of the standard normal distribution is given by

$$p(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), \quad -\infty < z < \infty \tag{8.115}$$

i.e., $Z \sim N(0,1)$. Conversely, if Z is a standard normal distribution, $Z \sim N(0,1)$, then X is a normal random variable with mean μ and variance σ^2 as shown below.

$$X = \sigma Z + \mu \tag{8.116}$$

Since the general form of the Gaussian probability function can be expressed in terms of the standard normal distribution, all subsequent formulae in this section are given for the standard form of the function. To evaluate the probability in Eq. (8.113), the error function erf(x), has to be evaluated. The standard normal distribution is tabulated (usually in the form of values of the CDF) and is available readily in most text books, and the other normal distributions are simple transformations of the standard normal distribution. Hence, one can use tabulated values of the CDF of the standard normal distribution to find values of the CDF of a general normal distribution. The cumulative distributions computed numerically for $x = 0$ and $\sigma = 1$ and given in the form of Gaussian tables give the cumulative probability that $z = z_1$, which is nothing but the area under the normal curve $p(z)$ up to z_1, computed as shown in the following equation and as depicted in Fig. 8.39. Table 8.5 gives $P(z)$ corresponding to a few often used values for the area under the bell curve. These values are useful in determining confidence intervals of the specified levels based on standard normal distribution. Figure 8.40 depicts a plot of the standard normal PDF.

Table 8.5 Gaussian table for computation of CDF

z	erf(z)	z	erf(z)	z	erf(z)	z	erf(z)
0.05	0.01994	0.80	0.28814	1.55	0.43943	2.30	0.48928
0.10	0.03983	0.85	0.30234	1.60	0.44520	2.35	0.49061
0.15	0.05962	0.90	0.31594	1.65	0.45053	2.40	0.49180
0.20	0.07926	0.95	0.32894	1.70	0.45543	2.45	0.49286
0.25	0.09871	1.00	0.34134	1.75	0.45994	2.50	0.49379
0.30	0.11791	1.05	0.35314	1.80	0.46407	2.55	0.49461
0.35	0.13683	1.10	0.36433	1.85	0.46784	2.60	0.49534
0.40	0.15542	1.15	0.37493	1.90	0.47128	2.65	0.49597
0.45	0.17365	1.20	0.38493	1.95	0.47441	2.70	0.49653
0.50	0.19146	1.25	0.39435	2.00	0.47726	2.75	0.49702
0.55	0.20884	1.30	0.40320	2.05	0.47982	2.80	0.49744
0.60	0.22575	1.35	0.41149	2.10	0.48214	2.85	0.49781
0.65	0.24215	1.40	0.41924	2.15	0.48422	2.90	0.49813
0.70	0.25804	1.45	0.42647	2.20	0.48610	2.95	0.49841
0.75	0.27337	1.50	0.43319	2.25	0.48778	3.00	0.49865

Fig. 8.39 Computation of CDF from standard normal distribution
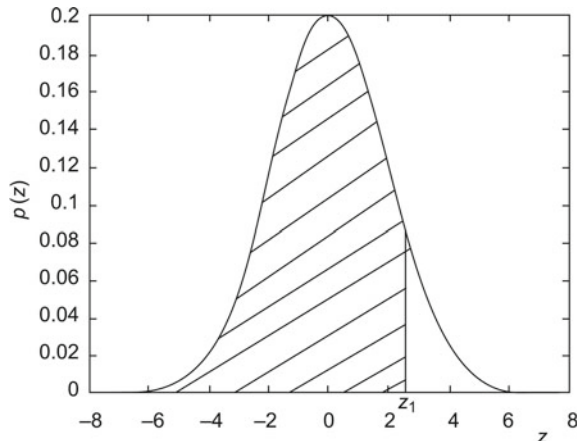
$$p(z = z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-(z^2/2)} \mathrm{d}z \qquad (8.117)$$

From this density curve, it is observed that values of z less than one standard deviation from the mean, account for about 68% of the total set of values, or in other words the probability that z is within one standard deviation of the mean is 0.683. Similarly, values of z within two standard deviations from the mean account for about 95%. Values of z that are less than three standard deviations from the mean account
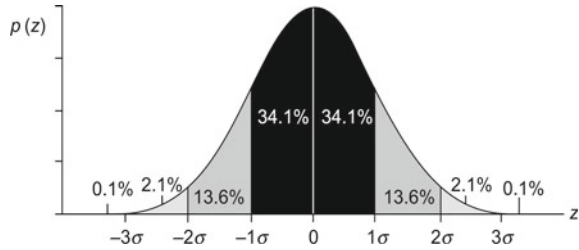
for about 99.7% of the entire population. This can be represented mathematically as

$$\text{Prob}\{|X - \mu| \le \sigma\} = 2\text{erf}(1) = 0.68268$$

$$\text{Prob}\{|X - \mu| \le 2\sigma\} = 2\text{erf}(2) = 0.95452 \tag{8.118}$$

$$\text{Prob}\{|X - \mu| \le 3\sigma\} = 2\text{erf}(3) = 0.9973$$

8.9.5 Statistical Descriptors for Random Signals

In any statistical method, a large amount of data is required to establish reliability. For example, to characterize road undulations over a certain stretch, a large number of records may have to be collected. The values typically used to characterize random signals in a statistical sense are mean, mean square value, skewness, kurtosis, and other higher moments and central moments.

8.9.5.1 Expected Value, Time Averaging, and Moments

Expectation or expected value is defined as the expected result in any trial, and is assumed to be the mean value of a very large number of trials. In the tossing of a symmetrical die, it can be assumed that all numbers from 1 to 6 will occur with equal frequency, provided that the number of throws is very large. Therefore, the expected value in this case is the mean of all the outcomes, i.e., 1, 2, 3, 4, 5, and 6.

$$E[N] = \left(\frac{1 + 2 + 3 + 4 + 5 + 6}{6}\right) = 3\frac{1}{2} \tag{8.119}$$

Here, E represents the expectation operator. It can also be written as

$$E[N] = \sum_{1}^{6} np(n) \tag{8.120}$$

Thus, for the case of discrete random variables, the expected value, which is the average value, can be generalized as

$$E[X] = \bar{x} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} x_i \tag{8.121a}$$

The concept of time averaging over a sufficiently long period of time T (ideally tending to infinity) may also be used and is often done so in the treatment of continuous-time random variables. The expected value of $x(t)$ or expectation or mean value denoted by $E[X] = \langle x(t) \rangle = \overline{x(t)}$ is defined as the average or mean value of the variable sampled over time T tending to infinity.

Here, $\langle \rangle$ denotes the expectation operator.

$$\text{Mean} = E[X] = \langle x(t) \rangle = \overline{x(t)} = \mu_x = \lim_{T \to \infty} \frac{1}{T} \int_0^T x(t) \mathrm{d}t \tag{8.121b}$$

The mean square value, designated by $\overline{x^2(t)}$ or expectation of $x^2(t)$, i.e., $E[X^2]$, may also be obtained by integrating $x^2(t)$ over a sufficiently long time T and taking its average value. This quantity is an indication of the energy present in the signal. The mean square value of a vibration signal is quite often used in machinery condition monitoring to give an indication of the severity of a fault. Mean square value can be applied to a single variable such as $x^2(t)$ or to a product of variables such as $x(t).y(t)$.

$$E[x^2(t)] = \overline{x^2(t)} = \lim_{T \to \infty} \frac{1}{T} \int_0^T x^2(t) \mathrm{d}t \tag{8.122}$$

Covariance is a quantitative measure of the degree to which two random variables $x(t)$ and $y(t)$ are correlated and is defined as

$$\sigma_{xy}^2 = \langle x(t).y(t) \rangle \tag{8.123}$$

If $x(t)$ and $y(t)$ are completely uncorrelated, σ_{xy}^2 will be zero; if they are not independent of each other, σ_{xy}^2 will be other than zero.

When one is interested in characterizing a signal in terms of its mean value and its fluctuation about the mean, a quantity called the variance σ^2, which is the mean squared value about the mean is used. This is a quantity of great importance when dealing with random signals and is obtained by doing a time averaging as shown below. σ is referred to as the standard deviation. σ^2 may be obtained from time integration and is given by the equation below.

$$\text{Variance} = \sigma^2 = \langle [x(t) - \langle x(t) \rangle]^2 \rangle = \lim_{T \to \infty} \frac{1}{T} \int_0^T (x - \bar{x})^2 \mathrm{d}t \tag{8.124}$$

This can be simplified to obtain an equation relating variance and mean as shown below.

$$\sigma^2 = \langle x^2(t) - 2x(t)\langle x(t)\rangle + \langle x(t)\rangle^2\rangle$$
$$= \langle x^2(t)\rangle - 2\langle x(t)\rangle^2 + \langle x(t)\rangle^2 \qquad (8.125)$$
$$= \langle x^2(t)\rangle - \langle x(t)\rangle^2$$
$$\text{i.e., } \sigma^2 = \overline{x^2} - (\overline{x})^2 = \overline{x^2} - \mu_x^2$$

From this equation, it is easily seen that variance σ^2 equals mean squared value minus square of the mean. For zero mean signals, the variance is the same as the mean square value and the square root of the variance, i.e., the standard deviation σ is the same as the root mean square value. This fact may come in handy when trying to validate experimentally obtained results. In practice, 'AC coupling mode' gives rise to a zero mean signal in an oscilloscope or signal analyser.

The mean and mean square values defined in terms of the time averages are related to the PDF as explained below. If a random variable X is given, and its distribution admits a PDF $p(x)$, then the first moment or expectation or expected value of X (if it exists) can be determined by computing the first moment of the area under the probability density curve about the y axis, that is the vertical axis passing through $x = 0$ (Fig. 8.41). First moment of the probability density curve about the $p(x)$ axis gives

$$E(X) = \overline{x} = \int_{-\infty}^{+\infty} xp(x)\mathrm{d}x \qquad (8.126)$$

The mean value $\overline{x}$ or the average value coincides with the centroid of the area under the probability density curve $p(x)$. One can recall from theory relating to properties of surfaces that the following equations give the location of the centroid denoted by $(\overline{x}, \overline{y})$. First moments of an area A about the y and x axes give

$$\overline{x} = \frac{1}{A} \int_0^A x\mathrm{d}A \qquad (8.127a)$$

$$\overline{y} = \frac{1}{A} \int_0^A y\mathrm{d}A \qquad (8.127b)$$

The area A for a probability density histogram is 1.

Similarly, the mean square value defined in terms of a time average is related to the PDF and may be determined from the second moment of the area under the probability density curve about the vertical axis $x = 0$.

Fig. 8.41 Computation of moments of area of PDH about $p(x)$-axis
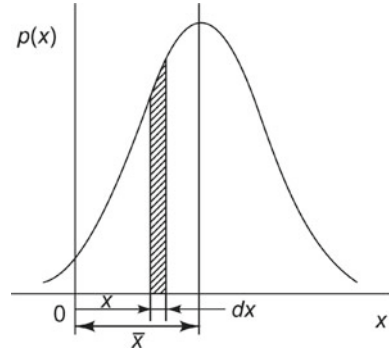
$$\text{Mean square value} = E(X^2) = \langle x^2(t) \rangle = \overline{x^2} = \int_{-\infty}^{+\infty} x^2 p(x)\mathrm{d}x \qquad (8.128)$$

This is analogous to the computation of the second moment of an area about $x = 0$ which gives the radius of gyration. The radii of gyration got from computation of the second moments of an area about the y and x axes are given in the following equations:

$$k_y^2 = \frac{1}{A} \int_0^A x^2 \mathrm{d}A \qquad (8.129a)$$

$$k_x^2 = \frac{1}{A} \int_0^A y^2 \mathrm{d}A \qquad (8.129b)$$

The variance σ^2, which is the mean squared value about the mean, is also known as the second central moment. It may be defined as the second moment of the area under the probability density curve about the vertical axis through $x = \mu_x$ or second moment about the mean.

$$E[(x - \mu_x)^2] = \sigma^2 = \overline{(x - \overline{x})^2} = \int_{-\infty}^{+\infty} (x - \overline{x})^2 p(x)\mathrm{d}x = \int_{-\infty}^{+\infty} (x - \mu_x)^2 p(x)\mathrm{d}x$$

$$(8.130)$$

In a similar fashion, higher order moments and central moments as well as cross-moments may be computed. In particular, the nth moment $E(X^n)$ of the probability distribution of a random variable X is given by

$$E(X^n) = \int_{-\infty}^{+\infty} x^n p(x)\mathrm{d}x \qquad (8.131)$$

Similarly, the higher order central moments may also be computed as

$$E[(X - \mu_x)^n] = \int_{-\infty}^{\infty} (x - \mu_x)^n p(x) \mathrm{d}x \qquad (8.132)$$

Among the higher order moments, the quantities of interest are skewness and kurtosis which correspond to orders 3 and 4, respectively. These quantities find use in machinery diagnostics.

8.9.5.2   Auto- and Cross-Correlations

Correlation is a measure of the similarity between two quantities. As applied to vibration or acoustic signals, correlation is a time domain analysis procedure useful for detecting hidden periodicities buried in a measured signal and for the measurement of the propagation time through a structure. Correlation gives information regarding a structure's spectral characteristics.

Autocorrelation The autocorrelation function is a quantity which in spite of being a time parameter, gives a lot of information about the spectral content of a signal. It is a measure of the general dependence of the value of a signal at some instant of time to values at other instants of time. Autocorrelation is the expected value of the product $x(t)x(t + \tau)$. It is a quantity that can be computed from a single record of sufficiently long duration T by multiplying the ordinates at time t and time $t + \tau$ (Fig. 8.42) and determining the average over all t.

$$R_{xx}(\tau) = \langle x(t)x(t + \tau) \rangle \lim_{T \to \infty} \frac{1}{T} \int_0^T [x(t)x(t + \tau)] \mathrm{d}t \qquad (8.133)$$

If averaging is done over a sufficiently long record, $R_{xx}(\tau)$ is independent of t and is a function of τ alone. When τ equals zero, the autocorrelation reduces to the mean squared value as shown below.

$$R_{xx}(0) = \sigma^2 = \overline{x^2} \qquad (8.134)$$

$R_{xx}(0)$ is the maximum value of the autocorrelation function. As all the products are squares, $R_{xx}(0)$ must be positive. When τ is very small, it can be expected that $x(t)$ and $x(t + \tau)$ are not much different from each other, and hence the number of positive products is larger than that of the negative products, resulting in a positive value of $R_{xx}(\tau)$, which will nevertheless be less than $R_{xx}(0)$. For increasing values of τ, the factors of each product become less and less related to each other and $R_{xx}(\tau)$ goes on oscillating between positive and negative values, before finally settling to zero. If fluctuations in $x(t)$ are very rapid, then $x(t)$ and $x(t + \tau)$ may have very little relationship. If on the other hand, the fluctuations are slow, then $x(t)$ and $x(t + \tau)$ are closely connected. Thus, the nature of the autocorrelation curve is dependent on the spectral content of the signal. When τ equals infinity, in the products $x(t)x(t + \tau)$ encompassing all t, it can be expected that there are almost an equal number of

Fig. 8.42 Computation of
autocorrelation
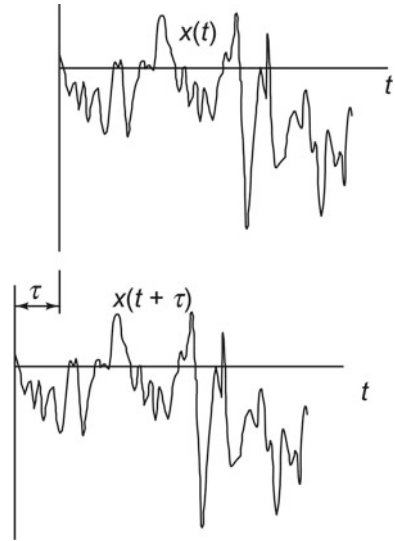
positive and negative values in the products, cancelling each other out and therefore
$R_{xx}(\infty)$ tends to zero. The autocorrelation function of highly random signals, such
as wide-band noise, shows a sharp spike at τ equal to zero that drops off rapidly with
$\pm\tau$. Therefore, except at τ equal to zero, wide-band random records have little or
no correlation. If we assume that we are dealing with a member function $x(t)$ of a
stationary ergodic random process, $R_{xx}(\tau)$ will also correspond to the autocorrelation
function of any other member function (Sect. 8.11).

Figure 8.43 shows the autocorrelation of some common functions. The first, sec-
ond, and third rows depict narrow-band random, broadband random, and sine wave
signals, respectively. The autocorrelation function of a sine wave is a cosine wave.
For a narrow-band signal, the autocorrelation is similar to that of a sine wave with
the envelope decaying slowly. The autocorrelation function of a wide-band random
signal is a sinc function. Figure 8.44 shows the setup for the computation of autocor-
relation.

Let us obtain the autocorrelation function of a signal with a non-zero mean. Let

$$x(t) = \xi(t) + \langle x(t) \rangle \tag{8.135}$$

Here, $\xi(t)$ is the fluctuating part of $x(t)$ and $\langle x(t) \rangle = \mu_x$ = mean value of $x(t)$.

$$\begin{aligned}
R_{xx}(\tau) &= \langle x(t)x(t+\tau) \rangle \\
&= \langle [\langle x(t) \rangle + \xi(t)][\langle x(t+\tau) \rangle + \xi(t+\tau)] \rangle \\
&= \langle x(t) \rangle^2 + R_{\xi\xi}(\tau)
\end{aligned} \tag{8.136}$$

This reduces to the above form for a stationary ergodic random process since $\langle x(t + \tau) \rangle = \langle x(t) \rangle$ and $\langle \xi(t + \tau) \rangle = \langle \xi(t) \rangle = 0$.
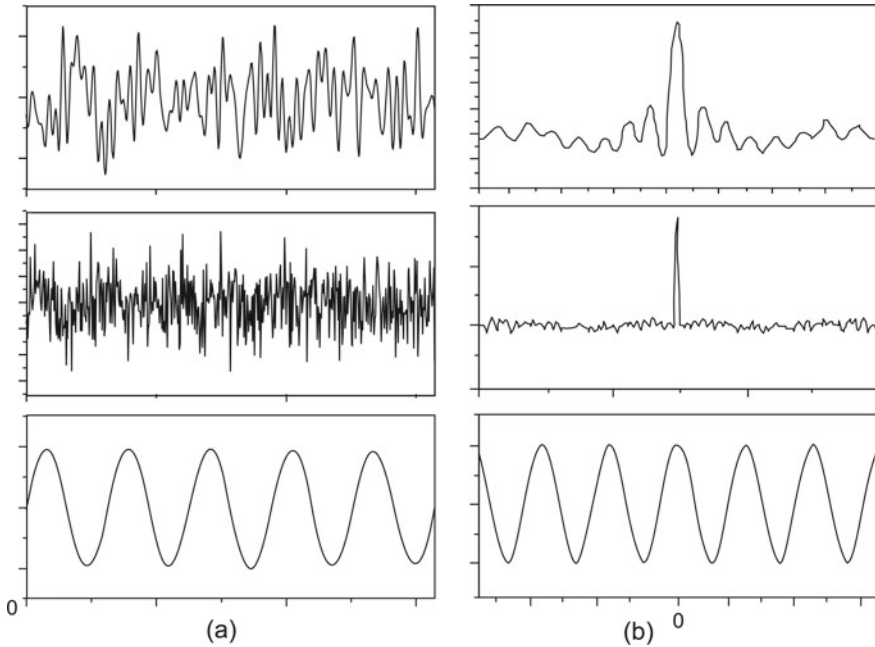
Fig. 8.43 Autocorrelation functions of different signals: **a** $x(t)$ versus t, **b** $R_{xx}(\tau)$ versus τ
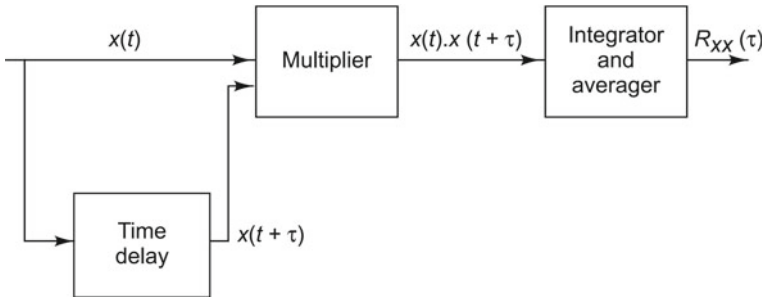
Fig. 8.44 Computation of autocorrelation

Cross-correlation The correlation between two records called the cross-correlation, is a measure of the similarity between the two records. A typical example could be $x(t)$ being the deflection at a point on a beam due to a load $f(t)$ at some other point on the same beam. Cross-correlation between two records $x(t)$ and $y(t)$ can be obtained by multiplying the ordinates of the two records at time t and time $t + \tau$ for all time t and determining the average value or expectation $\langle x(t)y(t + \tau)\rangle$ by dividing the sum of the products by the number of products. It is clear that cross-correlation will be highest if the two records are identical or similar. For dissimilar records, some products will be positive and some negative so that this sum will be smaller.
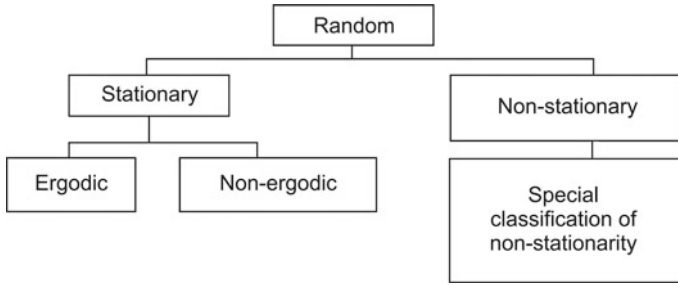
Fig. 8.45 Classification of random data

Cross-correlation between two quantities $x(t)$ and $y(t)$ may be calculated as

$$R_{xy}(\tau) = \langle x(t)y(t + \tau)\rangle \tag{8.137}$$

Correlation is related to spectral density through a Fourier relationship as shown later.

8.10   Classification of Random Data

Random processes may be classified as stationary and non-stationary as shown in Fig. 8.45. Ergodic and non-ergodic random processes form subsets of stationary random processes.

8.10.1   Stationarity

Let $x(t)$ be a random variable, say the random displacement of a vehicle traversing a rough road. This signal could be any other random quantity as well, such as random road undulations or jet noise. This quantity varies with time, but if the statistical characteristics of this time record are invariant with respect to time, then we may say that it is stationary. Consider a random process $\{x(t)\}$ consisting of N member functions or sample records $x_1(t)$ to $x_N(t)$ as shown in Fig. 8.46. The total collection of sample functions or sample records is called an ensemble and constitutes the random process $\{x(t)\}$.

The mean value of the signal across the ensemble at different instants of time $t_1, t_2, t_3, \ldots, t_n$ may be represented as $\mu_x(t_1), \mu_x(t_2), \mu_x(t_n)$. Here, $\mu_x(t_1)$ is computed as

$$\mu_x(t_1) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} x_k(t_1) \tag{8.138}$$
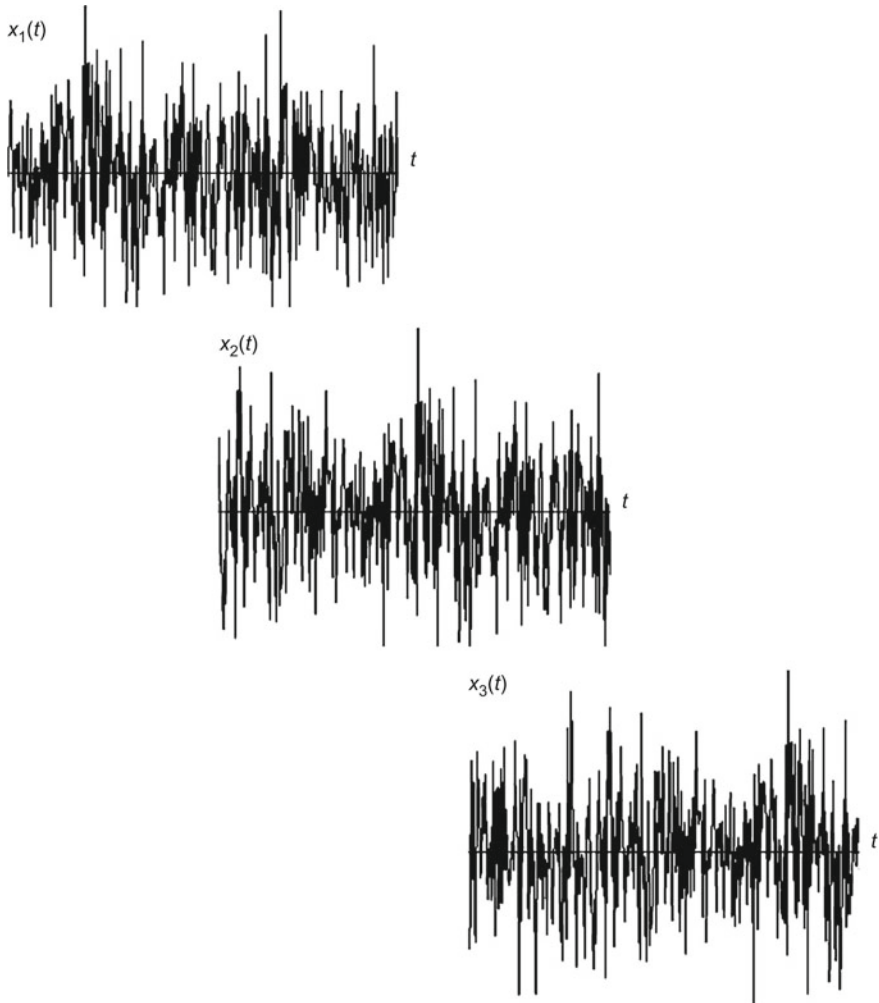
Fig. 8.46 Ensemble of records of a random phenomenon

Consider the scenario where

$$\mu_x(t_1) = \mu_x(t_2) = \mu_x(t_3) = \cdots = \mu_x(t_n) = \mu_x \qquad (8.139)$$

If these ensemble mean values are equal, i.e., if ensemble averages at time $t = t_1$, $t = t_2$, $t = t_3$, ..., $t = t_n$ for all the N sample functions are the same, then the random process is said to be invariant in the mean, or in other words, the ensemble average is independent of time t. It is possible to compute the ensemble autocorrelations at different instants of time as before. Ensemble autocorrelation at instant of time t_1 is

$$R_{xx}(t_1, \tau) = \lim_{N \to \infty} \sum_{k=1}^{N} x_k(t_1)x_k(t_1 + \tau) \tag{8.140}$$

Consider the case

$$R_{xx}(t_1, \tau) = R_{xx}(t_2, \tau) = R_{xx}(t_3, \tau) = \cdots = R_{xx}(t_n, \tau) = R_{xx}(\tau) \tag{8.141}$$

The random process is said to be invariant with respect to the autocorrelation function, or the ensemble autocorrelation is independent of time t. If conditions represented by Eqs. (8.139) and (8.141) are satisfied, then $\{x(t)\}$ is said to be a weakly stationary random process. If all possible higher order moments and joint moments are also time-invariant, then the random process is said to have strong stationarity.

8.10.2 Ergodicity

This is a subset of stationarity and requires that the time averages of all sample functions of duration T are equal. If $\mu_x(k)$ and $R_{xx}(\tau, k)$ which are the temporal averages over each sample function, do not differ when computed over different sample functions in the ensemble, then the random process is said to be ergodic. The sample averages as computed for the kth sample are shown below:

$$\mu_x(k) = \lim_{T \to \infty} \frac{1}{T} \int_0^T x_k(t)\mathrm{d}t \tag{8.142}$$

$$R_{xx}(\tau, k) = \lim_{T \to \infty} \frac{1}{T} \int_0^T x_k(t)x_k(t + \tau)\mathrm{d}t \tag{8.143}$$

Consider the situation where the ensemble averages are equal to the sample averages, i.e.,

$$\mu_x(1) = \begin{cases} \mu_x(2) = \mu_x(3) = \ldots \mu_x(N) = \mu_x \\ \mu_x(t_1) = \mu_x(t_2) = \mu_x(t_3) = \ldots \mu_x(t_n) \end{cases} \tag{8.144}$$

and

$$R_{xx}(\tau, 1) = \begin{cases} R_{xx}(\tau, 2) = R_{xx}(\tau, 3) = \cdots = R_{xx}(\tau, N) = R_{xx}(\tau) \\ R_{xx}(t_1, \tau) = R_{xx}(t_2, \tau) = R_{xx}(t_3, \tau) = \cdots = R_{xx}(t_n, \tau) \end{cases} \tag{8.145}$$

This is a case of the random process having both stationarity and ergodicity. A stationary, ergodic random signal is very convenient to handle, since one member function is sufficient to describe the entire random process and quantities like mean, variance, probability density, cumulative distribution, autocorrelation, power spectral density, etc., of one sample function are sufficient. Thus, for a stationary, ergodic random process, the statistical properties may be got from a single time function of sufficiently long duration. Though the assumption of stationarity and ergodicity may be a hypothetical one, it simplifies to a large extent the task of dealing with random variables. It is also to be mentioned that verification of weak stationarity can be taken as a justification of an assumption of strong stationarity for most practical problems.

8.11   Frequency Domain Representation of Random Signals

For periodic signals, we can have a Fourier series representation. A non-periodic signal, a transient signal, for example, can be expressed in the form of a Fourier series if it is considered to be periodic with infinite period. In this case, the discrete spectrum becomes a continuous curve as the Fourier series becomes the Fourier integral. Thus, the equations relating the time and frequency domain representations for a transient signal are given by the Fourier transform pair

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}t$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}\omega$$

$$(8.146)$$

$X(\omega)$ is called the Fourier transform of $x(t)$.

The two equations show near symmetry and they can be made more symmetric by writing the equations in terms of frequency f in Hz instead of circular frequency $\omega = 2\pi f$ in rad/s as shown.

$$X(f) = \int_{-\infty}^{\infty} x(t)\mathrm{e}^{-\mathrm{i}2\pi ft}\,\mathrm{d}t$$

$$x(t) = \int_{-\infty}^{\infty} X(f)\mathrm{e}^{\mathrm{i}2\pi ft}\,\mathrm{d}f$$

$$(8.147)$$

Application of Parseval's theorem to these equations gives some very useful results. Let the complex conjugate of $X(f)$ be denoted as $X^*(f)$.

$$X^*(f) = \int_{-\infty}^{\infty} x(t)e^{i2\pi ft}\,dt \tag{8.148}$$

We know that

$$X(f).X^*(f) = |X(f)|^2 \tag{8.149}$$

Thus,

$$\int_{-\infty}^{\infty} x^2(t)\,dt = \begin{cases} \int_{-\infty}^{\infty} x(t)\left[\int_{-\infty}^{\infty} X(f)e^{i2\pi ft}\,df\right]dt \\[2ex] \int_{-\infty}^{\infty} X(f)\left[\int_{-\infty}^{\infty} x(t)e^{i2\pi ft}\,dt\right]df \end{cases} \tag{8.150}$$

by changing the order of the integration. Therefore,

$$\int_{-\infty}^{\infty} x^2(t)\,dt = \begin{cases} \int_{-\infty}^{\infty} X(f).X^*(f)\,df \\[2ex] \int_{-\infty}^{\infty} |X(f)|^2\,df \end{cases} \tag{8.151}$$

Since negative frequencies have no physical meaning, this equation can be rewritten as

$$\int_{-\infty}^{\infty} x^2(t)\,dt = 2\int_{0}^{\infty} |X(f)|^2\,df \tag{8.152}$$

This result comes in handy when trying to define the PSD, which is the frequency domain representation of a random signal. A random signal cannot be expressed in the form of a Fourier transform since we have convergence issues in the computation. In order to have stationary properties, a random signal must be assumed to have an infinite length, in which case, the Fourier transform does not converge to a steady value. It is possible to define a new quantity called the PSD which has no convergence issues and which is applicable to all sample functions of a stationary, ergodic random process.

Fig. 8.47 Description of $x_T(t)$
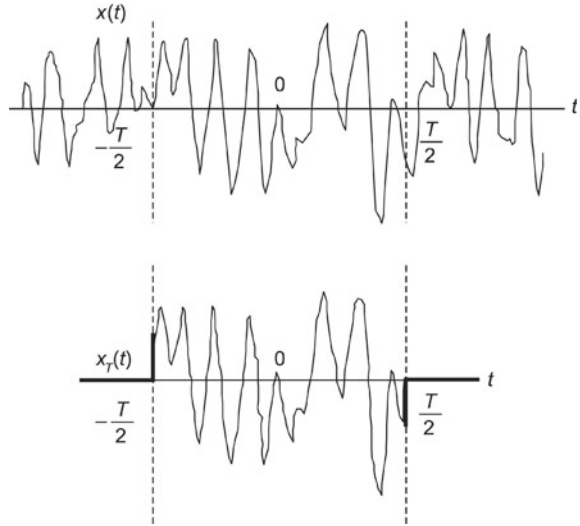
Let us consider a sample function $x(t)$ of a stationary random process. Since this exists from $t = -\infty$ to $t = +\infty$, its Fourier transform does not exist. We can, however, define a signal $x_T(t)$ identical to $x(t)$ over the interval $-(T/2) < t < (T/2)$ and zero elsewhere as shown in Fig. 8.47.

The mean square value of $x_T(t)$ is given by

$$\langle x_T^2(t)\rangle = \frac{1}{T} \int\limits_{-(T/2)}^{T/2} x_T^2(t)\mathrm{d}t = \frac{1}{T}\int\limits_{-\infty}^{\infty} x_T^2(t)\mathrm{d}t = \frac{2}{T}\int\limits_{0}^{\infty} |X_T(f)|^2\mathrm{d}f \qquad (8.153)$$

Letting $T \to \infty$, we can get an expression for the mean square value of $x(t)$.

$$\langle x^2(t)\rangle = \int\limits_{0}^{\infty} \lim_{T\to\infty}\left[\frac{2}{T}|X_T(f)|^2\right]\mathrm{d}f = \int\limits_{0}^{\infty} S_{xx}(f)\mathrm{d}f$$

where

$$S_{xx}(f) = \lim_{T\to\infty}\left[\frac{2}{T}|X_T(f)|^2\right] \qquad (8.154)$$

Here, $S_{xx}(f)$ is called the PSD or simply (auto) spectral density of $x(t)$ and represents the distribution of the harmonic content of the signal. It is represented in terms of energy/Hz (Fig. 8.47).

Thus, voltage PSD would be represented in terms of V^2/Hz. The units for PSDs of current, vibration acceleration, and pressure would be A^2/Hz, g^2/Hz, and Pa^2/Hz, respectively. $S_{xx}(f)\Delta f$ is the mean square value of the signal passed through a narrow bandpass filter of bandwidth Δf. The area under the PSD curve gives the root mean square value of the signal. We have already seen that the latter quantity may be obtained from a time domain description $x(t)$ or from a probabilistic description $p(x)$. A third method is now seen to exist in terms of PSD $S_{xx}(f)$,

$$\text{i.e., } \sigma^2 = \langle x^2(t) \rangle = \int\limits_{-\infty}^{\infty} x^2 p(x)\mathrm{d}x = \int\limits_{0}^{\infty} S_{xx}(f)\mathrm{d}f \qquad (8.155)$$

One salient aspect of the PSD is that since it is obtained from the modulus $|X(f)|$, it does not contain any phase information and is therefore applicable to a whole range of functions $x(t)$ which might have been generated by a physical system. Hence, it can be expected that different member functions $x(t)$ of a stationary ergodic random process $\{x(t)\}$ will have a common spectral density $S_{xx}(f)$. Therefore, the spectral content of all member functions $x(t)$ of a stationary ergodic random process may be defined by a single PSD $S_{xx}(f)$. However, a knowledge of $S_{xx}(f)$ alone is not sufficient to define a random process.

A random signal which has a constant spectral content over a range of frequencies is called white noise, through an imperfect analogy with white light. Though the concept of a white noise signal extending to infinite frequency is not physically meaningful, in practice, we do have what are called band-limited white noise signals extending over a limited range of frequencies and with a definite cut-off frequency. Thus, band-limited white noise is quite often used in vibration testing to determine the first few modes of a system with the knowledge that the response of the system at higher modes is very small. $S_{xx}(f)$ can be measured using a bandpass filter. Figure 8.48 shows typical PSDs of some common signals: they are from the top, narrow-band noise, broadband noise, and a harmonic signal.

In general, knowing the PSD, the mean square value can be determined, but nothing can be said about the PDF or cumulative distribution. If, however, a random process is known to be Gaussian, with PDF in the form given in Eq. (8.111), then the PDF can be written knowing the value of mean μ and standard deviation σ. This is the reason why scientists and engineers grab the opportunity to model a random process as Gaussian, with the process displaying the slightest hint of Gaussian properties (Fig. 8.49).
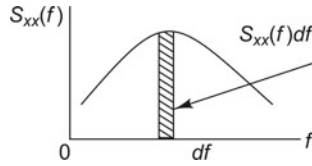
Fig. 8.48 Typical PSD plot
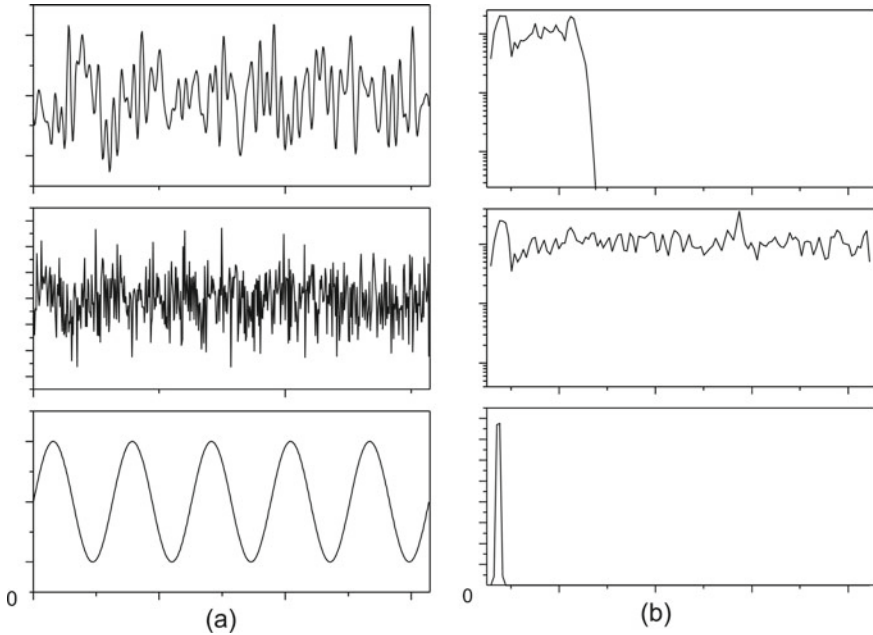
Fig. 8.49 PSDs of common signals: **a** $x(t)$ versus t, **b** $S_{xx}(f)$ versus f

8.11.1   Relationship Between Autocorrelation $R_{xx}(\tau)$ and Power Spectral Density $S_{xx}(f)$

We have intuitively deduced in Sect. 8.9.5.2 that autocorrelation is related to spectral content in a qualitative manner. The mathematical expression relating the two may be obtained as described below. We shall start with the function $x_T(t)$ coinciding with $x(t)$ over the range $-(T/2) < t < (T/2)$ and zero elsewhere as before (Fig. 8.47).

Autocorrelation $R_{xx}(\tau)$ can be defined as shown below, as the autocorrelation of function $x_T(t)$, mean value being taken over interval T.

$$R_{xx}(\tau) = \lim_{T \to \infty} R_{x_T x_T}(\tau) = \lim_{T \to \infty} \langle x_T(t) x_T(t + \tau) \rangle \qquad (8.156)$$

Taking Fourier transform of $R_{x_T x_T}(\tau)$ and making the simplification $s = t + \tau$, we get

$$\int_{-\infty}^{\infty} R_{x_T x_T}(\tau) e^{-i2\pi f \tau} d\tau = \begin{cases} \int\limits_{-\infty}^{\infty} \frac{1}{T} \left[\int\limits_{-\infty}^{\infty} x_T(t) x_T(t + \tau) dt \right] e^{-i2\pi f \tau} d\tau \\[2mm] \frac{1}{T} \int\limits_{-\infty}^{\infty} \left[\int\limits_{-\infty}^{\infty} x_T(t) x_T(t + \tau) e^{-i2\pi f \tau} dt \right] d\tau \\[2mm] \frac{1}{T} \int\limits_{-\infty}^{\infty} \left[\int\limits_{-\infty}^{\infty} x_T(t) x_T(t + \tau) e^{i2\pi f t} e^{-i2\pi f(t+\tau)} dt \right] d\tau \\[2mm] \frac{1}{T} \int\limits_{-\infty}^{\infty} \left[\int\limits_{-\infty}^{\infty} x_T(t) x_T(s) e^{i2\pi f t} e^{-i2\pi f s} dt \right] ds \\[2mm] \frac{1}{T} \int\limits_{-\infty}^{\infty} x_T(t) e^{i2\pi f t} dt \int\limits_{-\infty}^{\infty} x_T(s) e^{-i2\pi f s} ds \\[2mm] \frac{1}{T} X_T^*(f) \cdot X_T(f) \\[2mm] \frac{1}{T} |X_T(f)|^2 \end{cases}$$

$$(8.157)$$

Letting $T \to \infty$ so that $R_{x_T x_T}(\tau)$ becomes $R_{xx}(\tau)$ and using Eqs. (8.154) and (8.157), we have

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} \frac{1}{2} S_{xx}(f) e^{i2\pi f \tau} df \qquad (8.158)$$

i.e., the autocorrelation function and the PSD are related through a Fourier transform relationship. This equation is called the Wiener–Khinchin equation. Spectral density may be easier to measure, while autocorrelation may be more convenient for time domain analysis.

8.12   Response of an SDOF System

> **INTERESTING FACTS:**   We often study signals and systems to get an idea about system behaviour. When studying a system, we have to deal with three quantities: (i) input to the system, (ii) system behaviour characterized in terms of its impulse response function or frequency response function or system parameters, and (iii) output of the system. Typically, the three operations of common interest are: given any two, how do you find the third? When dealing with signals and systems, it is very important to know what approximations are to be made regarding the signals as well as the system. This is largely dictated by the objective of the exercise. One may model the input as deterministic if one is interested in finding the steady-state response. On the other hand, if one is interested in finding the fluctuations in the output due to fluctuations in the input, then the input of concern would be the fluctuations in the input, which would no longer be deterministic. Wherever possible and within allowable error limits, we simplify and try to model periodic signals as simple harmonic, nearly periodic signals as periodic and random signals as stationary and ergodic to reduce computational effort. Similarly, in the modelling of a system, for reasonably small inputs, it may be assumed that the system is operating in its linear range and it is then possible to apply the principle of superposition. But for very large inputs, for which a linear relationship between the inputs and outputs no longer exists, the assumption of a linear system would be wrong. We always try to make simplifications in modelling systems, a stable, linear time-invariant (LTI) model being a very convenient assumption that makes computations fairly simple. We generally start with the simplest model of a system, namely single-input single-output (SISO) model and once sufficient confidence has been built, we can graduate to more complex models such as single-input multiple-output (SIMO) models or multi-input multi-output (MIMO) models.

In the following discussions, we shall try to derive the response of an SDOF system to transient and random excitations. These concepts will come in handy in understanding the modal analysis procedures that will be discussed in Chap. 9. However, before we do so, it is imperative to understand how the SDOF system responds to a harmonically varying load. Then we may go on to study transient and random loads. The assumptions made in all the following analyses are that the system is linear, time-invariant, and stable.

8.12.1 Response to Harmonic Excitation

This has been discussed at length in Sect. 2.2.6.2 for an SDOF system and in Sect. 2.4.5 for an MDOF system. Let us refer to Fig. 2.9a which represents a viscously damped SDOF system. Let $\alpha(\omega)$ be its complex frequency response function (FRF), or receptance, to be more specific. We shall derive an expression for this letting the harmonic loading to be represented by $f(t) = F_0 e^{i\omega t}$. The dynamic equation of the system is

$$m\ddot{x} + c\dot{x} + kx = F_0 e^{i\omega t} \tag{8.159}$$

Since the response of an LTI system to harmonic loading is harmonic at the same frequency as the excitation, Eq. (8.159) may be simplified as

$$(-m\omega^2 + i\omega c + k)X_0 = F_0 \tag{8.160}$$

$$\text{or} \quad X_0 = \begin{cases} \dfrac{F_0}{(k - m\omega^2) + i\omega c} \\ \alpha(\omega)F_0 \end{cases} \tag{8.161}$$

The above equation is a statement of the fact that a harmonic function is an eigenfunction of an LTI system and $\alpha(\omega)$ is the eigenvalue or complex FRF of the system.

$$\alpha(\omega) = \frac{1}{(k - m\omega^2) + i\omega c} \tag{8.162}$$

The response $x(t)$ of the system is described by the equations

$$\begin{aligned} x(t) &= \frac{F_0}{(k - m\omega^2) + i\omega c} e^{i\omega t} \\ &= \frac{F_0(k - m\omega^2 - i\omega c)}{(k - m\omega^2)^2 + \omega^2 c^2} e^{i\omega t} \\ &= \frac{F_0 e^{i\phi}}{\sqrt{(k - m\omega^2)^2 + \omega^2 c^2}} e^{i\omega t}, \text{ where } \tan\phi = \frac{\omega c}{k - m\omega^2} \end{aligned} \tag{8.163}$$

Working in terms of actual frequency f instead of circular frequency ω, we get

$$\alpha(f) = \frac{1}{k - m4\pi^2 f^2 + i2\pi f c} \tag{8.164}$$

This quantity $\alpha(f)$ is the response of the SDOF system to a complex force of unit modulus and proportional to $e^{i\omega t}$ and is called the receptance of the system. The

receptance provides a complete description of the displacement response of the LTI system to harmonic force excitation.

The above result may be obtained in an alternate fashion as described below. We know that the response to a force $f(t) = F_0 e^{i2\pi ft}$ is given by the convolution integral

$$x(t) = \begin{cases} \int\limits_0^\infty h(\tau) F_0 e^{i2\pi f(t-\tau)} \mathrm{d}\tau \\ F_0 e^{i2\pi ft} \int\limits_0^\infty h(\tau) e^{-j2\pi f\tau} \mathrm{d}\tau \end{cases} \tag{8.165}$$

where $h(t)$ is called the Impulse Response Function as discussed in Sect. 8.4.3 and is the response to a unit impulse. However, due to the eigenfunction property of $e^{j2\pi ft}$, we know from Eq. (8.161) that

$$x(t) = \alpha(f) F_0 e^{i2\pi ft} \tag{8.166}$$

Therefore,

$$\alpha(f) = \int\limits_0^\infty h(\tau) e^{-i2\pi f\tau} \mathrm{d}\tau \tag{8.167}$$

i.e., $\alpha(f)$ is the Fourier transform of $h(t)$ if $h(t)$ is assumed to be 0 for $\tau < 0$, i.e., for a causal system. For any other arbitrary signal, which can be represented as the superposition of a number of harmonic signals, i.e., for an excitation which can be expressed in the form of a Fourier series, the total response can be obtained by superimposing the individual responses to all harmonic excitations, taking phase information into account.

8.12.2 Response to Transient Excitation

We often come across situations where a system is subjected to impulsive disturbances (large-amplitude, short duration forces of the order of a few seconds). Typical examples are excitations during forging, punching, and blasting operations. Duhamel's integral or convolution integral is used for the solution of transient vibration problems in linear systems and is based on the principle of superposition of responses to a sequence of impulses. Any impulsive loading may be expressed as a superposition of amplitude scaled and time-shifted unit impulses, i.e.,

$$F(t) = I\delta(t) \tag{8.168}$$

Fig. 8.50 Arbitrary load
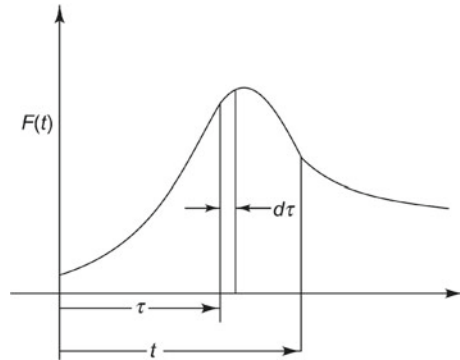
where I is the amplitude of the impulse and $\delta(t)$ is the Dirac δ function described in Eqs. (8.13) and (8.14). For an arbitrary loading of the form shown in Fig. 8.50, it is possible to compute the response of an LTI system as discussed below.

The loading itself may be expressed in the form of a convolution integral as shown in Eq. 8.169.

$$F(t) = \int_0^\infty \delta(\tau)F(t-\tau)\,\mathrm{d}\tau \qquad (8.169)$$

The total response to this loading is obtained using the convolution integral

$$x(t) = \int_0^\infty h(\tau)F(t-\tau)\,\mathrm{d}\tau \qquad (8.170)$$

i.e., the response is expressed as the superposition of amplitude-scaled, time-shifted IRFs. It is also to be remembered that $h(t)$ is the inverse Fourier transform of $H(\omega)$; $h(t)$ and $H(\omega)$ represent the system parameters in the time and frequency domains, respectively.

8.12.3 Response-to-Random Loading

Any random excitation that we may input will be idealized as a stationary, ergodic random process, making the response also stationary and ergodic. It is to be borne in mind that the random response which is calculated will have no more details than that present in the mathematical description of the excitation function. Let us consider a randomly varying force $F(t)$. Let autocorrelation function of the force be denoted by $R_{FF}(\tau)$ and its auto-spectral density by $S_{FF}(f)$. Let the displacement output of the SDOF system to this force be denoted by $x(t)$ with autocorrelation $R_{xx}(\tau)$ and

spectral density $S_{xx}(f)$. The assumption behind this analysis is that the system is an LTI system and that the input and output random signals are stationary, ergodic random processes with zero mean. It is also assumed that the input is applied at a single point and we are concerned with the output displacement at a single point. The aim of the exercise is to determine $R_{xx}(\tau)$ in terms of $R_{FF}(\tau)$ and $S_{xx}(f)$ in terms of $S_{FF}(f)$. The response $x(t)$ of the system may be written in terms of the convolution integral

$$x(t) = \int_0^\infty h(\tau_1)F(t - \tau_1)d\tau_1$$

(8.171)

$$x(t + \tau) = \int_0^\infty h(\tau_2)F(t + \tau - \tau_2)d\tau_2$$

The autocorrelation of the response is given by

$$R_{xx}(\tau) = \langle x(t)x(t + \tau) \rangle$$

$$= \lim_{T \to \infty} \frac{1}{T} \int_0^T \left[\int_0^\infty h(\tau_1)F(t - \tau_1)d\tau_1 \right] \left[\int_0^\infty h(\tau_2)F(t + \tau - \tau_2)d\tau_2 \right] dt$$

$$= \int_0^\infty h(\tau_1) \int_0^\infty h(\tau_2) \langle F(t - \tau_1)F(t + \tau - \tau_2) \rangle d\tau_2 d\tau_1$$

(8.172)

$$= \int_0^\infty h(\tau_1) \int_0^\infty h(\tau_2) \langle F(t)F(t + \tau_1 - \tau_2 + \tau) \rangle d\tau_2\, d\tau_1$$

$$= \int_0^\infty h(\tau_1) \int_0^\infty h(\tau_2) R_{FF}(\tau_1 - \tau_2 + \tau) d\tau_2\, d\tau_1$$

The last of the above equations follows from the fact that a change in the origin of t makes no difference to the averaged quantity if the process is stationary. The spectral density may be obtained by taking the Fourier transform of $R_{xx}(\tau)$.

$$S_{xx}(f) = 2 \int_{-\infty}^\infty R_{xx}(\tau)e^{-i2\pi f\tau}d\tau$$

$$= 2 \int\limits_{-\infty}^{\infty} \left[\int\limits_{0}^{\infty} h(\tau_1) \int\limits_{0}^{\infty} h(\tau_2) R_{FF}(\tau_1 - \tau_2 + \tau) d\tau_2 \, d\tau_1 \right] e^{-i2\pi f \tau} d\tau$$

$$= 2 \int\limits_{0}^{\infty} h(\tau_1) \int\limits_{0}^{\infty} h(\tau_2) \left[\int\limits_{-\infty}^{\infty} R_{FF}(\tau_1 - \tau_2 + \tau) e^{-i2\pi f \tau} d\tau \right] d\tau_2 \, d\tau_1 \quad (8.173)$$

$$= 2 \int\limits_{0}^{\infty} h(\tau_1) \, e^{i2\pi f \tau_1} \int\limits_{0}^{\infty} h(\tau_2) e^{-i2\pi f \tau_2}$$

$$\times \int\limits_{-\infty}^{\infty} R_{FF}(\tau_1 - \tau_2 + \tau) e^{-i2\pi f (\tau_1 - \tau_2 + \tau)} d(\tau_1 - \tau_2 + \tau) d\tau_2 \, d\tau_1$$

$$= 2 \int\limits_{0}^{\infty} h(\tau_1) \, e^{i2\pi f \tau_1} \, d\tau_1 \int\limits_{0}^{\infty} h(\tau_2) e^{-i2\pi f \tau_2} \, d\tau_2$$

$$\times \int\limits_{-\infty}^{\infty} R_{FF}(\tau_1 - \tau_2 + \tau) e^{-i2\pi f (\tau_1 - \tau_2 + \tau)} d(\tau_1 - \tau_2 + \tau)$$

Therefore,

$$S_{xx}(f) = \begin{cases} H^*(f) H(f) \frac{1}{2} S_{FF}(f) = |H(f)|^2 S_{FF}(f) \\[2mm] |\alpha(f)|^2 S_{FF}(f) \end{cases} \quad (8.174)$$

This is an elegant and simple equation relating the PSD of the response to the PSD of excitation through the transfer function. The equation states that the PSD of the response at any frequency is the product of the PSD of the exciting force at that frequency and square of the modulus of the receptance at that frequency. It is worthwhile noting that phase information is lacking in this equation and hence obtaining $H(f)$ which in the above exercise is the receptance $|\alpha(f)|$ is not difficult. Once the PSD of the response is obtained, the mean square value of the response can also be got.

8.12.3.1   Response Involving Cross-Correlations

Figure 8.51 shows a system with two random excitations $P(t)$ and $Q(t)$ and response $x(t)$. Let $\alpha_{xP}(f)$ and $\alpha_{xQ}(f)$ (denoted simply by α_{xP} and α_{xQ}) be the receptances for harmonic excitation at locations P and Q and let the corresponding IRFs be $h_{xP}(t)$ and $h_{xQ}(t)$, respectively.

Fig. 8.51 Response of a
system to random inputs
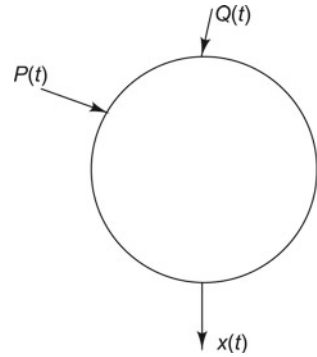
Let autocorrelations of $x(t)$, $P(t)$, and $Q(t)$ be $R_{xx}(\tau)$, $R_{PP}(\tau)$, and $R_{QQ}(\tau)$, respectively. Then the autocorrelation of the response may be expressed as

$$R_{xx}(\tau) = \langle x(t)x(t+\tau)\rangle$$

$$= \left\langle \int_0^\infty [h_{xP}(\tau_1)P(t-\tau_1) + h_{xQ}(\tau_1)Q(t-\tau_1)]d\tau_1 \right.$$

$$\left. \times \int_0^\infty [h_{xP}(\tau_2)P(t-\tau_2+\tau) + h_{xQ}(\tau_2)Q(t-\tau_2+\tau)]d\tau_2 \right\rangle$$

$$= \left\langle \int_0^\infty h_{xP}(\tau_1)P(t-\tau_1)d\tau_1 \int_0^\infty h_{xP}(\tau_2)P(t-\tau_2+\tau)d\tau_2 \right\rangle$$

$$+ \left\langle \int_0^\infty h_{xP}(\tau_1)P(t-\tau_1)d\tau_1 \int_0^\infty h_{xQ}(\tau_2)Q(t-\tau_2+\tau)d\tau_2 \right\rangle$$

$$+ \left\langle \int_0^\infty h_{xQ}(\tau_1)Q(t-\tau_1)d\tau_1 \int_0^\infty h_{xP}(\tau_2)P(t-\tau_2+\tau)d\tau_2 \right\rangle$$

$$+ \left\langle \int_0^\infty h_{xQ}(\tau_1)Q(t-\tau_1)d\tau_1 \int_0^\infty h_{xQ}(\tau_2)Q(t-\tau_2+\tau)d\tau_2 \right\rangle$$

$$R_{xx}(\tau) = \int_0^\infty h_{xP}(\tau_1) \left[\int_0^\infty h_{xP}(\tau_2) \langle P(t)P(t+\tau_1-\tau_2+\tau)\rangle d\tau_2 \right] d\tau_1$$

$$+ \int_0^\infty h_{xP}(\tau_1) \left[\int_0^\infty h_{xQ}(\tau_2) \langle P(t)Q(t+\tau_1-\tau_2+\tau)\rangle d\tau_2 \right] d\tau_1$$

$$+ \int_0^\infty h_{xQ}(\tau_1) \left[\int_0^\infty h_{xP}(\tau_2) \langle Q(t)P(t + \tau_1 - \tau_2 + \tau) \rangle \, d\tau_2 \right] d\tau_1$$

$$+ \int_0^\infty h_{xQ}(\tau_1) \left[\int_0^\infty h_{xQ}(\tau_2) \langle Q(t)Q(t + \tau_1 - \tau_2 + \tau) \rangle \, d\tau_2 \right] d\tau_1 \quad (8.175)$$

The first and the last terms in the expansion are autocorrelation functions and may be written as $R_{PP}(\tau_1 - \tau_2 + \tau)$ and $R_{QQ}(\tau_1 - \tau_2 + \tau)$, respectively. The corresponding quantities in the second and third terms are cross-correlation functions $R_{PQ}(\tau_1 - \tau_2 + \tau)$ and $R_{QP}(\tau_1 - \tau_2 + \tau)$, respectively. Hence, Eq. (8.175) may be simplified as

$$R_{xx}(\tau) = \int_0^\infty h_{xP}(\tau_1) \left[\int_0^\infty h_{xP}(\tau_2) R_{PP}(\tau_1 - \tau_2 + \tau) d\tau_2 \right] d\tau_1$$

$$+ \int_0^\infty h_{xP}(\tau_1) \left[\int_0^\infty h_{xQ}(\tau_2) R_{PQ}(\tau_1 - \tau_2 + \tau) d\tau_2 \right] d\tau_1$$

$$+ \int_0^\infty h_{xQ}(\tau_1) \left[\int_0^\infty h_{xP}(\tau_2) R_{QP}(\tau_1 - \tau_2 + \tau) d\tau_2 \right] d\tau_1$$

$$+ \int_0^\infty h_{xQ}(\tau_1) \left[\int_0^\infty h_{xQ}(\tau_2) R_{QQ}(\tau_1 - \tau_2 + \tau) d\tau_2 \right] d\tau_1 \quad (8.176)$$

To determine the PSD of the response, we use the relationship

$$S_{xx}(f) = 2 \int_{-\infty}^\infty R_{xx}(\tau) e^{-i2\pi f \tau} \, d\tau$$

$$= 2\Im\{R_{xx}(\tau)\} \quad (8.177)$$

$$= \alpha_{xP}{}^* \alpha_{xP} S_{PP}(f) + \alpha_{xP}{}^* \alpha_{xQ} S_{PQ}(f) + \alpha_{xQ}{}^* \alpha_{xP} S_{QP}(f)$$

$$+ \alpha_{xQ}{}^* \alpha_{xQ} S_{QQ}(f)$$

Here, the cross-PSD terms are of the form

$$S_{PQ}(f) = 2 \int_{-\infty}^\infty R_{PQ}(\tau) e^{-i2\pi f \tau} \, d\tau \quad (8.178)$$

Thus, it is seen that the response PSD is a function of auto- and cross-PSDs of input.

Case (i): If $P(t)$ and $Q(t)$ are independent, then $S_{PQ}(f) = S_{QP}(f) = 0$ and therefore $S_{xx}(f)$ reduces to the form

$$S_{xx}(f) = |\alpha_{xP}|^2 S_{PP}(f) + |\alpha_{xQ}|^2 S_{QQ}(f) \qquad (8.179)$$

Case (ii): If $P(t)$ and $Q(t)$ are directly correlated so that $Q(t) = kP(t)$ where k is a constant, then the auto- and cross-correlations can be written as

$$
\begin{aligned}
R_{PQ}(\tau) &= \langle P(t)Q(t+\tau)\rangle = \langle P(t)kP(t+\tau)\rangle = kR_{PP}(\tau) \\
R_{QP}(\tau) &= \langle kP(t).P(t+\tau)\rangle = kR_{PP}(\tau) \\
R_{QQ}(\tau) &= \langle kP(t).kP(t+\tau)\rangle = k^2 R_{PP}(\tau)
\end{aligned}
\qquad (8.180)
$$

The PSD of the response is therefore

$$
\begin{aligned}
S_{xx}(f) &= \alpha_{xP}^*\alpha_{xP} S_{PP}(f) + \alpha_{xP}^* k S_{PP}(f) + \alpha_{xQ}^*\alpha_{xP} k S_{PP}(f) \\
&\quad + \alpha_{xQ}^*\alpha_{xQ} k^2 S_{PP}(f) \\
&= (\alpha_{xP}^* + k\alpha_{xQ}^*)(\alpha_{xP} + k\alpha_{xQ}) S_{PP}(f) \\
&= |\alpha_{xP} + k\alpha_{xQ}|^2 S_{PP}(f)
\end{aligned}
\qquad (8.181)
$$

Case (iii): If the two forces are directly correlated, in which case $S_{PP}(f) = S_{QQ}(f) = S(f)$, the response PSD is

$$
S_{xx}(f) =
\begin{cases}
|\alpha_{xP} + \alpha_{xQ}|^2 S(f) \\
[|\alpha_{xP}|^2 + |\alpha_{xQ}|^2 + 2|\alpha_{xP}||\alpha_{xQ}| \cos\phi] S(f)
\end{cases}
\qquad (8.182)
$$

where ϕ is the phase difference between the two receptances at a particular frequency f. When $\phi = \pm\frac{\pi}{2}$, $\cos\phi = 0$ and this amounts to uncorrelated loadings.

Case (iv): If $Q(t)$ reproduces $P(t)$ after a lag τ_0 so that $Q(t) = P(t+\tau_0)$, the cross-correlation is

$$
R_{PQ}(\tau) =
\begin{cases}
\langle P(t)Q(t+\tau)\rangle \\
\langle P(t)P(t+\tau_0+\tau)\rangle = R_{PP}(\tau_0+\tau)
\end{cases}
\qquad (8.183)
$$

The corresponding cross-spectral densities are given by the expressions

Fig. 8.52 Multiple random loadings
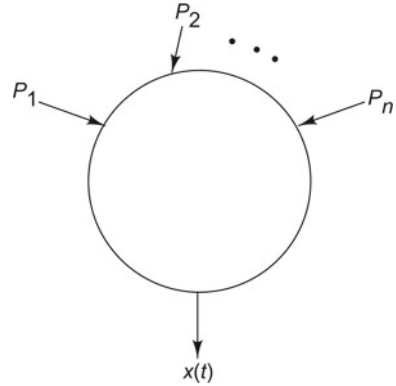
$$S_{PQ}(f) = 2 \int_{-\infty}^{\infty} R_{PP}(\tau_0 + \tau) e^{-i2\pi f \tau} d\tau$$

$$= 2e^{i2\pi f \tau_0} \int_{-\infty}^{\infty} R_{PP}(\tau_0 + \tau) e^{-i2\pi f(\tau_0 + \tau)} d(\tau_0 + \tau) \qquad (8.184)$$

$$= e^{i2\pi f \tau_0} S_{PP}(f)$$

Obviously,

$$S_{QQ}(f) = S_{PP}(f) \qquad (8.185)$$

The PSD of the response is given by

$$S_{xx}(f) = (\alpha_{xP}{}^* \alpha_{xP} + e^{i2\pi f \tau_0} \alpha_{xP}{}^* \alpha_{xQ} + e^{-i2\pi f \tau_0} \alpha_{xQ}{}^* \alpha_{xP} + \alpha_{xQ}{}^* \alpha_{xQ}) S_{PP}(f) \qquad (8.186)$$

Case (v): We often have to find the response to more than two loadings. Let us consider the case when n random loads act simultaneously, as shown in Fig. 8.52.

In general, for n forces $P_1, P_2, \ldots, P_n$ acting on a system to produce responses $x_1(t), x_2(t), \ldots, x_n(t)$, at a particular location, the total response $x(t)$ at that location is given by

$$z(t) = x_1(t) + x_2(t) + x_3(t) + \cdots + x_n(t) \qquad (8.187)$$

The autocorrelation of the response takes the form shown below

$$R_{zz}(\tau) = \sum_{r=1}^{n} \sum_{s=1}^{n} R_{x_r x_s}(\tau) \qquad (8.188)$$

The response PSD is of the form

$$S_{xx}(f) = \sum_{r=1}^{n} \sum_{s=1}^{n} \alpha_{xP_r}{}^{*} \alpha_{xP_s} S_{P_r P_s}(f) \tag{8.189}$$

It is seen that the autocorrelation of the response has terms containing the autocorrelations of the input forces, as well as all possible combinations of cross-correlations between the inputs. This is true also of the auto-spectral density of the output.

Many of the fundamental concepts of signal processing discussed in this chapter will have to be used when making measurements and while conducting the experiments described in Chaps. 10–12.

List of Symbols

Symbol	Meaning
a	constant
$a_1, a_2, a_3, \ldots, a_N$	bits or constants
A	constant, operator
A_k	constant
A, A_{ref}	amplitude-like quantities
b	constant
B	frequency bandwidth (Hz)
B_h, B_k, B_x	constants
c	viscous damping coefficient (N.s/m)
c_k	Fourier series coefficients
C	constant
$D_0, D_1, D_2, \ldots$	bits
f	frequency (Hz)
f_{min}, f_{max}	minimum and maximum frequencies (Hz)
f_s	sampling frequency $= 1/T_s$ (Hz)
$f(t)$	continuous-time signal
$f(t)$, $F(t)$	force (N)
F_0	harmonic force amplitude (N)
$h[n]$	impulse response function of discrete-time system
$h(t)$	impulse response function of continuous-time system
$h_r(t)$	impulse response function of reconstruction filter
$h_{xP}(t)$ and $h_{xQ}(t)$	IRFs for harmonic excitations $P(t)$ and $Q(t)$
$H(\omega)$	frequency response function
$H_r(\Omega)$	frequency response function of reconstruction filter
i	$\sqrt{-1}$
I	amplitude of impulse
k	scaling factor

Symbol	Meaning
k	spring stiffness (N/m)
k_x, k_y	radii of gyration about x- and y-axes (m)
K	constant
m	mass (kg)
n	sample number, number of bits
N	number of points/lines in DFT computation
$p(b/a)$	probability of B given that A has occurred
$p(n)$	probability density function of discrete variable n
$p(x)$	probability density function of continuous variable x
$P_1, P_2, \ldots, P_n$	forces (N)
$P(n)$	probability distribution/cumulative distribution function of discrete variable n
$P(x)$	probability distribution/cumulative distribution of continuous variable x
$P(x, y)$	joint cumulative distribution function
P, P_{ref}	power-like quantities
$R_{PP}(\tau), R_{QQ}(\tau)$	autocorrelation of $P(t), Q(t)$
$R_{PQ}(\tau), R_{QP}(\tau), R_{x_r x_z}(\tau)$	cross-correlations
$R_{xx}(\tau)$	autocorrelation of $x(t)$
$R_{xy}(\tau)$	cross-correlation between $x(t)$ and $y(t)$
$s(t)$	sampling signal
$S(\Omega)$	Fourier transform of sampling signal $s(t)$
$S_{PQ}(f), S_{QP}(f)$	cross-spectral densities
$S_{xx}(f)$	auto-spectral density/ PSD of $x(t)$
t	time (s)
T	total time duration of data (s)
T_s	clock/sampling rate/sampling interval (s)
x	quantization error, continuous variable
x_d	output of comparator
x_r	reference signal
x_{FS}	full-scale output of DAC
$x[n]$	nth sample of sequence
$x(t)$	continuous-time signal $x(t)$ or displacement (m)
$x_1(t), x_2(t), x_3(t)$	continuous-time signals
$x_0(t)$	output of sample-and-hold circuit
$x_e(t)$	even/conjugate symmetric part of real/complex $x(t)$
$x_o(t)$	odd/conjugate anti-symmetric part of real/complex $x(t)$

Symbol	Meaning
$x_r(t)$	reconstructed signal
$x_s(t)$	sampled signal $= x_c(t) \cdot s(t)$
$x_w(t)$	windowed signal $= x(t) \cdot w(t)$
$\{x[n]\}$	sequence
$\{x(t)\}$	random process of N member functions $x_1(t),$ $x_2(t), \ldots, x_N(t)$
X_0	harmonic displacement amplitude
$X[k]$	kth sample of Fourier transform of $x[n]$
$X(e^{i\omega})$	Fourier transform of sequence $x[n]$
$X_s(f), X_s(\omega)$	Fourier transform of sampled signal $x_s(t)$
$X(\omega), X_1(\omega), X_2(\omega)$	Fourier transforms of $x(t), x_1(t), x_2(t)$
$X_c(\omega)$	Fourier transform of $x_c(t)$
$X(\omega), Y(\omega)$	Fourier transforms of $x(t)$ and $y(t)$
$y[n]$	nth sample of sequence
$y(t)$	continuous-time signal or displacement (m)
$w(t)$	windowing signal
$\alpha(f)$	receptance FRF
$\alpha_{xP}(f), \alpha_{xQ}(f)$	receptance FRFs for harmonic excitations $P(t)$ and $Q(t)$
$\delta(t)$	unit impulse function/Dirac delta function
$\delta[n]$	unit sample sequence
Δf	bandwidth of a narrow-band, bandpass filter (Hz)
$\Delta t_1, \Delta t_2, \Delta t_3, \ldots$	time intervals (s)
ε	very, small arbitrary quantity
μ	mean
μ_x	mean value of x
$\xi(t)$	fluctuating part of $x(t)$
σ	standard deviation
σ_x	standard deviation of x
σ_x^2	variance of x
$\phi(t)$	continuous-time signal
$\phi(\omega)$	phase angle at frequency ω (rad)
ω	circular frequency for continuous-time signal (rad/s)
ω	circular frequency for discrete-time signal (rad/sample)
ω_c	cut-off frequency (rad/s)
Ω	frequency (rad/s)
Ω_c	cut-off frequency (rad/s)
Ω_N	highest frequency in a signal (rad/s)
Ω_s	sampling frequency (rad/s)

Special Operators

Notation	Operation
E	expectation operator
$E[X]$	expected value of X
$E(X^n)$	nth moment of probability density function of random variable X
$\Im\{\ \}$	Fourier transform
$\Im^{-1}\{\ \}$	inverse Fourier transform
Im	imaginary part
Re	real part
$\langle\ \rangle$	expectation operator
$x(t) * h(t)$	convolution of $x(t)$ with $h(t)$
$X^*(\omega)$	complex conjugate of $X(\omega)$
$X_R(\omega)$	real part of $X(\omega)$
$X_I(\omega)$	imaginary part of $X(\omega)$
$\bar{x}$	mean value of x
$x(t) \leftrightarrow X(\omega)$	$x(t)$ has a Fourier transform $X(\omega)$

Abbreviations

ADC	analogue-to-digital converter or A/D converter
BIBO	bounded-input bounded-output
CDF	cumulative distribution function
C/D	continuous-to-discrete
DAC	digital-to-analogue converter or D/A converter
DAS/DAQ	data acquisition system
DFT	discrete Fourier transform
FFT	fast Fourier transform
FRF	frequency response function
FS	full-scale
IRF	impulse response function
LSB	least significant bit
LTI	linear time-invariant
MSB	most significant bit
PC	personal computer
PDF	probability density function
PSD	power spectral density

RPM   revolutions per minute
SE   single-ended
SISO   single-input single-output
SNR   signal-to-noise ratio
SONAR   sound navigation and ranging
USB   universal serial bus

Questions

1. What is the need for digital conversion of analogue signals?
2. If an SNR of 90 dB is required in an ADC, how many bits should it have?
3. Draw the flipped versions of sine and cosine signals.
4. Draw $A\delta(t - t_0)$ where $A = 4$ and $t_0 = 8$.
5. When a signal $x(t)$ is multiplied by the signal in question 4 what happens to it?
6. What do you mean by causal and anti-causal signals?
7. How can you check for the stability of a system from its IRF?
8. How can you check for causality of a system from its IRF?
9. Give 2 examples each of even and odd signals.
10. What is meant by time inversion of a signal?
11. What is meant by time scaling of a signal? Explain with equations.
12. When is a system said to be memoryless?
13. Draw the time-delayed and time-advanced versions of a signal.
14. What does duality between time and frequency domains imply?
15. The impulse response function $h(t)$ of a system can be used to find out if the system is (i) linear, (ii) stable, (iii) causal and (iv) memoryless. Tick the appropriate choices. Give reasons.
16. What is the advantage in being able to model a system as LTI?
17. What is the Fourier transform of $\delta(t)$?
18. What is the difference between Fourier series and Fourier transform? For which class of signals are they used?
19. Give expressions relating the input, system parameters and output in the time and frequency domains.
20. Why is the response of a system under periodic excitation obtained as a summation of several harmonic responses?
21. What is meant by unit impulse?
22. What is meant by the steady state response of a system?
23. How do you estimate quantization noise?
24. How can you reduce quantization noise in an ADC?
25. Is the frequency response function a property of the system or does it depend on the input as well?
26. Give examples of correlated and uncorrelated signals.

27. What is the Duhamel integral? What is its use?
28. The response of a system under arbitrary force can be found by summing the responses due to several elementary impulses. State whether true or false.
29. What do you mean by higher harmonics?
30. Differentiate between bandpass filters used for machinery fault diagnostics and human ride comfort.
31. What equation would you use for finding the response of a system to random road excitation in the frequency domain?
32. Differentiate between anti-aliasing filter and reconstruction filter.
33. What is meant by impulse response function?
34. What is the advantage of frequency domain computations over those in the time domain?
35. Let $x(t) = 8\cos(2\pi t) + 6\sin(8\pi t) + 4\cos(22\pi t) + 7\sin(66\pi t)$ be sampled for A/D conversion. What sampling frequency would you suggest for this signal?
36. What is the need for windowing in signal processing?
37. What is aliasing? How can it be prevented?
38. What do you mean by a stationary, ergodic random process?
39. What do you mean by Dirac delta function? What are its properties?
40. What advantage do you get by being able to model a random process as stationary and ergodic?
41. How can you obtain the FRF of a system starting from the time domain?
42. How do you define temporal and spatial PSDs?
43. What is a linear time-invariant system?
44. How do you model a linear time-invariant system?
45. How can you obtain the Frequency Response Function of a system in the frequency domain?
46. Given an input description in the time domain, what equation would you use to find the response of a system as a function of time?
47. Differentiate between constant percentage narrow-band analysis and constant bandwidth analysis.
48. If the FRF of a system is known, can its output be found for all types of inputs?
49. What is the difference between deterministic and random signals? Give two examples of each.
50. How can you express a periodic signal in terms of its harmonics?
51. In a particular measurement made using a 10-bit ADC, the signal to noise ratio was found to be 60 dB. How can you increase it to 100 dB?
52. What is the Fourier transform of the signal $x(t) = \delta(t + 0.5) + \delta(t - 0.5)$?
53. The parameters of a system may be found from (a) $h(t)$, (b) $H(j\omega)$, (c) $h(t)$ and $x(t)$, (d) none of the above. Tick appropriate choices. Why?
54. What is an anti-aliasing filter?
55. What are the operations involved in the reconstruction of a sampled signal (i) in the time domain? (ii) in the frequency domain?
56. What is meant by the expected value of a random signal?
57. What class of signals can be represented by the Fourier transform?
58. What does the Fourier transform for modulation state?

59. What does the Fourier transform for convolution state?
60. Define probability density function and cumulative distribution function.
61. Give examples of self-windowing functions.
62. When do you resort to zooming in signal processing?
63. How would you model a random input for frequency domain studies?
64. What is meant by Parseval's theorem?
65. What is the need for averaging in signal processing?
66. What properties of W_N^{kn} are exploited in the computation of the FFT?
67. What laws does the convolution integral obey?
68. What do you mean by the standard normal distribution? How is it different from the Gaussian distribution?
69. What is meant by power spectral density?
70. What is a random process?
71. Give an example of a random process with uniform distribution.
72. How do you find the RMS value of a signal? What is its significance?
73. What does the Wiener-Khinchin relation state?
74. Starting from the power spectral density, is it possible to get the corresponding time record?
75. How are the variance, mean and mean square values of a random signal related?
76. What parameters can be used to characterize a random signal?
77. How can you get the response of an LTI system to transient excitation?
78. Give an example of two correlated variables.
79. How do you define cross-spectral density between two signals?
80. How are the random input to an LTI system system and its random output related?
81. What are the assumptions made in the computation of the FFT?

Explain with Reasons if TRUE or FALSE

82. The mean square value and spectral density function of a random signal are related.
83. For any signal $x(t)$, $R_{xx}(\tau) = R_{xx}(-\tau)$.
84. The joint density function of two random variables is always $p(x, y) = p(x).p(y)$.
85. The mean square value and variance of a signal are always the same.
86. The mean square value of a signal can be found from the time domain, from its probability density function and from its power spectral density.
87. The power spectral density has the phase information intact.

Fill in the Blanks

88. A 12-bit tracking ADC with a full scale voltage of 10 V will show a digital code of _____ corresponding to an input voltage of 2.5 V.
89. If a signal is a rectangular function in one domain, it is a _____ function in the other domain.

90. If the mean value of a Gaussian random acceleration is $4\,g$ and rms value is $9\,g$, then its probability density function can be written as _____.

91. If a discrete-time LSI system is such that the output signal is always identical to the input signal with a unit-sample delay, then the unit impulse response $h[n]$ is _____.

92. Any continuous-time periodic signal can be represented in terms of the _____.

93. The Gaussian distribution is characterized by its _____ and _____ only.

94. If $x(t)$ is a member function of a random process, the value of autocorrelation $R_{xx}(0)$ is _____.

95. The resolution of a 10-bit ADC with full scale voltage of 20 V is _____.

96. Define signal to noise ratio in measurement. It can be increased by _____.

Bibliography

1. Alexander, W., & Williams, C. M. (2017). *Digital signal processing: Principles, algorithms and system design*. Academic Press.
2. Bendat, J. S. (1993). *Engineering applications of correlation and spectral analysis*. New York: Wiley-Interscience.
3. Bendat, J. S., & Piersol, A. G. (2000). *Random data: Analysis and measurement procedures*. New York: Wiley-Interscience.
4. Bertsekas, D. P. (2002). *Introduction to probability*. Belmont: Athena Scientific.
5. Bhattacharyya, S. S., Deprettere, E. F., Leupers, R., & Takala, J. (2019). *Handbook of signal processing systems*. Springer International Publishing.
6. Billingsley, P. (1979). *Probability and measure*. New York: Wiley.
7. Blandford, D.,& Parr, J. (2012). *Introduction to digital signal processing*. Pearson.
8. Carlson, G. E. (1998). *Signal and linear system analysis*. New York: Wiley.
9. Chaparro, L., & Akan, A. (2018). *Signals and systems using MATLAB*. Academic Press.
10. Chung, K. L. (2000). *A course in probability theory revised*. Boston: Academic Press.
11. Crandall, S. H., & Mark, W. D. (1963). *Random vibration in mechanical systems*. New York: Academic Press.
12. Damelin, S. B., & Miller, W. Jr. (2011). *The mathematics of signal processing (Cambridge texts in applied mathematics)*. Cambridge University Press.
13. DeGroot, M. H., & Schervish, M. J. (2001). *Probability and statistics*. Boston: Addison Wesley.
14. Elishakoff, I. (1999). *Probabilistic theory of structures*. New York: Dover Publications.
15. Feller, W. (1991). *An introduction to probability theory and its applications* (Vol. 2). New York: Wiley.
16. Gallager, R. G. (1996). *Discrete stochastic processes*. Boston: Kluwer Academic.
17. Garcia, A. L. (1994). *Probability, statistics and random processes for electrical engineering*. Reading: Addison-Wesley.
18. Giron-Sierra, J. M. (2016). *Digital signal processing with MATLAB examples, volume 1: Signals and data, filtering, non-stationary signals, modulation (Signals and communication technology)*. Springer.
19. Harris, S. L., & Schilling, R. A. (2013). *Introduction to digital signal processing using MATLAB*. Cengage.
20. Haykin, S., & Kosko, B. (2001). *Intelligent signal processing*. New York: IEEE Computer Society Press.
21. Hsu, H. P. (1993). *Signals and systems, Schaum's outline series*. New York: McGraw-Hill.

22. Hsu, H. P. (1997). *Schaum's outline of probability, random variables and random processes*. New York: McGraw-Hill.
23. Ibrahim, R. A. (2008). *Parametric random vibration*. New York: Dover Publications.
24. Karlin, S., & Taylor, H. M. (1975). *A first course in stochastic processes*. New York: Academic Press.
25. Karlin, S., & Taylor, H. M. (1981). *A second course in stochastic processes*. New York: Academic Press.
26. Kehtarnavaz, N., Sehgal, A., & Parris, S. (2018). *Smartphone-based real-time digital signal processing: (Synthesis lectures on signal processing)* (2nd ed.). Morgan & Claypool.
27. Kolmogorov, A. N. (1956). *Foundations of the theory of probability*. New York: Chelsea Publishing Company.
28. Lathi, B. P. (2005). *Linear systems and signals*. USA: Oxford University Press.
29. Lathi, B.P. and Green, R.A., Essentials of Digital Signal Processing, Cambridge University Press, 2014.
30. Leis, J. W. (2012). *Digital signal processing using MATLAB for students and researchers*. Wiley.
31. Lin, Y. K. (1976). *Probabilistic theory of structural dynamics*. New York: Krieger.
32. Lindner, D. K. (1999). *Introduction to signals and systems*. Asia: McGraw-Hill Education.
33. Lipschutz, S., & Lipson, M. (2000). *Schaum's outline of theory and problems of probability*. New York: McGraw-Hill.
34. Lutes, L. D., & Sarkani, S. (2004). *Random vibrations: Analysis of structural and mechanical systems*. Burlington: Elsevier Butterworth-Heinemann.
35. Lyons, R. G. (1997). *Understanding digital signal processing*. Reading: Addison-Wesley.
36. Lyons, R. G., & Fugal, D. L. (2014). *The essential guide to digital signal processing (Essential guide series)*. Prentice Hall.
37. Manolakis, D. G., & Ingle, V. K. (2011). *Applied digital signal processing: Theory and practice*. Cambridge University Press.
38. McMlellan, J. H., Schafer, R., & Yoder, M. (2016). *Digital signal processing*, First Global Edition. Pearson.
39. Miller, S. L., & Childers, D. G. (2004). *Probability and random processes: With applications to signal processing and communications*. Amsterdam: Elsevier Science.
40. Mitchell, L. D. (1985). Signal processing and the fast Fourier transform (FFT) analyzer. *Experimental Techniques*, *9*, 3s–15s (ISSN 0732-8818).
41. Newland, D. E. (2005). *An introduction to random vibrations, spectral & wavelet analysis*. New York: Dover Publications.
42. Nigam, N. C. (1983). *Introduction to random vibrations*. Cambridge: MIT Press.
43. Nigam, N. C., & Narayanan, S. (1994). *Applications of random vibrations*. New Delhi: Narosa Publishing House.
44. Oppenheim, A. V., Hamid, S., & Willsky, A. S. (1996). *Signals and systems*. Englewood Cliffs: Prentice-Hall.
45. Oppenheim, A. V., & Schafer, R. W. (1999). *Discrete-time signal processing*. Upper Saddle River: Prentice-Hall.
46. Papoulis, A., & Pillai, S. U. (2002). *Probability, random variables and stochastic processes*. New York: McGraw-Hill Science/Engineering/Math.
47. Parker, M. (2010). *Digital signal processing 101: Everything you need to know to get started*. Newnes.
48. Preumont, A. (1994). *Random vibration and spectral analysis*. Boston: Kluwer.
49. Proakis, J. G., & Manolakis, D. G. (1996). *Digital signal processing: Principles, algorithms, and applications*. Englewood Cliffs: Prentice-Hall.
50. Rabiner, L. R., & Gold, B. (1975). *Theory and application of digital signal processing*. Englewood Cliffs: Prentice-Hall.
51. Rao, R. P. (2008). *Signals and systems*. New Delhi: Tata McGraw Hill Publishing Company Limited.
52. Rawat, T. K. (2014). *Digital signal processing*. Oxford University Press.

53. Roberts, J. B., & Spanos, P. D. (2003). *Random vibration and statistical linearization*. New York: Dover Publications.
54. Robson, R. D. (1963). *An introduction to random vibration*. Amsterdam: Edinburgh University Press, Elsevier Publishing Company.
55. Ross, S. M. (2005). *First course in probability*. Upper Saddle River: Prentice Hall.
56. Ross, S. M. (2008). *Stochastic processes*. New York: Wiley.
57. Rossi, G. B. (2014). *A probabilistic theory of measurement with applications*. Netherlands: Springer.
58. Sircar, P. (2016). *Mathematical aspects of signal processing*. Cambridge University Press.
59. Smith, S. W., The scientist and engineer's guide to digital signal processing. www.dspguide.com.
60. Solnes, J. (1997). *Theory of stochastic processes and random vibrations*. New York: Wiley.
61. Soong, T. T., & Grigoriu, M. (1993). *Random vibration of mechanical and structural systems*. Englewood Cliffs: Prentice Hall.
62. Spiegel, M. R., & Stephens, L. (2007). *Schaum's outline of statistics (Schaum's outline series)*. New York: McGraw-Hill.
63. Spiegel, M. R., Schiller, J. J., & Srinivasan, R. A. (2000). *Schaum's outline of probability and statistics*. New York: McGraw-Hill.
64. Stanley, W. D., Dougherty, G. R., & Dougherty, R. (1984). *Digital signal processing*. Virginia: Reston.
65. Stark, H., & Woods, J. (2011). *Probability, statistics, and random processes for engineers*. Pearson.
66. Tocci, R. J. (1988). *Digital systems: Principles and applications*. New Delhi: Prentice Hall of India.
67. Vetterli, M., Kovacevic, J., & Goyal, V. K. (2014). *Foundations of signal processing*. Cambridge University Press.
68. Wirsching, P. H., Paez, T. L., & Ortiz, K. (2006). *Random vibrations: Theory and practice*. New York: Dover Publications.
69. Yang, C. Y. (1986). *Random vibrations of structures*. New York: Wiley Press.
70. Ziemer, R. E., Tranter, W. H., & Fannin, R. (2002). *Signals and systems: Continuous and discrete*. New Delhi: Prentice-hall of India Pvt. Ltd.

Chapter 9
Basics of Experimental Modal Analysis

DID YOU KNOW that it was Robert Hooke, who through his pioneering experiments, was the first to observe the nodal patterns associated with the vibrations of glass plates? He ran a violin bow along the edge of a plate covered with flour to see the nodal patterns emerge. Chladni repeated Hooke's experiments and published in 1787 his findings on drawing a bow over a piece of metal whose surface was covered with sand. The plate was bowed till it reached resonance, causing the sand to form nodal lines or lines of zero vibration. The patterns formed by these lines are what are now called Chladni figures. In 1802, while studying the vibration of plates, Chladni came up with the method of sprinkling sand on a rigid vibrating plate to find its mode shapes in various modes of vibration. In 1809, he demonstrated the beauty and intricacy of the modal patterns of the vibrating plates in the French Academy. Napoléon Bonaparte was so impressed that he presented a sum of 3,000 Francs to the academy to be awarded to the first person who came up with a satisfactory mathematical theory of the vibration of plates.

9.1   Introduction

Experimental modal analysis deals with the determination of modal parameters such as natural frequencies, damping properties, and mode shapes of a structure through experiments. This was once considered an area requiring expert knowledge, but with the advent of personal computers and the development of inexpensive, user-friendly software packages for modal parameter extraction, as well as signal processing software and hardware, it has evolved into a common tool readily available in vibration toolboxes and accessible to most test engineers. In any experimental modal analysis procedure, modal parameters are to be estimated from the measured frequency

response functions (FRFs), i.e., from the output response and input force data and the quality of an experimental modal model is only as good as the quality of the FRFs. Hence, several important experimental aspects are to be considered while conducting modal tests in order to obtain valid modal data. Choice of excitation of the structure, in terms of excitation location and mechanism, excitation signal, frequency range and amplitude of the excitation force, fixing/mounting of test structure so as to minimize exciter/test structure interactions are to be looked into. Besides, selection of appropriate transducers and their positioning and mounting, as well as aspects related to signal processing, have to be considered and are described in this chapter.

Though the entire process of obtaining the required data is experimental in nature, many mathematical techniques are involved in the computation of FRFs and modal parameter extraction. Experimental modal analysis is, therefore, a difficult application, especially where high accuracy is required; besides an in-depth knowledge of structural dynamics is also required. Therefore, mathematical theory, as found necessary, has been presented in this chapter to obtain a logical development of these techniques. Starting with the basics of the representation and properties of various forms of FRFs of a single-degree-of-freedom (SDOF) linear dynamic system, we progress to FRF representation of multi-degree-of-freedom (MDOF) systems, since this is how most physical structures are modelled. The discussion of undamped cases is followed by those with viscous and structural damping. Finally, parameter estimation techniques are presented to gain insight, first into the single-mode concept, followed by concepts related to MDOF systems. Modal testing, with its structural dynamics and modal parameter estimation theories, is based on an ideal physical model, i.e., the system is linear, time-invariant, causal, and observable. However, most structures deviate from this ideal behaviour and the resulting problems have to be anticipated and dealt with in any modal test setup. It is hoped that the material presented here will give the reader a sound understanding of modal analysis concepts so as to enable him/her to confidently embark on a modal analysis exercise.

9.2 Important Experimental Aspects of Modal Testing

DID YOU KNOW that it was only in the 1960s that the modern era of experimental modal analysis began? This was the decade when measured force and motion could be accurately recorded. Many technologies that were developed earlier in the 1900s matured and became somewhat integrated by the 1960s. The theoretical background of modern modal test methods was developed in the 1930s, 1940s, and 1950s, and was well established in the literature by 1963. The development of the tracking filter during the late 1950s was a major technical development that pushed nascent experimental modal analysis methods into the modern era. Besides, the experimental modal analysis could take off only after sensors were readily available and sufficiently stable and accurate to measure

force and vibration response. By the 1960s, sensors were commercially available and well accepted for vibration testing. With the development of the bondable strain gauge, modern strain-based load cells and accelerometers for measuring force and motion were possible; later piezoelectric load cells and accelerometers came into the market. Finally, the modern era of experimental modal analysis began when equipment was readily available to measure the force and response captured. This required equipment capable of measuring both magnitude and phase or the real and imaginary parts of harmonic signals. With the development of dynamic analyzers, tracking filters, and transfer function analyzers, commercial equipment was finally available.

Some of the important practical aspects to be considered in modal analysis are the boundary conditions of the test structure, minimization of exciter/test structure interaction, choice of exciters/shakers, problems in the measurement of excitation force, difficulties encountered in impact testing, sensing techniques and fixtures, as well as the selection of excitation signals for modal testing. These have been discussed in the sections to follow.

9.2.1 Support Conditions of Test Structure

For structures such as buildings, towers, machinery, vehicles, etc., modal testing is done to evaluate system dynamics, or from the point of view of redesign or modification of actual fixing conditions in order to improve their vibration isolation characteristics or fatigue lives. The fixture location and the point of application of excitation force (either through hammer or shaker) require careful consideration. The support and fixing conditions used for a test structure largely dictate the success of a modal test. Often, the fixture can distort the test input so badly that the specimen is overtested at some frequencies and undertested at others, causing the results to be meaningless. Hence, any basic test setup used for modal analysis should be carefully planned, taking into account the type of the structure to be tested and the level of accuracy of the desired results. It is important to find out all the fixity conditions in which the structure may be used under actual operating conditions and choose the appropriate fixtures required to obtain the desired support constraints or boundary conditions. If the structure is not properly prepared, then there is the degradation of the overall structural characteristics obtained. This affects subsequent analyses such as structural modification, finite element correlation, and substructure coupling.

The main decision to be taken is whether to conduct the test in the (i) free condition or (ii) fixed/grounded condition or (iii) actual fixity conditions at site. Analytically, the first two boundary conditions can be achieved very easily by either allowing all degrees-of- freedom for the first case, or arresting the DOF at the support for the second case. However, it is not so easy to simulate these conditions while testing. Free condition implies that the structure is not attached to the ground anywhere and is freely suspended or floating in space. With such a condition, the structure

exhibits rigid body behaviour ideally at low frequencies close to zero frequency and does not undergo any bending or flexing. The aeroplane or rocket in flight is an example of this free condition. In practice, however, the rigid body frequencies (three translational and three rotational) will not be zero. Testing with this boundary condition, therefore, allows determination of the rigid body modes and hence the mass and inertia properties of the structure. Physically, achieving a truly free support is not realizable, so the structure must be held in some manner. This may be achieved by suspending the structure using very soft, highly elastic bands or suspension cords or by placing the test structure on very soft sponge, cushion, springs or specially designed suspensions. Due to such an arrangement, the structure will be constrained to some extent and the rigid body modes will, therefore, have low frequencies. If a sufficiently resilient support is used, the frequencies corresponding to the rigid body modes will be quite low as compared to the frequencies of the bending modes and will thus have negligible influence. As a guideline in testing with free supports, the frequency of the highest rigid body mode should be less than one tenth that of the first flexible mode. The excitation frequency band during modal testing should be carefully chosen to avoid exciting the rigid body modes which can cause failure of the suspension cords.

A fixed boundary condition implies that all motion: displacements as well as rotations, at the clamped edge, or at the support, are forced to zero. This condition is much more difficult to achieve than a free boundary condition, since most structures have some amount of flexibility at the clamps. The base to which the structure is attached will typically have some motion of its own due to flexibility of the bolted, riveted or welded connections. To ascertain whether this motion is negligible or not, FRFs may be measured at the base over the frequency range of interest and compared with the corresponding response of the structure. If the former is negligibly small, the structural data may be taken to be accurate. Figure 9.1 shows free and fixed boundary conditions for a gear wheel and beam, respectively.

From a practical point of view, the actual boundary conditions chosen for a test may depend on the weight and size of the structure. For instance, it is not possible to support a large structure weighing a few tonnes in a free test state. Also, it may

(a)   (b)

Fig. 9.1 Boundary conditions in modal testing: **a** free, **b** fixed

not be convenient to fix certain structures in the clamped condition. There are also situations requiring both test conditions as in the case of a space vehicle. It is desirable to test this structure in a free condition to simulate its operating environment in space. However, it is also required to do the test in the grounded condition to study dynamic behaviour in the launch environment. Sometimes the boundary conditions can lead to uncertainties and non-linear behaviour of the test structures, and therefore, fixing conditions should be carefully considered. Besides, it is to be remembered that the clamping mechanism also adds damping to the structure due to movement in the fixture joints.

9.2.1.1   Minimizing Exciter/Test Structure Interaction

The shaker physically interacts with the structure through the force transducer, altering the system dynamics. In the case of lightweight structures, the load cell adapter may lead to considerable mass loading, causing the measured force to be greater than that actually applied. Since this extra mass is between the structure and the load cell, the latter senses it as a part of the structure. The shaker should transmit forces in line with the main axis of the load cell. However, it is possible that there is rotation about the other two axes. To minimize this problem, the exciter is connected to the load cell through a slender rod called the stinger, which allows the load to move freely in other directions. This stinger is essentially a rod with a strong axial stiffness like a truss, but weak bending and shear stiffness. Figure 3.28 shows the positioning of the stinger in modal testing. It is also desirable to isolate the main body of the exciter from the structure to prevent reaction forces from being transmitted back to the structure. There are a few ways in which this could be done:

 (i) By supporting the shaker on a mechanically isolated foundation.
 (ii) By mounting the shaker on a solid floor with the structure suspended from above.
(iii) By suspending the shaker and using an inertial mass attached to the shaker to generate a measurable force.

The concept of impedance is useful in understanding the interaction between the shaker and the test specimen. Mechanical impedance, in simple terms, is defined as the force required to produce a desired motion (velocity to be precise). A force applied to a test structure by means of an exciter drops close to zero at the natural frequency when the vibration amplitude becomes high. Such reduction in force is due to the fact that, at resonance, the test structure becomes highly compliant and vibrates with large amplitudes, even with very low input forces. The exciter then uses all the available energy to accelerate its own mechanical parts, leaving very little force for driving the test structure. The signal level of the force may, in fact, drop to the noise floor in the instrumentation. Complex systems, with several natural frequencies, load the vibration exciter each time the excitation frequency coincides with a natural frequency. The lighter the mass of the moving elements in the exciter, the less this problem will be. An ideal exciter is one with infinite mechanical impedance; however, this is not the case in practice. This problem may also be alleviated by effective

mechanical de-coupling of the structure under test and the modal exciter, through appropriate test fixtures as discussed above, minimizing the change in mechanical impedance caused by the exciter. This is required so that when modal parameters are later extracted from the vibration data around resonance, force drop-offs are reduced, thereby increasing the accuracy of the FRFs.

If the input is not a force applied to the mass being tested, but instead is a motion (displacement) applied to the base, the result is that instead of the zero force required to maintain motion, an oscillating force of specific magnitude is needed to maintain the motion of the test mass. Therefore, base excitation of a system exhibits high impedance. In the limit, if the test mass approaches infinity, the impedance also approaches infinity.

It is clear that the concept of infinite impedance is inherent in any modal test program, so that the dynamic input to the specimen is not altered by the specimen itself. The design of the fixture is directly tied to the need for a shaker with infinite impedance. The fixture must be as stiff as possible, so that it is not deflected by the force and transfers motion from the exciter to the specimen with zero distortion at all amplitudes and frequencies. Generally, however, a fixture is never stiff enough to result in the natural frequency of the test system being above the range of test frequencies. Instead, the natural frequency generally lies within the test frequency range. Test specifications typically require that the dynamic input to a test specimen be kept at a constant level, regardless of fixture loading. Electromagnetic and electrohydraulic vibration exciters handle this problem with servo-controllers. With other exciters, the modal analyst has to take all precautions. To reduce the impedance mismatch between the shaker and the structure, the following methods may be adopted:

 (i) Using shakers with different coil sizes depending on the structure.
(ii) Using a shaker with a constant current amplifier and positioning the shaker at a point with large effective mass.

9.2.2 Choice of Exciters/Shakers

There are many exciters that may be used for experimental modal analysis. They include hydraulic shakers, electrodynamics shakers, unbalanced mass exciters, inertial shakers, etc.; with such exciters, the specimen is fixed onto the shaker table and tested for identifying resonance behaviour. In the case of unbalanced mass exciters, the force produced is proportional to the square of the rotational speed of the unbalanced mass. Since these exciters are capable of producing large forces, they are used to test structures like bridges, masts, towers, dams, foundations, etc., the shaker itself offering negligible mass loading in these cases. For imparting forces to lightweight structures, or for modal analysis of machines at site, the impact hammer is a very useful device. The difficulties involved in a hand-controlled hammer are occurrences of multiple impacts and poor repeatability of impact conditions, namely the force amplitude and direction. These problems can be got rid of to a large extent by using

electric hammers. An interesting development in the field of exciters is that of inertial shakers (reaction mass shakers) and non-contact electromagnetic exciters for rotating machines. The latter are particularly convenient for exciting lightweight and rotating structures. They do not have the problems associated with impact testing using a hammer. They can be used to transmit forces to rotating discs, as in vibration control or for estimation of the stiffness and damping of bearings. These exciters are also useful for evaluating the dynamic response of impellers, circular saws, etc. Inertial shakers are useful for obtaining the response of a structure in field conditions within the shortest test time possible, without the use of any additional test fixtures. They can very conveniently be used for modal testing of a ship hull or automobile body, which requires fixing the exciter with arbitrary orientations. Chapter 3 has dealt exhaustively with excitation mechanisms with Tables 3.4 and 3.6 summarizing salient features of intrusive and non-intrusive excitation techniques, respectively.

9.2.2.1 Problems in the Measurement of Excitation Force

Measurement of the input force is one of the most crucial aspects of a successful modal test. Errors originating here will not only affect each and every FRF obtained, but will also typically be 'hidden', and unlike those in the measured response, be difficult to detect visually. Obtaining accurate and reliable force measurements is therefore very important in order to ensure that the extracted modal parameters are correct. With a conventional force transducer used in a modal test setup, the measured force will not be directly proportional to and in phase with the excitation drive signal from the signal generator due to the masses of the moving parts of the exciter, as well as the mass of force transducer, which are not negligible. If, however, the moving parts of the excitation drive train are considered to be part of the structure being tested, then the force can be indirectly computed by sensing the current to the electrodynamic shaker as discussed in Sect. 3.10.1.1. The forces measured may also be erroneous due to the asymmetric construction of the force transducer and the influence of lateral degrees-of-freedom of motion at the contact point of the transducer and test structure. Mass cancellation circuits may also be used for this.

9.2.3 Impact Testing and Difficulties

Impact testing is probably the most popular excitation technique used in experimental modal analysis. Its popularity is due largely to its ease of use, ability to adapt to difficult testing environments, and self-windowing potential for minimal signal processing errors. All these aspects have been described in Sect. 3.11.4. All modal parameter extraction methods used for obtaining FRFs with shaker excitation have been adapted to impact testing also. Impacts may be imparted using an ordinary instrumented hammer or a modal punch or an electric hammer. The challenge in impact testing of large structures is in supplying an impact with sufficient energy

to uniformly excite the test structure. Since most of the dynamic characteristics of interest in large structures are at low frequencies, a means of restricting the input force spectrum is an important consideration. Electrodynamic shakers can also be used to deliver impulsive loads to test objects by feeding them with voltage pulses. This method has conventionally been used for shock qualification of components; however, there is no reason why such shocks cannot be used in modal testing.

One major problem with impact testing is the generation of multiple impacts while hitting. This can be eliminated by the use of a force window which is nothing but a rectangular window. It suppresses the second and subsequent pulses, if any, in the measured force that is used in the FRF computation. However, the response to the suppressed pulses exists in the output spectrum used in the FRF calculation, as a result of which the computed FRF does not represent the system accurately. Therefore, multiple impacts should be avoided. In practice, it is usually easier to simply discard a record with a double hit from the ensemble of records and this is done by most modal software packages. If a few averages are taken for an impact measurement, even one unsatisfactory record can adversely affect the averaged set of measurements.

Another problem in impact testing involves the use of non-linear structures. If impact forces are high, they can overdrive the system and aggravate its non-linear response. Every effort should be made to linearize the system when making FRF measurements using impact testing for estimation of modal parameters, since all the theory behind FRF computation and modal parameter extraction is based on the assumption of a linear time-invariant system. Non-linearities should be eliminated from the test system as far as possible. Static preloads applied with soft springs can tighten clearances and constrain the system into a comparatively linear state. The low stiffness of preload in series with the stiffness of the structure causes a negligible effect on the system dynamics.

Yet another problem is impacting at locations which are not accessible with a conventional impact hammer. A punch impactor may be used under such circumstances to precisely locate impacts, to minimize the variance of the impact location, and to impact in a skewed direction (non-orthogonal to the coordinates). The punch impactor should be fitted with a hard tip, typically a metal or plastic tip. It should be calibrated in the configuration in which it is to be tested, since the sensitivity depends on the tip-structure combination.

9.2.4   Sensing Techniques

In Chap. 3, we discussed various types of transducers which may be used in vibration testing. Displacement, velocity or acceleration transducers should be used, depending on whether we require receptance, mobility or acceleration FRFs (described in Sect. 9.3.1). Mass loading due to the sensor could be a severe problem for highly flexible structures, especially if the sensor mass is considerable compared to the effective mass of a particular mode of the structure. Mass loading can also be reduced by

choosing lightweight transducers or even better, non-intrusive sensors which cause minimal mass loading on the structure. Details of non-intrusive sensors are given in Sect. 3.8. There are many transducers which are suitable for non-contact vibration response measurement and many of these have the desirable small target footprints. In modal analysis, the use of miniature accelerometers and strain gauges for sensing the response is popular; they are not truly non-intrusive, but they hardly alter the system dynamics. Many of the non-contact response measurement techniques allowing direct interpretation of experimental data are optical in nature and have very good resolution, dynamic range and flexibility necessary for modal analysis applications. Some of these are fibre optic sensors, laser holography, laser Doppler vibrometer (LDV), laser triangulation techniques, and line scan cameras. There are other non-optical techniques using proximity probes of both inductive and capacitive type and acoustic sensors. A lot of information is given in Chap. 3 regarding a whole range of transducers of the contact type, which may also be used for modal analysis after taking into account their salient features.

Considerations regarding the choice, location, and orientation of transducer and transducer mounting have to be looked into carefully. Local conditions of a structure may drastically vary, influencing the local stiffness at the transducer location and hence choice of location is very important. The transducer should not be mounted at or close to a node of any structural mode of interest. For point FRF measurements, the force and response transducers should be at the same point. This may be done using an impedance head, which has both force gauge and the accelerometer combined in a single housing. It is important to note the direction of the force and response, and when at a location both translational and rotational DOFs are present, the errors in the mobility ratio can be considerable. Most of the present modal analysis techniques use translational DOF. Bi-and triaxial transducers are also available to perform detailed modal analysis.

Due to mass loading effects of a contact type of transducer, the vibration measured turns out to be different from what would be produced in the absence of loading. Fortunately, technology exists for performing effective non-contact modal analysis of such structures, but careful consideration must be given to the entire test system compatibility. For example, consider the read/write head and assembly on a typical computer hard disc, the dynamic response of which is critical to the performance of the system. The mass of the entire read/write mechanism being quite small (approximately 30–40 gm), the mounting of either an excitation or response transducer can alter the system dynamics. The mass loading of even the lightest accelerometer (less than a gram) can significantly alter the system response.

9.2.5 Selection of Excitation Signals for Modal Testing

The choice of excitation system, as well as excitation function, can make the difference between a good measurement and a bad one. Once the excitation system, i.e., the shaker or excitation mechanism that best suits the application has been decided

on, the next step in the measurement process involves selecting an excitation function
such as sine sweep, random, transient, etc. The excitation signal generally depends
on the type of test to be conducted. Though the level of non-linearities can be mea-
sured and characterized effectively with sine sweeps, a random function is most
suitable to estimate the best linear model of a non-linear system. The amount of
damping and the density of the modes within the structure can also dictate the use of
specific excitation functions. The exciter control electronics must therefore have a
signal generator and a gain control, dependent on the feedback from the test object.
Computer-aided vibration control systems which are simple to operate and interac-
tive are commercially available for shaker control. They utilize highly developed and
powerful digital signal processing hardware and/or software. The excitation signal
can be one of the following:

  (i) stepped sinusoidal
 (ii) slow sinusoidal sweep
(iii) periodic
(iv) true random

- sine on random
- random on random

 (v) pseudo-random
(vi) transient

- chirp random
- sine chirp
- burst chirp
- external pulse
- shock

Most of the present-day analyzers come built in with many/all the options mentioned
above. Stepped and swept sinusoidal, as well as periodic and random excitation
signals, can also be got from an oscillator or signal/function generator. The latter
signal can also be got from a random noise generator.

9.2.5.1 Stepped Sine Excitation

In this method, the signal given to the exciter is a discrete sinusoidal voltage or current
with a fixed amplitude and frequency. In order to encompass the entire frequency
range of interest, the command signal frequency is stepped from one discrete value
to another in such a way as to provide the necessary frequency resolution in the FRF
plot. Care should be taken to ensure that the excitation frequency is not changed
abruptly and that sufficient time is given for the system to settle down to steady state
condition. Especially if one is in the vicinity of resonance, or if one is measuring
lightly damped or closely spaced modes, care should be taken to avoid transient

effects by delaying each measurement and ensuring that steady state conditions have been attained before the readings are taken.

The major advantage with stepped sine excitation is that it gives the vibration engineer flexibility in choosing the frequency resolution. Near resonance, tests can be conducted at a larger number of frequencies and away from resonance, excitations can be given at a few frequencies only. It is especially important to have closely spaced points around resonances if one is interested in measuring modal damping. Considering that excitation is always imparted at a single well-known frequency, it is possible to use simple measuring techniques for obtaining FRFs and doing away with the need for sophisticated equipment like Fast Fourier transform (FFT) analyzers. It is possible to optimize the measurement process, irrespective of whether one is doing the whole process manually or is using computer control. Though sine testing gives the best signal-to-noise ratio (SNR), it is extremely slow for broadband measurements.

Stepped sine excitation is ideally suited for the evaluation of non-linearities since it is very easy to control the amplitude, frequency, and phase of the excitations. The presence of non-linearities in the system may be found out by analyzing the dependence of the response level on the excitation level and the manifestation of higher order harmonic components in response to sinusoidal excitation of a single frequency. This testing technique is typically used for large, complex structures and for cases where there is a high level of background noise present. For large structures requiring a multiple excitation/response measurement, one has to use a controller which defines a phase/amplitude relationship between the different shakers. This relationship has to be maintained throughout the entire sweep. This method of testing is called the normal mode method.

9.2.5.2 Slow Sine Sweep

This involves the use of a sweep oscillator to provide a sinusoidal command signal, the frequency of which is varied continuously and slowly throughout the frequency range of interest. Swept sine capability allows conventional sine testing of non-linear structures where high energy at a single frequency is necessary. Figure 9.2 shows a typical sine sweep signal.

As in the case of stepped sine excitation, it is imperative in this case also to ensure that the frequency sweep is done sufficiently slowly to enable the system to settle down to steady state conditions before measurements are made. If the sweep rate is too high, the FRF gets distorted and the true resonance peak gets shifted to the right (higher frequency) with increasing sweep rate. A good way to check whether a sweep rate is alright, is to measure twice, once sweeping upwards, and the second time downwards through the frequency range. It can be said that the sweep rate is correct, if both the sweeps give rise to the same FRF. It is possible to prescribe an optimum sweep rate for a given structure, taking into account its damping levels. A good practical approach to guarantee that we can very closely approach the desired condition is by using a logarithmic sweep rate. The ISO standards on 'methods
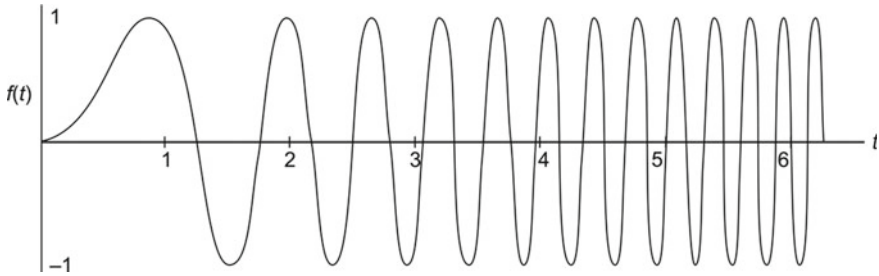
Fig. 9.2 Sine sweep

for the experimental determination of mechanical mobility' prescribe the maximum allowable linear and logarithmic sweep rates through a resonance. These are given by the formulae below.

Linear sweep

$$S_{\max} < 54(f_r)^2(\xi_r)^2 \cdot \text{Hz/minute} \tag{9.1a}$$

or

$$S_{\max} < 216(f_r)^2(\zeta_r)^2 \text{ Hz/minute} \tag{9.1b}$$

Logarithmic sweep

$$S_{\max} < 78(f_r)^2(\xi_r)^2 \text{ octaves/minute} \tag{9.2a}$$

or

$$S_{\max} < 310(f_r)^2(\zeta_r)^2 \text{ octaves/minute} \tag{9.2b}$$

Here $S_{\max}$ denotes the maximum allowable sweep rate, f_r the resonant frequency of the mode under study, ζ_r its damping ratio, and ξ_r the corresponding structural damping loss factor.

9.2.5.3 Periodic Excitation

A complex periodic input signal is sometimes used for obtaining FRFs. Such a signal contains not one, but a superposition of several sinusoidal signals simultaneously. With such an excitation signal, a spectrum analyzer can extract the response to each of these sinusoidal components simultaneously. Two types of periodic signals are typically used; both types are generated within most analyzers in order to ensure perfect synchronization with the analysis. One is a synthesized signal with various sinusoidal components having ordered amplitude, frequency, and phase relationships, as in a square wave. The other is of the pseudo-random type, which is discussed later.

9.2.5.4 True Random Excitation

True random signal is the most popular excitation signal for shakers since the measurement time is minimal. The choice of this excitation signal can easily be imposed by the dynamics of the test structure. This signal may be obtained from signal generators or random noise generators, which are commercially available. Several types of noise signals are typically used, namely uniform white noise, Gaussian white noise, pink noise, and periodic random noise. Ideal white noise has equal power per unit bandwidth, resulting in a flat power spectral density (PSD) across the frequency range of interest. Pink noise, also called weighted white noise, on the other hand, contains equal energy per measurement band and thus has an octave or one-third octave band energy level which is constant across frequency bands with the lower frequencies having a higher energy level. In practical measurements involving white noise, an infinite number of samples are required for achieving a flat power spectral density. Hence, when making measurements with white noise excitation, the power spectra are usually averaged; the larger the number of averages, the flatter is the resulting power spectrum.

The terms uniform and Gaussian refer to the nature of the probability density function (PDF) of the noise. For uniform white noise, the PDF curve is flat, while for Gaussian white noise, the PDF is bell-shaped. Figure 9.3 shows the time record of white noise. Figures 8.37a and 8.38 show the PDFs of uniform white noise and Gaussian white noise, respectively. Gaussian white noise may be obtained by passing uniform white noise through a linear filter.

This signal has the advantage that both its peak-to-root-mean-square ratio (crest factor) and root mean square (RMS) value can be controlled. It also has a reasonably good signal-to-noise ratio (SNR). A random function is most suitable for obtaining the best linear model of a non-linear system. Typically in testing, the true random
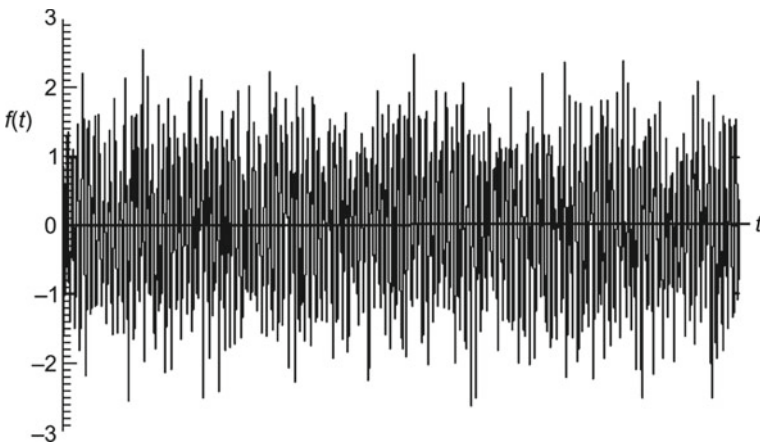
Fig. 9.3 Time record of white noise
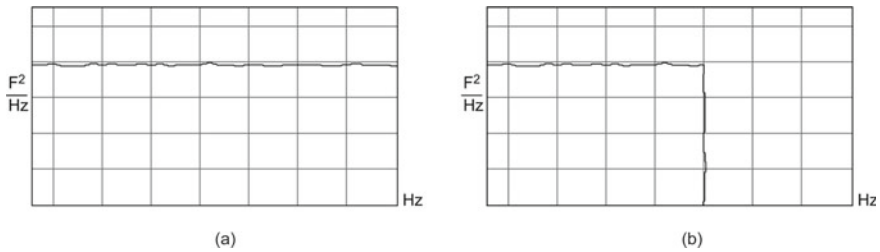
Fig. 9.4 PSD of random noise: **a** white noise, **b** band-limited white noise

excitation signal is sent out during a time period T to the shaker by the controller and the structure's response is measured synchronously. During the next time period T, a different random excitation record (with the same RMS level and crest factor, however) is sent out. Hence, the excitation level at a given frequency is different during each period T, allowing the test engineer to get the best linear approximation of a non-linear system by using an averaging process. Though averaging helps in determining the best linear model and in reducing uncorrelated noise, it increases the total analysis time. This technique also has the disadvantage that there is leakage in the analysis window due to the non-periodic nature of the signal in this window. This problem can be overcome by using pseudo-random excitation which is periodic in the analyzer window considered. Band-limited random noise allows exciter energy to be concentrated in a small frequency range for improved SNR. We have already seen the time history of true random noise (Fig. 9.3). Figure 9.4 shows the PSDs of true random (white noise) and band-limited white noise force signals.

9.2.5.5 Pseudo-Random Excitation

Pseudo- or periodic random noise (PRN) consists of a summation of all possible sinusoids with frequencies that can be represented as a fundamental frequency and higher harmonics in the number of samples being processed. PRN is a superposition of sinusoidal signals with equal energy at all frequencies and with random phases. The major advantage of this type of excitation signal is its exact periodicity in the analyzer bandwidth, resulting in zero leakage errors. PRN is thus self-windowing and there is no need to do windowing before spectral analysis. The main difference between PRN and white noise is that PRN does not have a continuous PSD curve, as does white noise, but has energy only at discrete frequencies corresponding to harmonics of a fundamental frequency. This fundamental frequency is equal to the sampling frequency divided by the number of samples. The most favourable feature of PRN is that the FRF of a linear system may be computed with just a single time record, instead of averaging a number of FRFs over several time records, as is done for non-periodic random noise sources. Figure 9.5 shows the PSD of force in the form of PRN.
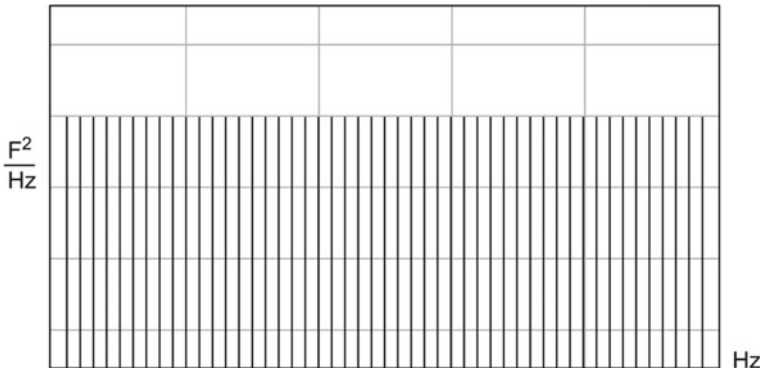
Fig. 9.5 PSD of periodic random force

9.2.5.6 Transient Excitation

Transient excitation is different from steady-state excitation, in that, it poses the constraint that the response signal should have died by the end of the sampling window. This type of excitation is essentially of short duration and requires that lightly damped structures have a rather long sampling window, limiting the frequency range that can be studied. With transient excitation, as in the case of random excitation, it is the practice to make a large number of repeated measurements under nominally identical conditions, and then to do averaging, resulting in averaged FRF estimates. Though a large number of averages tremendously improves the SNR, it may diminish any advantage of the time taken for analysis, which is a potential attraction of the transient excitation technique. Most analyzers provide the user with a transient excitation signal called the chirp signal which is nothing but a rapid swept sine excitation from the lowest to the highest frequency in the spectrum over the relatively short sampling time window. This signal is so-called because it sounds like the chirping of a bird. Chirp signals encompass transient signals such as chirp random, sine chirp and burst chirp. A burst chirp signal is similar to the chirp, except that it is turned off before the end of the sampling window. Chirp signals, like all the previously mentioned excitation signals, can be fed to an attached shaker. We have already seen in Sects. 3.11.4 and 9.2.3 that a transient input can be imparted to a system using an instrumented hammer also. Though all the transient signals mentioned above are very similar from an analysis point of view, the sine chirp signal has the advantage that it offers greater controllability in terms of the amplitude and frequency content of the input and allows larger vibration energy, while burst noise excitation provides decreased test time and reduced leakage errors. Figure 9.6 shows some typical transient signals.
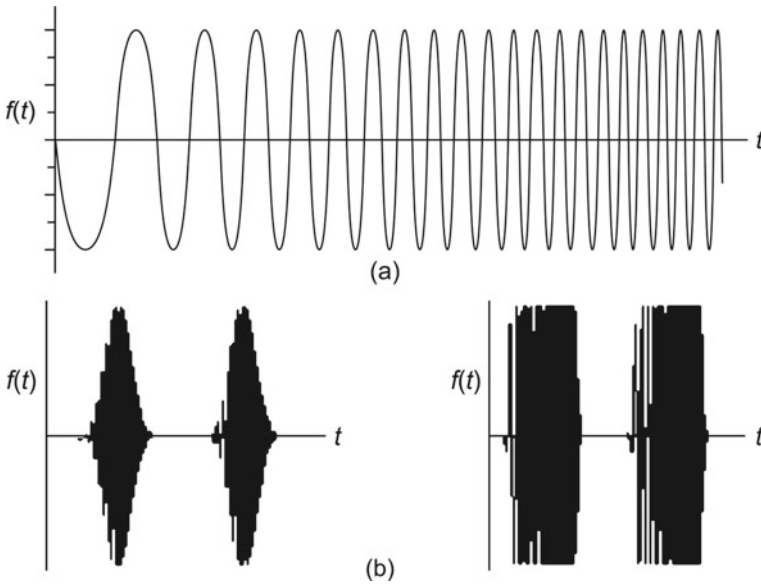
Fig. 9.6   Transient signals for excitation: **a** sine chirp, **b** burst signals

9.2.5.7   Burst Random Excitation

A true random signal is used in burst random testing, but it is turned off well before the end of the finite-length sampling window. This ensures that both the excitation and response signals are completely contained within the sampling window T and that the measured signals are periodic in the window, making them self-windowing in nature and doing away with the need for windows. The advantage of the burst signal is that the structure has returned to rest before the next burst is given. The final FRF is obtained as the average with multiple burst excitations. This is the most appropriate excitation function for systems with closely coupled and or lightly damped modes, since it allows measurements in a leakage free manner. This excitation technique offers a very good compromise between measurement accuracy and speed. It is quite fast and like the true random signal, offers a very good linear approximation of a non-linear system. However, one cannot reduce the duration of the random signal too much, since a minimum amount of energy (or RMS level) is needed for effective excitation of the test structure. Figure 9.7 shows the zoomed waveform of a typical burst random signal. Electrodynamic shakers can also be used to deliver burst loads to test structures, though this method is used more for shock qualification of components. Table 9.1 gives a comparison of various excitation signals.

Table 9.1 Comparison of various excitation signals

Excitation signal	Test time	SNR	Crest factor	Controlled frequency content	Controlled amplitude content	Removes distortion	Reduces leakage	Characterizes non-linearity
Sine steady state	Very long	Very high	High	Yes	Yes	No	No	Yes
True random	Good	Fair	Fair	Yes	No	Yes	No	No
Periodic in analyzer window								
Pseudo-random	Very good	Fair	Fair	Yes	Yes	No	Yes	No
Random	Fair	Fair	Fair	Yes	No	Yes	Yes	No
Fast sine	Fair	High	High	Yes	Yes	No	Yes	Yes
Transient in analyzer window								
Impact	Very good	Low	Low	No	No	No	Yes	No
Burst sine	Very good	High	High	Yes	Yes	No	Yes	Yes
Burst random	Very good	Fair	Fair	Yes	No	Yes	Yes	No
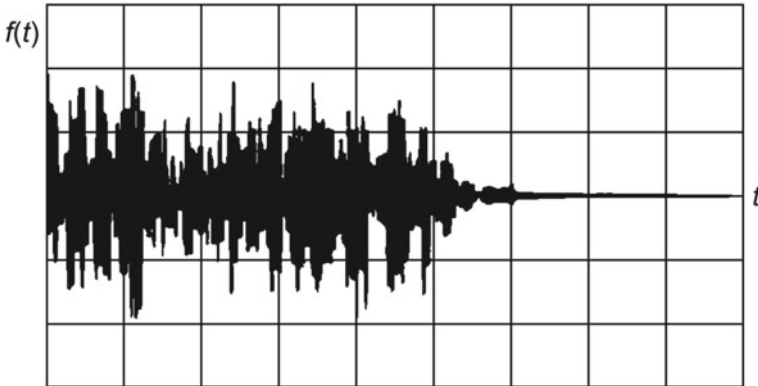
Fig. 9.7 Zoomed version of burst random signal

9.2.6 Applications and Features of a Signal Generator

The signal generator or function generator is a device that generates the input to the power amplifier which drives the electrodynamic shaker/shakers, and hence controls the nature of vibration energy input to the test structure, in terms of excitation waveform and signal level. Basically, the signal generator should be able to deliver different kinds of signals, the choice of which have been seen in the preceding sections. Most generators also produce user-definable waveforms with any signal shape such as triangle or square or arbitrary shape outputs. User defined waveforms can also be swept up and down in frequency similar to a sinusoidal signal.

Typical Technical Specifications of a Signal Generator

Output
  Impedance: 50 Ω
  Amplitude: 0.1–10 V peak

Typical outputs
  White noise: with flatness of ±1 dB up to 40 kHz and multiple
uncorrelated channels
  Band-limited noise: with filter roll off characteristics >20 dB/octave
  Burst: All waveform types with continuous or single shot operation
  Fixed frequency: Sine, square and triangular waveforms
    Frequency range: 1 Hz–40 kHz
    Frequency accuracy: 0.2%
    Frequency stability: 0.2%
    Total harmonic distortion: <2%

Swept frequency
  Sweep types: Linear or logarithmic
  Sweep rate: Programmable with pause, sweep up and sweep down

User defined waveform (arbitrary waveform)
  Buffer size: 8192 samples of 12-bit data

9.3   Representation and Properties of FRF Data of SDOF and MDOF Systems

DID YOU KNOW that most present-day experimental modal analysis procedures are based on frequency domain methods and use the FRF? By the 1950s, the FRF was being used in a number of research areas in sound and vibration in order to understand the dynamics of mechanical systems. By the early 1960s, 'phase separation methods' were well developed and were beginning to be accepted as experimental methods used by automotive, aircraft, and machine tool industries. In these methods, frequency response functions were measured sequentially at a number of discrete frequencies, or by slowly sweeping across analogue frequencies using a single shaker, and the data was analyzed assuming one mode or a limited number of modes in a small frequency band. The need to use both magnitude and phase to separate closely spaced modes was understood. The development of the tracking filter revolutionized the measurement of FRFs and narrowband response spectra. The transfer function analyzer was later developed; this instrument coupled a tracking filter with a sweep oscillator, log voltmeters, phase meter, and x-y plotter in one package, automating the measurement of FRFs.

Having seen the experimental aspects related to modal testing, we now move on to the theory behind modal analysis. In Sect. 2.4.5, the expressions for the FRF characterizing the dynamic properties of a linear time-invariant system have been derived. In the present section are discussed various mathematical forms of the FRF and ways of presenting or graphically displaying them. The plotting of FRF data is an involved process for the reason that there are three quantities, frequency along with two parts of the complex FRF, amplitude ratio and phase angle between the force and the response, and these cannot be fully displayed on a standard x-y graph. Thus, a full representation of an FRF in a single plot can only be made by using a three-dimensional display as illustrated in Fig. 9.8. This is obviously not a convenient way of graphically representing the FRF. Any simple x-y graph can only show two of the three quantities and so there are different possibilities available, all of which are used from time to time.

9.3.1   Graphical Display of FRF Data for SDOF Systems

The FRF of an SDOF system may be defined in terms of receptance, mobility and accelerance or inertance. These quantities are described for an undamped system in

Fig. 9.8 Three-dimensional
plot of the receptance of an
SDOF system
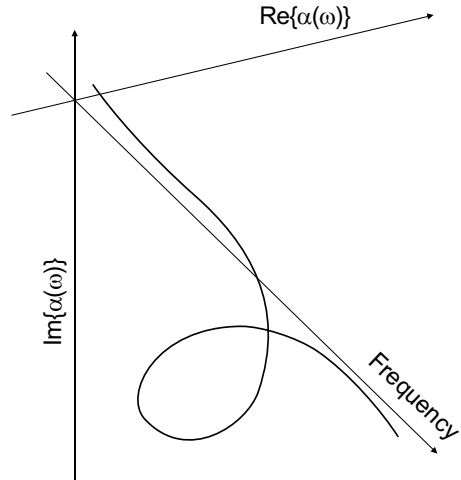
the following discussion. Let $x(t)$, $v(t)$, and $a(t)$ represent the displacement, velocity,
and acceleration responses to a harmonic exciting force $f(t)$. Receptance is defined
as the output displacement to unit input harmonic force and may be described by the
expression below for an undamped system.

$$\alpha(\omega) = \frac{Xe^{i\omega t}}{Fe^{i\omega t}} = \frac{1}{k - \omega^2 m} \tag{9.3}$$

For sinusoidal vibration, we have the following simple harmonic relationships for
the force and response.

$$f(t) = Fe^{i\omega t} \tag{9.4}$$

$$x(t) = Xe^{i\omega t} \tag{9.5}$$

$$\dot{x}(t) = i\omega Xe^{i\omega t} = Ve^{i\omega t} \tag{9.6}$$

and

$$\ddot{x}(t) = -\omega^2 Xe^{i\omega t} = Ae^{i\omega t} \tag{9.7}$$

Mobility is defined as the output velocity to unit input force and has been obtained
based on the above harmonic relationships.

$$Y(\omega) = \frac{Ve^{i\omega t}}{Fe^{i\omega t}} = i\omega \frac{X}{F} = i\omega\alpha(\omega) \tag{9.8a}$$

or

$$|Y(\omega)| = \omega|\alpha(\omega)| \tag{9.8b}$$

and

$$\theta_Y = \theta_\alpha - \frac{\pi}{2} \tag{9.8c}$$

Inertance or accelerance is the ratio of acceleration response to unit input harmonic force as shown below.

$$A(\omega) = \frac{Ae^{i\omega t}}{Fe^{i\omega t}} = -\omega^2 \alpha(\omega) \tag{9.9a}$$

or

$$|A(\omega)| = \omega^2 |\alpha(\omega)| \tag{9.9b}$$

and

$$\theta_A = \theta_\alpha - \pi \tag{9.9c}$$

Receptance, mobility, and inertance are the main formats in which an FRF may be displayed, depending on whether displacement, velocity or acceleration measurements have been made. There do exist other possible ways of displaying FRFs in terms of functions which are inverses of the above as shown below

$$\text{Dynamic stiffness} = \frac{\text{Force}}{\text{Displacement}} \tag{9.10a}$$

$$\text{Mechanical impedance} = \frac{\text{Force}}{\text{Velocity}} \tag{9.10b}$$

$$\text{Apparent mass} = \frac{\text{Force}}{\text{Acceleration}} \tag{9.10c}$$

However, the inverse formats are used only in special cases, since they can give rise to considerable confusion and error, especially for MDOF systems. Table 9.2 summarizes all the FRF parameters, both standard and inverse, along with the various notations utilized.

Receptance for a damped SDOF system is given by the expression

$$\alpha(\omega) = \frac{Xe^{i\omega t}}{Fe^{i\omega t}} = \frac{1}{k - \omega^2 m + i\omega c} \tag{9.11a}$$

Table 9.2 Definitions of frequency response functions

Response parameter	Standard FRF (response/force)	Inverse FRF (force/response)
Displacement	Receptance, admittance dynamic compliance, dynamic flexibility	Dynamic stiffness
Velocity	Mobility	Mechanical impedance
Acceleration	Inertance, accelerance	Apparent mass
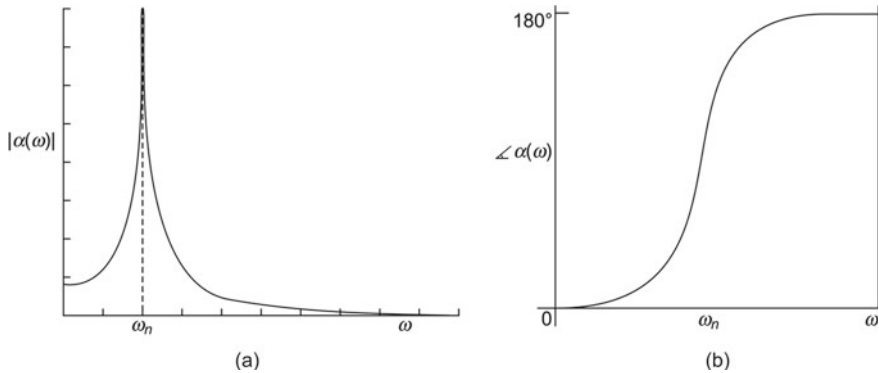
Fig. 9.9 Receptance plot: **a** magnitude, **b** phase

The angle of lag ϕ is given by

$$\tan \phi = \frac{\omega c}{k - m\omega^2} \tag{9.11b}$$

Figure 9.9 shows a typical receptance plot.

Many real structures have a frequency dependent damping called structural or hysteretic damping which has been discussed in Sect. 2.2.5. Here h denotes a complex stiffness and h/k is the structural damping loss factor ξ. The receptance in this case is described by

$$\alpha(\omega) = \frac{1}{k - \omega^2 m + ih} \tag{9.12a}$$

$$\tan \phi = \frac{h}{k - m\omega^2} \tag{9.12b}$$

The three most commonly used graphical presentations are as follows:

(i)   Bodé plots, consisting of two graphs: modulus of FRF versus frequency and phase of FRF versus frequency.
(ii) Real part of FRF versus frequency and imaginary part versus frequency (two plots) and
(iii) Nyquist plot: a single plot consisting of real part versus imaginary part; this representation does not explicitly contain frequency information.

9.3.1.1   Bodé Plots

This style of displaying FRF data applies to both damped and undamped systems, while the other forms are applicable only to damped systems and are sensitive to the
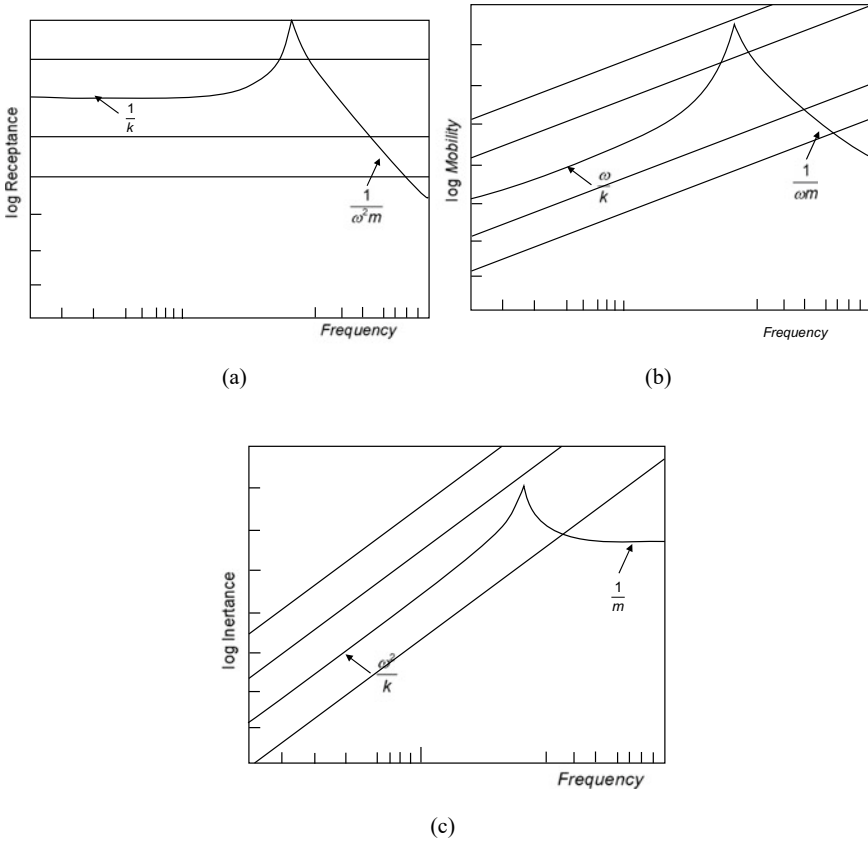
Fig. 9.10 Log-log FRF representations of an undamped SDOF system: **a** log-log receptance, **b** log-log mobility, **c** log-log inertance

type of damping. Irrespective of the type of FRF used, a wide range of values result, typically of the order of 1 to 10000, both in the FRF magnitude, as well as frequency, and these may be compressed using a logarithmic scale, otherwise the smaller values would get submerged if linear scales are used. Plots for receptance, mobility, and inertance of an undamped SDOF system are shown in Fig. 9.10a–c, respectively, with logarithmic scales for both the x- and y-axes. Data which would get displayed as curves with linear scales, become asymptotic to the straight mass and stiffness lines on a plot depicting log modulus versus log frequency. This is advantageous as it provides a simple means of checking the validity of a plot and also allows for easy identification of the mass and stiffness characteristics of the system under study.

Let us consider a rigid mass m, free in space upon which a harmonic force f is applied. The corresponding equation of motion is given by Newton's law

$$m\ddot{x} = f \tag{9.13}$$

Then, the receptance of this system would be

$$\alpha(\omega) = -\frac{1}{\omega^2 m} \tag{9.14a}$$

and the magnitude of $\alpha(\omega)$ in logarithmic terms would be

$$\log|\alpha(\omega)| = -\log(m) - 2\log(\omega) \tag{9.14b}$$

A log-log plot of the magnitude of $\alpha(\omega)$ as a function of frequency is, therefore, a straight line with a slope of -2. Similarly, the receptance of a simple isolated massless spring is given by

$$\alpha(\omega) = \frac{1}{k} \tag{9.15a}$$

Therefore,

$$\log|\alpha(\omega)| = -\log(k) \tag{9.15b}$$

Similarly, the magnitudes and logarithms of magnitudes of the FRF function represented in terms of mobility and accelerance may be derived and are shown in Table 9.3.

From Table 9.3 it is seen that mass and stiffness properties will always appear as straight lines on log modulus versus log frequency plots. Hence, any typical FRF plot can be divided into the following three regions:

(i) a low-frequency straight-line region characterized by the stiffness
(ii) a high-frequency straight-line region characterized by the mass
(iii) the resonant region with its abrupt magnitude and phase variations characterized by the damping in the system.

From Fig. 9.10, which shows the log-log FRF representations of an undamped SDOF system, and Table 9.3, it is clear that when $\omega \ll \omega_r$, the frequency response is approximately equal to an asymptote called the stiffness line and has a slope of 0, 1 or 2 for displacement, velocity, and acceleration responses, respectively. When $\omega \gg \omega_r$, the frequency response is approximately equal to an asymptote called the mass line

Table 9.3 Frequency responses of mass and stiffness elements

FRF parameter	Mass	Stiffness
Receptance $\alpha(\omega)$	$\frac{-1}{\omega^2 m}$	$1/k$
$\log|\alpha(\omega)|$	$-\log(m) - 2\log(\omega)$	$-\log(k)$
Mobility $Y(\omega)$	$\frac{-i}{\omega m}$	$\frac{i\omega}{k}$
$\log|Y(\omega)|$	$-\log(m) - \log(\omega)$	$\log(\omega) - \log(k)$
Inertance $A(\omega)$	$1/m$	$\frac{-\omega^2}{k}$
$\log|A(\omega)|$	$-\log(m)$	$2\log(\omega) - \log(k)$

and has a slope of -2, -1 or 0 for displacement, velocity, or acceleration responses, respectively. The relevant FRF characteristics of simple mass and spring elements may be superimposed on these log-log plots as a grid of lines to enable finding out the mass and stiffness of the system under consideration.

9.3.1.2 Real Part and Imaginary Part Versus Frequency

Receptance FRF of a lightly damped SDOF system has been shown in Fig. 9.11 in the form of real and imaginary parts versus frequency, and from these, it can be seen that the phase change in the vicinity of the resonance frequency causes a sign change in one part, and a peak (maximum or minimum) value in the other parts. It is not possible to use logarithmic scales in this case, as in the modulus plots, because both positive and negative values have to be shown. Due to this reason, this display format is not as widely used as the others, even though it is available as a standard feature in most analyzers.

9.3.1.3 Nyquist Plot

The Nyquist or Argand plane plot or vector response plot consists of a plot of real part of FRF versus the imaginary part and is widely used as it helps display the resonance region in detail. This display format consists of only a single plot, hence the missing frequency information should be added at appropriate points on the curve, usually at regular frequency intervals. Figure 9.12a and b show the Nyquist plots of an SDOF system with viscous and structural damping, respectively. A striking feature of these plots is that only points very close to resonance are clearly identifiable, while those away are very crowded. Indeed, it is this feature (of favourably distorting the plot so
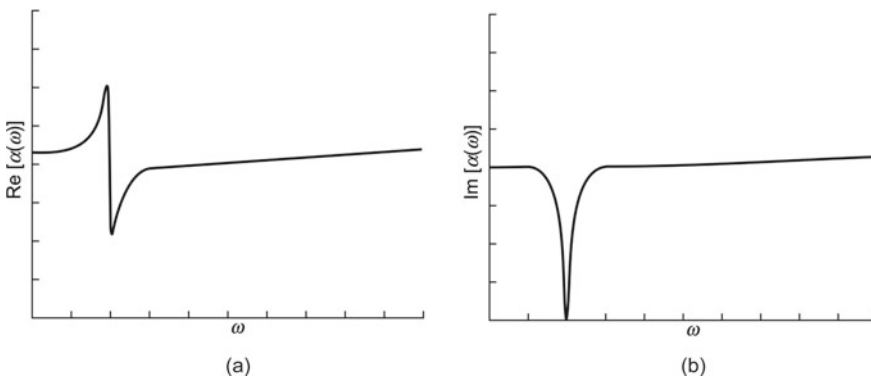
Fig. 9.11 Receptance plots of FRFs for a damped SDOF system: **a** real part, **b** imaginary part
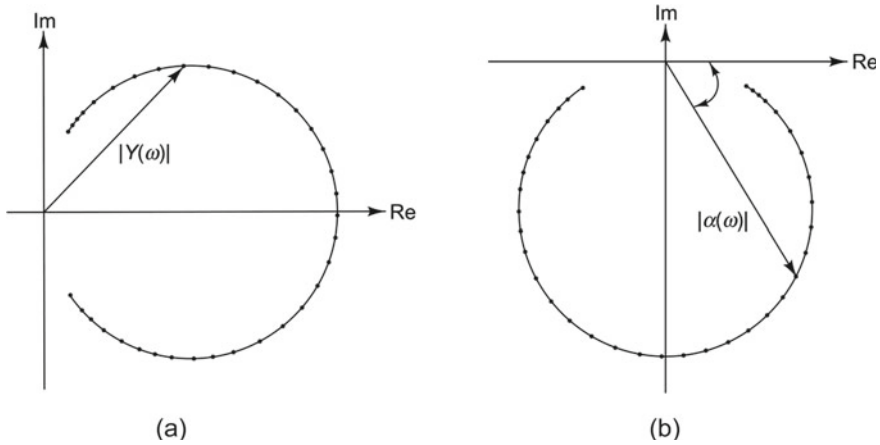
Fig. 9.12 Nyquist plots for SDOF system: **a** mobility with viscous damping, **b** receptance with structural damping

as to focus on the resonance region) that makes the Nyquist plot very attractive for modal analysis.

It is clear from Fig. 9.12 that these FRFs resemble circles. The mobility $Y(\omega)$ with viscous damping traces out an exact circle, while it is the receptance $\alpha(\omega)$ that does so for hysteretic or structural damping. In the other FRFs, the deviation from a circular locus heavily depends on the amount of damping present and becomes negligible for small damping.

9.3.1.4 Properties of SDOF FRF Plots

The basic geometric properties of three specific FRF plots of an SDOF system, which are very convenient forms of representation, are described below. They are as follows:

 (i) Log-log plot of modulus of mobility versus frequency (Fig. 9.10b).
 (ii) Nyquist mobility plot for viscous damping (Fig. 9.12a).
(iii) Nyquist receptance plot for hysteretic damping (Fig. 9.12b).

In a log-log plot of modulus of mobility, it can be seen that for light damping (less than 1%, say), the curve is more or less symmetrical about the resonance frequency. As far as the Nyquist plots are concerned, it is seen that two particular cases, namely, (i) mobility for a viscously damped system and (ii) receptance for a hysteretically damped system, trace out exact circles as frequency ω is changed from 0 to ∞. This may be proved as shown below for hysteretic damping. Receptance with this damping is described by Eq. (9.12a).

Fig. 9.13 Nyquist plot of
receptance
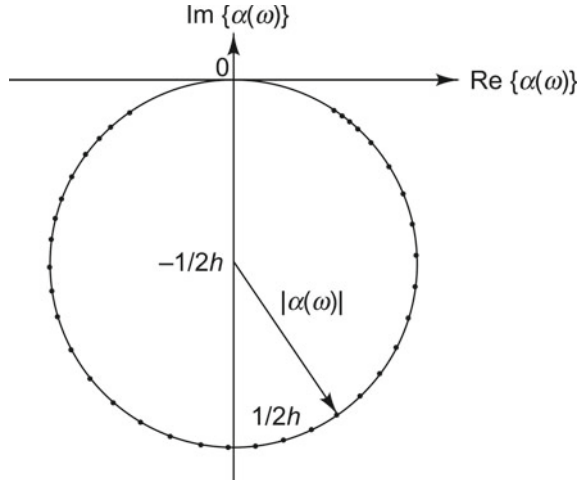
It follows from this equation that

$$\mathrm{Re}(\alpha) = \frac{k - \omega^2 m}{[(k - \omega^2 m)^2 + (h^2)]} \tag{9.16a}$$

$$\mathrm{Im}(\alpha) = \frac{h}{[(k - \omega^2 m)^2 + (h^2)]} \tag{9.16b}$$

It is clear that

$$(\mathrm{Re})^2 + \left(\mathrm{Im} + \frac{1}{2h}\right)^2 = \left(\frac{1}{2h}\right)^2 \tag{9.16c}$$

Thus, the Nyquist plot of receptance for a hysteretically damped SDOF system traces a circle of radius $(1/2h)$ with centre at $(0, -1/2h)$ as depicted in Fig. 9.13. It is to be noted that Nyquist plots are not useful when damping is very low.

9.3.2 Characteristics and Presentation of MDOF FRF Data

The general form of the receptance FRF expression for an MDOF system is

$$\alpha(\omega) = \sum_{r=1}^{N} \frac{C_r}{\omega_r^2 - \omega^2 + iD_r} \tag{9.17}$$

where C_r is a constant representing the contribution to the response from the rth mode and D_r is a damping term. This is an extension of the FRF expression for the much simpler SDOF system studied earlier. It is important to understand the properties of this type of function to get an understanding of the modal analysis process. Hence,

the various means of displaying the information it contains are explained in this section. There are three main options, using displacement, velocity or acceleration response to produce receptance, mobility or inertance (or accelerance), respectively, as in the case of the SDOF system. The mobility and inertance matrices are related to the receptance matrix as shown below.

$$[Y(\omega)] = i\omega[\alpha(\omega)] \qquad (9.18)$$

$$[A(\omega)] = i\omega[Y(\omega)] = -\omega^2[\alpha(\omega)] \qquad (9.19)$$

The FRFs of MDOF systems are more complex than their SDOF counterparts. A general receptance term $\alpha_{jk}(\omega)$ is given by

$$\alpha_{jk}(\omega) = \left(\frac{x_j}{f_k}\right); \quad f_m = 0, \ m = 1, \ N; \ \neq k \qquad (9.20)$$

These FRFs may also be expressed in the form of the inverses of the standard receptance, mobility, and inertance as in the case of the SDOF system. However, deriving these inverse properties from the standard form is not straightforward. The velocity response of a structure to an exciting force may be got in terms of the mobility matrix using the equation

$$\{\dot{x}\} = [Y(\omega)]\{f\} \qquad (9.21)$$

The above equation may be rewritten in terms of the impedance matrix (inverse of mobility) as

$$\{f\} = [Z(\omega)]\{\dot{x}\} \qquad (9.22)$$

It is to be noted, however, that a typical element in the mobility matrix is not related through a simple relationship to its counterpart in the impedance matrix, i.e.

$$Y_{jk}(\omega) = Y_{kj}(\omega) \neq \frac{1}{Z_{jk}(\omega)} \qquad (9.23)$$

because

$$Y_{kj}(\omega) = \left(\frac{\dot{x}_k}{f_j}\right); \quad f_m = 0, \ m = 1, \ N; \ \neq j \qquad (9.24)$$

and

$$Z_{jk}(\omega) = \left(\frac{f_j}{\dot{x}_k}\right); \quad v_m = 0, \ m = 1, \ N; \ \neq k \qquad (9.25)$$

From a practical point of view, while it is possible to measure the mobility FRF by applying an excitation force at a single point, keeping the forces at all other points zero, it is almost impossible to measure impedance keeping all locations grounded except one. Hence, only the standard FRFs can be measured directly. Besides, it is to be noted that an element in the impedance matrix is not obtained as the inverse

of the corresponding element in the mobility matrix, but from an inversion of the matrix itself.

Some common terms used in modal analysis parlance are driving point or simply point FRF and transfer FRF. A point receptance is obtained by keeping the response and excitation coordinates identical, and is measured using an impedance head which has in-built force and response transducers. Point FRFs constitute the diagonal terms of the FRF matrix. A transfer mobility term is one for which the response and excitation coordinates are different and constitutes an off-diagonal term of the FRF matrix. Transfer mobilities are further classified as direct and cross-mobilities; in the former, the response and excitation coordinates are the same, say, both along x-axis, while in the latter, they are different (for example, one along x-axis and the other along y-axis.).

9.3.2.1 FRF Representation of Undamped MDOF System

The receptance for an undamped MDOF system is given by the equation

$$\alpha_{jk}(\omega) = \sum_{r=1}^{N} \frac{_rA_{jk}}{\omega_r^2 - \omega^2} = \frac{_1A_{jk}}{\omega_1^2 - \omega^2} + \frac{_2A_{jk}}{\omega_2^2 - \omega^2} + \frac{_3A_{jk}}{\omega_3^2 - \omega^2} + \cdots + \frac{_NA_{jk}}{\omega_N^2 - \omega^2}$$
(9.26)

Here $_rA_{jk}$ is called the modal constant or modal participation factor for the rth mode with excitation at location k and response at location j; $\omega_1, \omega_2, \omega_3, \ldots, \omega_r$, etc., are the natural frequencies of the 1st, 2nd, 3rd, ..., rth modes. The participation factor indicates the amount each mode contributes to the total response at a particular point.

The individual terms in the FRF may be plotted with log-log axes as described earlier. Figure 9.14 shows the receptance plot corresponding to Eq. (9.26); each term in the series expression for the FRF is shown as a separate curve in the figure and the sum of the individual modal contributions is also shown superimposed. It is, however, difficult to predict the exact nature of the curve from the individual curves, since only the magnitude information is available in the log plots and the phase information is missing. It is thus unclear from the plot whether in each section of a curve (each frequency region) corresponding to one term of Eq. (9.26), the FRF is actually positive or negative in sign and this decides the shape of the total FRF when terms are added together.

Comments on Shape of FRF The shape of the total FRF at any frequency depends on the signs of all the contributing terms at that frequency, which, in turn, depend on the signs of the modal constants. It is also interesting to see what determines the sign of the modal constant. It should be remembered that the modal constant $_rA_{jk} = (_r\psi_j)(_r\psi_k)$ where $(_r\psi_j)$ is the mass-normalized displacement of the jth coordinate (response coordinate) in the rth mode (refer Sect. 2.4.4) and $(_r\psi_k)$ is the mass-normalized displacement of the kth coordinate (excitation coordinate) in the rth mode, or in other words, $(_rA_{jk})$ is the product of two eigenvector elements, one at the response point and the other at the excitation point. In order to visualize eigenvectors,
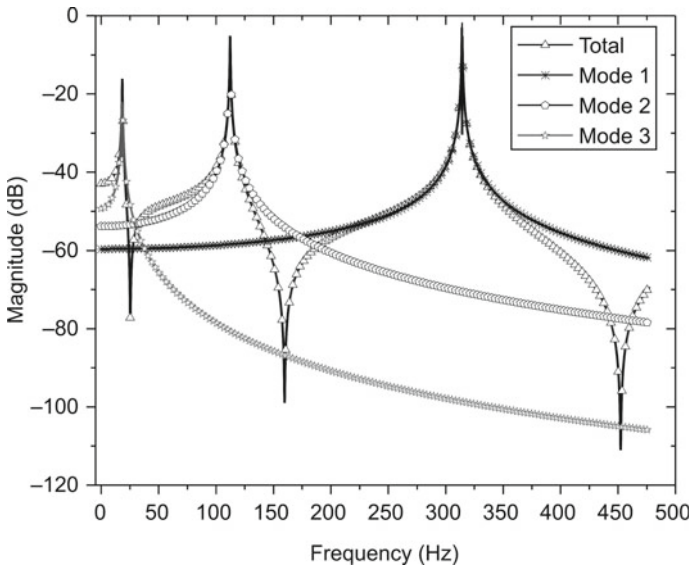
Fig. 9.14 Typical receptance FRF for MDOF system showing contributions of individual modes and their superposition

mode shapes corresponding to the first three modes of a cantilever beam are shown in Fig. 9.15a. $(_1\psi_j)$ and $(_1\psi_k)$ are marked in the figure corresponding to mode 1 for two locations j and k. So also $(_2\psi_j)$ and $(_2\psi_k)$ are indicated for the second mode and $(_3\psi_j)$ and $(_3\psi_k)$ for the third mode. Based on these, the general trends to be expected in point and transfer FRFs are described below.

Point FRF For point mobility, the modal constant for every mode is positive since it is of the form $(_r\psi_j)^2$. This implies that for a point FRF, in the frequency region between two natural frequencies, the signs of the two terms contributing to these modes are negative, leading to cancellation, and therefore, there must compulsorily be an antiresonance (sharp dip) following every resonance (Fig. 9.15b).

Transfer FRF In the case of a transfer FRF, at frequencies where all terms have the same sign and are thus additive, the magnitude of the total FRF curve is higher than the magnitude of any of the individual components, but since the FRF plot uses a logarithmic scale, the contribution of modes other than the predominant one at these frequencies is relatively small. Hence, the total FRF curve is only slightly higher than that for the predominant term. Thus, in the immediate vicinity of a resonant frequency, the contribution of that term is much greater than that of the others, with the result that the total is, in effect, almost the same as that one term. This implies that physically, the response of an MDOF system at one of its natural frequencies is totally dominated by that mode, the participation of the other modes being negligible.

   If two consecutive modes have the same sign for the modal constants, then at some frequency between the natural frequencies of those two modes, there will be an antiresonance, because one term becomes positive and the other negative,
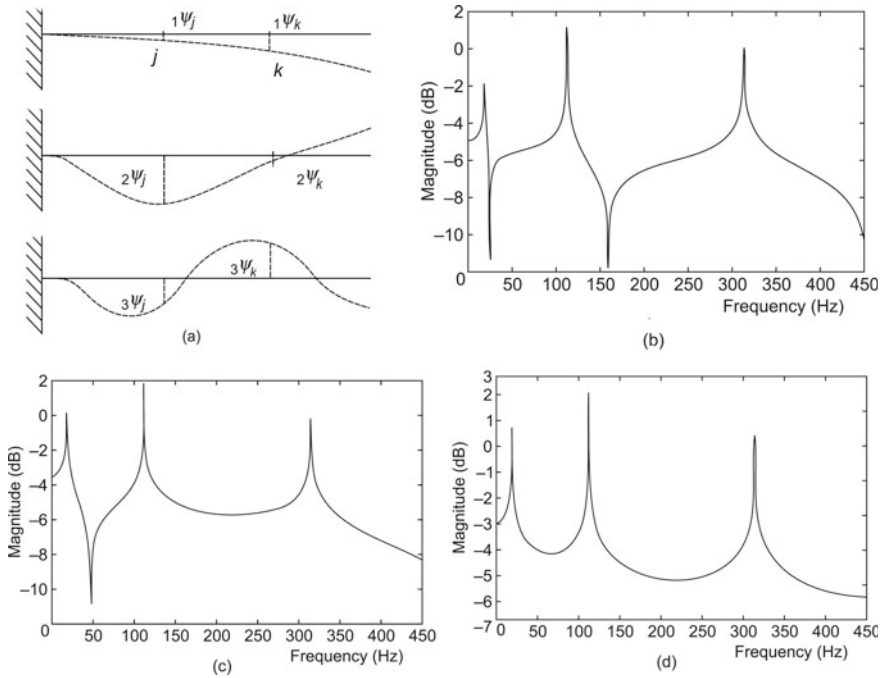
Fig. 9.15 Point and transfer FRF receptance plots for a cantilever beam with damping: **a** mode shapes corresponding to first three modes, **b** point FRF at a point close to the fixed end of the beam, **c** transfer FRF at an intermediate point, **d** transfer FRF at the free end of the beam

cancelling each other out. The most important feature of the antiresonance is that the magnitude is very small and there is an associated phase change. If, on the other hand, the modal constants have opposite signs, then these two adjacent terms in the series will have the same sign leading to addition, and hence there will not be an antiresonance, but just a minimum. In MDOF systems for which the transfer FRF has a large number of terms, the modal constant will sometimes be positive and sometimes negative, depending on the distance between the excitation and response locations. The FRF plot, therefore, shows a mixture of antiresonances and minima (Fig. 9.15c). The further apart the two locations in question are, the higher is the possibility that the two eigenvector elements will alternate in sign as one progresses through the modes. It can be expected that a transfer FRF between two locations far apart on the structure will exhibit less antiresonances, than one between two locations physically closer. A clear example of this is given in Fig. 9.15d.

One important precaution to be taken while making measurements is that neither the excitation nor the response coordinate should coincide with a node for any of the modes, giving rise to the condition $_r A_{jk} = (_r\psi_j)(_r\psi_k) = 0$. If this happens, then the mode in question will not appear as a resonance on the FRF plot because the only contribution at or near $\omega = \omega_r$ will be from the off-resonant modes and this can be expected to be very small.

9.3.2.2   Display of FRFs for Damped MDOF Systems

The general appearance of the FRF (Bodé) plot for damped systems is similar to that
for undamped systems, except that the resonances and antiresonances are rounded
by the inclusion of damping. Besides, the phase angles are no longer exactly 0° or
180°. Figures 9.15b–d show mode shapes and FRFs obtained from a finite element
(FE) model of a damped cantilever beam with 3 DOF, where the force location is
near the fixed end. Rayleigh (proportional) damping with $\alpha = 0.01$ and $\beta = 0$ has
been assumed for the simulation.

   The Nyquist or Argand diagram gives an exact circle for the SDOF case. The
mobility FRF is a true circle with viscous damping, while the receptance FRF is a
true circle with hysteretic damping. There are typically as many lobes or circles in
the FRF as the number of modes. With proportional damping, the modal constants
are real quantities, being identical to those for the undamped version, and will result
in circles symmetric about the Im(FRF) axis. If the signs of the terms of the FRF are
the same, then all the circles lie in the upper half or lower half, depending on the sign.
If they have opposing signs, then some of the circles will lie in the upper half, while
others will lie in the lower half of the complex plane. In the case of non-proportional
damping, the individual modal circles are no longer 'upright', but are rotated by an
amount depending on the phase of the complex modal constants; due to this, the
resonance points are no longer at the bottom (or top) of the corresponding circles.
Figure 9.16 shows a Nyquist plot of the point FRF displayed in Fig. 9.15a. Plots of
the real and imaginary parts of the same Fig. 9.16 as a function of frequency are
shown in Fig. 9.17.

Fig. 9.16 Nyquist
receptance plot for damped
three DOF system
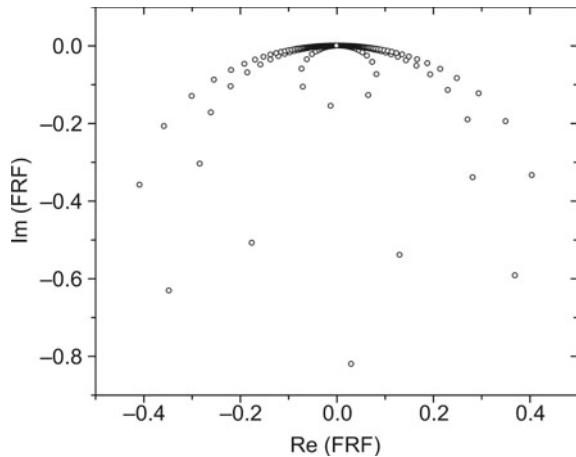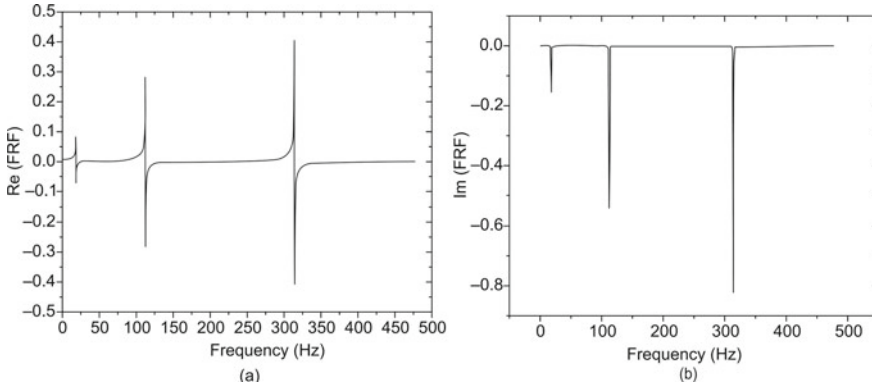
Fig. 9.17 Receptance FRFs plots for damped three DOF system: **a** real part, **b** imaginary part

9.4 Obtaining FRFs with True Random Excitation

DID YOU KNOW that at the heart of digital noise generation is the random number generator? Each time a generation command is given in any of the standard software packages, this generates a new random number having a value between zero and one, with an equal probability between these two values. For most applications, digital noise with a Gaussian probability distribution is used; hence many algorithms create a normally distributed noise signal with an arbitrary mean and standard deviation. The mathematical basis for this algorithm is the Central Limit Theorem which states that a sum of random numbers becomes normally distributed as the number of random numbers added together becomes very large. This theorem does not stipulate that individual random numbers should be from any particular distribution, or that they should be from the same distribution. This theorem also explains why normally distributed signals are so widely prevalent in nature. Most random number generators operate by starting with a seed, a number between zero and one. The seed is passed through an algorithm, resulting in a new number between zero and one, which is taken as the random number. This is then stored and used as the seed for the next random number generation. Thus, a continuous sequence of random numbers can be generated, starting with the same seed. If the program is run multiple times with the same seed, the same random number sequence will be generated. If a different random number sequence is required, reseeding of the random number generator has to be done. A common way of doing this is to use the time indicated by the clock as the seed, thus obtaining a new sequence each time the program is run.

The measurement of an FRF as has been described involves measurement of the input force and output response as a function of frequency with all the phase information

intact. A modal test usually involves single or multipoint excitation; while the former is simpler, the latter method has far-reaching implications due to the flexibility it offers in terms of the combination of position and orientation of the exciter and pickup. The input force may be imparted through a shaker or an impact hammer. FRF measurements have to be made at a sufficiently large number of DOFs to get accurate mode shapes. All modal parameters are extracted from these FRFs and are, therefore, only as good as the FRFs, even if very elegant parameter estimation techniques are used. Modal parameter extraction from the FRFs may be done using any of the curve fitting techniques available. It has to be ensured that driving point FRFs (or simply point FRFs with the excitation and response measured at the same DOF) are included in order to calibrate (scale) the modal model.

When excitation is imparted using an impact hammer, we have what is called roving impact excitation, the hammer being moved around the structure to excite every DOF, with the accelerometer response DOF being fixed and serving as a reference DOF. With shaker excitation, the excitation DOF is usually fixed, serving as the reference DOF. We then have roving response measurements by moving the response transducer around the structure. This causes a varying mass loading of the structure and corrective action for this must be taken while extracting modal parameters. If instead of one roving accelerometer, however, there are many accelerometers simultaneously fixed at all the DOFs, the mass loading at each DOF will be the same in all the FRFs, providing better consistency of results and minimizing testing time. Figure 9.18a shows a typical modal analysis test setup with roving excitation and Fig. 9.18b shows the setup with fixed shaker and roving transducer.

9.4.1   Single- and Multiple-Input System Response Models

Both Single-Input Single Output (SISO) and Single-Input Multiple Output (SIMO) analyses involve excitation at a single point. Figure 9.19 shows the system configuration for the SIMO case.

Typically, a random signal is fed to a shaker which excites the system. Alternately, a broad band input may be imparted through an impact hammer. The input force and displacement/velocity/acceleration response signals are measured and fed to a real-time analyzer which processes the time domain data to produce FRFs. The equation relating the input force to the output response of a linear system is given by

$$Y_i = H_{ij} X_j \qquad\qquad (9.27a)$$

where Y_i is the output spectrum at the ith DOF, X_j is the input spectrum at the jth DOF and H_{ij} is the FRF relating the two. Thus

$$H_{ij} = \frac{Y_i}{X_j} \qquad\qquad (9.27b)$$
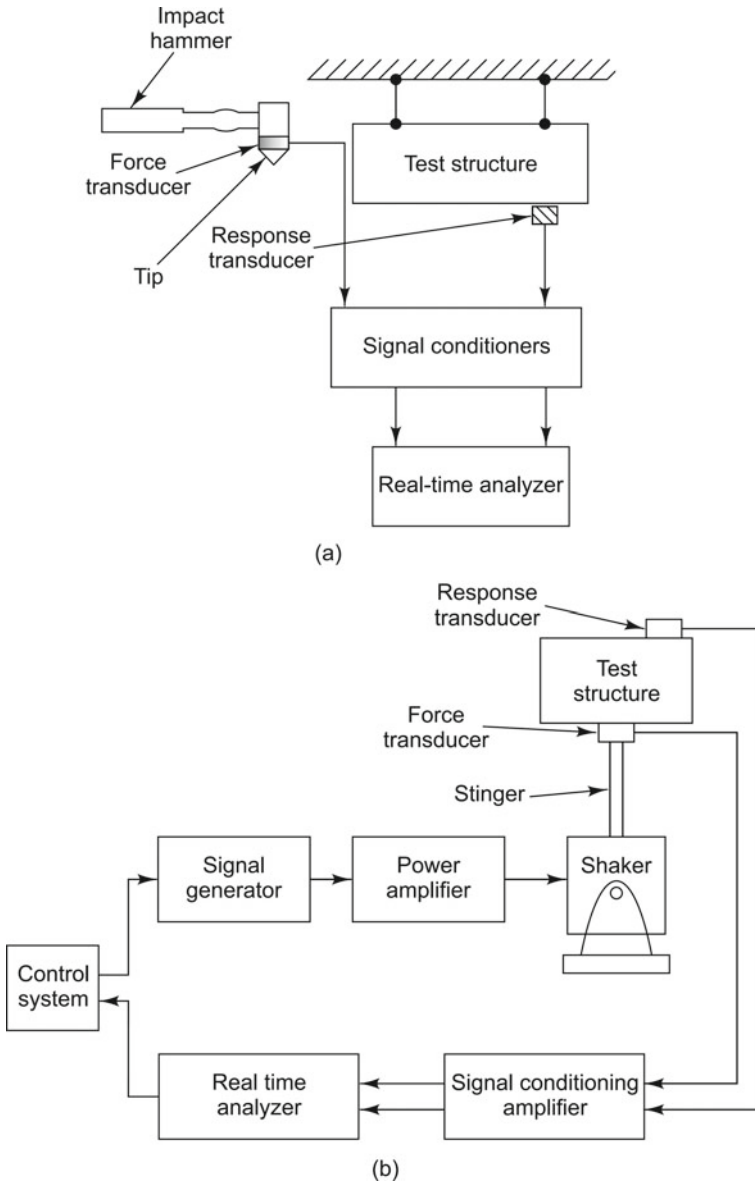
Fig. 9.18 Typical modal analysis test arrangements: **a** roving hammer excitation, **b** fixed shaker excitation

with the input forced to zero at all DOFs other than j. There are different classical estimators for estimating the optimal FRF H_{ij}. One of the most common ways of determining the FRF with single point excitation is through the use of the expression
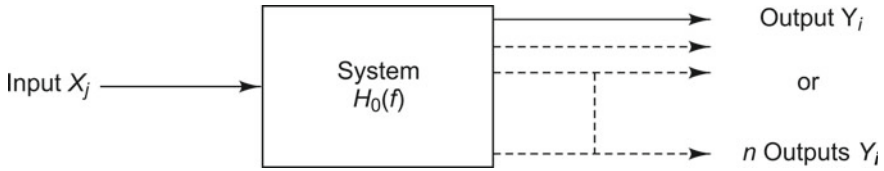
Fig. 9.19 System configuration for SIMO

given below

$$S_{yy}(\omega) = |H(\omega)|^2 S_{xx}(\omega) \tag{9.28}$$

Here $S_{yy}(\omega)$ and $S_{xx}(\omega)$ represent the auto-power spectral densities of the response and excitation, respectively, and $H(\omega)$ is the FRF linking the output response y of the structure to the input force x. Though this is a convenient expression, it has the drawback that it does not provide a complete description of the random vibration, since it lacks the phase information of the FRF. Alternate ways of describing the response are based on cross-correlation between the excitation and the response and are of the form given below

$$S_{xy}(\omega) = H(\omega)S_{xx}(\omega) \tag{9.29a}$$

or

$$S_{yy}(\omega) = H(\omega)S_{yx}(\omega) \tag{9.29b}$$

In this expression, $S_{yx}(\omega)$ is the cross-spectral density between the output and input signals, i.e., the complex conjugate of $S_{xy}(\omega)$. Most commercially available spectrum analyzers have the facility to estimate these various quantities. However, in spite of using random signals (which typically require very lengthy records for a complete description) for excitation, their evaluation is based on finite-length data and these lead to errors. Errors also creep in from the measured forces and responses. Hence, to understand how much error has crept in, the results may be cross-checked by obtaining different FRF estimates such as $H_1(\omega)$, $H_2(\omega)$, etc., using more than one equation. Details regarding various FRF estimators, and what is called coherence is discussed in the next section.

9.4.2 Issues Involved in Multiple-Input Multiple- Output (MIMO) Analysis

The analytical expressions extended to the case of simultaneous multiple excitations have not only the auto- and cross-spectral densities of and between the different inputs, but also the cross-spectral densities between inputs and outputs as discussed in Sect. 8.12.3.1. For uncorrelated inputs, the input/output expression would take the
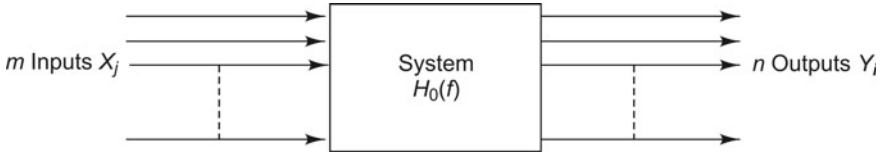
Fig. 9.20 System configuration for MIMO testing
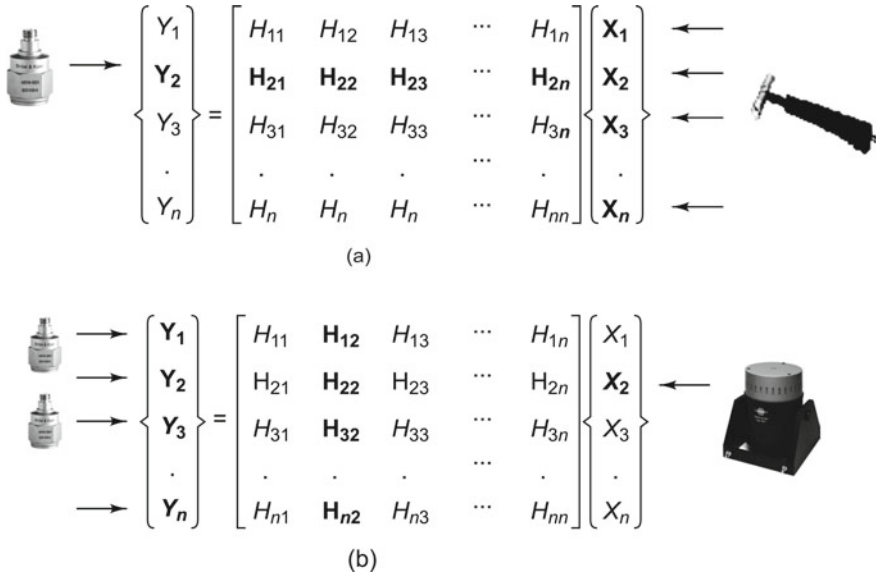
Fig. 9.21 Experimentally obtaining the H matrix: **a** single fixed response measurement, roving impact excitation, **b** single fixed shaker excitation, roving response measurement

matrix form

$$[S_{xy}(\omega)] = [S_{xx}(\omega)][H(\omega)] \tag{9.30}$$

where $[S_{xx}(\omega)]$ is the matrix of the auto- and cross-spectral densities of and between the different input DOFs, and $[S_{xy}(\omega)]$ is the matrix of the cross-spectral densities between inputs and outputs. The assumption behind this computation is that at each frequency of interest, all the forces are non-zero and that no pair of forces is fully correlated. Figure 9.20 shows the system configuration for the MIMO testing.

An important question to be considered in Multiple-Input Multiple-Output (MIMO) analysis is whether we need all the individual FRFs in the $[H]$ matrix or whether a few are sufficient to establish the modal model as shown in Fig. 9.21a and b.

The answer to this question will tell us how many auto- and cross-spectral density terms need to be measured to estimate these FRFs. From a theoretical point of view,

only one row or one column of the FRF matrix $[H]$ may suffice, but from a practical point of view, it may not be enough.

9.4.2.1 Single Reference Modal Analysis

The assumption behind single reference modal analysis is that the selected reference DOF does not correspond to a nodal position for any of the modes. This DOF thus contains information about all the modes and ensures accurate modal parameter extraction. Hence, in such cases, FRF data are processed with data measured from only one reference DOF, i.e., measurement of only one row or one column of the $[H]$ matrix (Fig. 9.21) is sufficient. The identification of a proper reference DOF may require some initial trial and error, or results of finite element (FE) analysis if available, could be used. With a roving impact excitation (Fig. 9.21a), only one response DOF is needed, i.e., only one accelerometer point along a predetermined direction which serves as the reference DOF. In this case, we need to determine only one row of the FRF matrix. With a fixed shaker test (Fig. 9.21b), the location and direction of the shaker serve as the reference DOF and we need only one column of the FRF matrix to get the complete modal information required. A test in which multiple outputs are sensed in each measurement is called a Single-Input Multiple-Output (SIMO) test configuration.

9.4.2.2 Multireference Modal Analysis

In cases where it is not possible to find a proper reference DOF, measurements with more than one reference DOF will be required, i.e., measurements of more than one row or more than one column of the FRF matrix $[H]$. This happens if a structure exhibits local modes with predominant modal deflections at different parts of the structure and in complex structures made of different parts having different structural properties. It then becomes impossible to find a DOF where all the modes have a sufficiently large contribution as to enable proper modal extraction. Multireference testing also becomes necessary in symmetrical structures with more than one mode at the same frequency, requiring the measurement of as many rows or columns as the number of modes (or more) at the same frequency. Here, it is also to be ensured that the mode shapes to be extracted look different in the reference DOFs. With measurements over many rows or columns of the FRF matrix, it becomes possible to enhance different modes by computation of linear combinations of these rows or columns. Thus, the sum of the two rows or the two columns of a symmetrical structure measured with two reference DOFs at symmetrical locations may help in highlighting the symmetrical modes, while the difference may enhance the antisymmetrical modes. In the case of a roving impact test, as many rows can be obtained as the number of response DOFs measured. With a fixed shaker excitation, a multireference data set could be acquired simply by measuring one column at a time, by keeping the shaker first in one position (one reference DOF), then in another position, and so on.

9.4.2.3   Multiple-Input Multiple-Output Modal Test

Instead of obtaining a multireference data set by measuring one column at a time as suggested earlier, a better solution could be obtained through the use of multiple shakers to perform a 'multiple-input multiple-output (MIMO)' test. The multiple reference impact testing (MRIT) method or MIMO method has been used successfully, especially in large and complex structures and structures with heavy damping such as aircraft, wind turbines, large industrial fans, bridges, etc. Reduced cost of electronics leading to multiple channel data acquisition systems, inexpensive transduction systems, and multiple measurement parameter estimation algorithms have made multipoint excitation possible. The MIMO method uses controlled multiple shaker excitation or multipoint impact testing. Single point excitation does not work here for the reasons that the excitation force applied at a single point may not excite all parts of the structure owing to its large size and if the excitation level is increased in an effort to do so, the system might be driven into the non-linear range of vibration, resulting in deterioration in the estimation of the FRFs. Besides, results got from finite element analysis of large structures also require verification, and for this, modal tests based on multipoint excitation (MIMO tests) come in handy. Many exciters (reference DOfs) are used to simultaneously excite the structure at multiple DOFs, resulting in simultaneous measurements of many columns of the FRF matrix. The response transducers must be roved around or if there are many transducers available, simultaneous measurements may be made at all the response DOFs. Another aspect to be looked into in MIMO studies is that the excitation forces applied to the structure are random and uncorrelated in order to satisfy the mathematical assumptions made in the analysis. Typically, uncorrelated random excitation signals such as pure random, burst random or periodic random signals are used for excitation. Signal generators with multiple signal sources allow one to perform multiple point random modal tests. System non-linearities can also give rise to low coherence and their effects can be minimized by appropriate choice of excitation levels and sufficient averaging. Generally, for MIMO testing 10 to 20 exciter positions or roving impact hammer positions and up to 100s or 1000s of response locations may be used depending on the requirement. It is required to have a multichannel FFT analyzer or modal analysis system, with the basic minimum being a 2 channel analyzer. Triaxial or uniaxial accelerometers may be used to make FRF measurements.

The most attractive feature of MIMO testing is that it provides a better distribution of the input force energy over the structure, as compared to single point excitation, resulting in a more uniform vibration response over the structure. Excitation at multiple locations also simulates more closely the excitation that the structure experiences during real-life operation. Simultaneous measurement of FRFs at more than one column provides increased consistency in the data set which is an important consideration in polyreference modal extraction procedures or when linear superpositions of columns are computed for modal enhancement. Besides, this method results in shorter measurement time, as compared to sequential column measurements.

There is another method called multipoint sine dwell test or normal mode testing method (or phase resonance method as it was called in the 1960s) that uses fixed fre-

quency sinusoidal excitation with multiple shakers. Signal generators with multiple signal sources allow one to tune the excitation frequency of the system to a resonance frequency, analyzing one mode at a time. The amplitude and phase of the force signals should be set such that they match those of the mode being excited, i.e., the force signals are set according to the mode shape of the mode under consideration. This exercise is very time consuming and is, therefore, not very popular.

Yet another form of modal analysis is the so-called operational/ambient/output-only modal analysis. This method involves extraction of modal parameters from response data alone, without knowledge of the input excitation forces. This method is convenient for parameter estimation of large structures like bridges and towers where it is difficult to artificially excite the structure and determine the responses.

9.4.3   Estimation of Frequency Response Functions

Ideally the FRF should give a very accurate representation of the actual structural response. But quite often the FRFs are distorted by noise in the input or output or both, and hence a lot of research has been done on obtaining good FRFs. Most FFT analyzers use the conventional H_1 estimator to estimate the FRF. There is also the H_2 FRF which is an improvement over the H_1 estimator. There are other ways of estimating FRFs as well, such as H_c and H_v. Four methods of estimating FRFs are presented in this section. Let us first start with the derivation of the H_1 FRF. For this, consider a model of the measurement system shown in Fig. 9.22 which contains noise in both the input and output.

We shall use the following notations for the subsequent derivations.

$u(t)$ = the true input signal to the system (not measured signal)
$v(t)$ = the true output of the system (not measured signal)
$n(t)$ = the uncorrelated noise in the input
$m(t)$ = the uncorrelated noise in the output
$x(t)$ = the input signal measured by the FFT analyzer, which is the sum of the true input and uncorrelated noise, i.e.
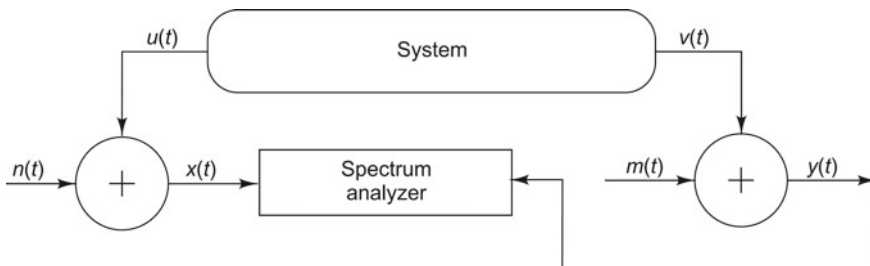
Fig. 9.22 System measurement model

$$x(t) = u(t) + n(t) \tag{9.31}$$

$y(t)$ = the output signal measured by the FFT analyzer, which is the sum of the true output and uncorrelated noise, i.e.

$$y(t) = v(t) + m(t) \tag{9.32}$$

S_{uu} = power spectral density of the true input
S_{vv} = power spectral density of the true output
S_{nn} = power spectral density of random measurement noise in the input
S_{mm} = power spectral density of random measurement noise in the output
S_{xx} = power spectral density of the measured input, i.e.

$$S_{xx} = S_{uu} + S_{nn} \tag{9.33}$$

S_{yy} = power spectral density of the measured output, i.e.

$$S_{yy} = S_{vv} + S_{mm} \tag{9.34}$$

S_{xy} = cross-spectral density between measured input x and measured output y
S_{uv} = cross-spectral density between the true input u and the true output v.

The true FRF is denoted by $H_0(f)$ and is defined as shown below.

$$H_0(f) = \frac{S_{uv}}{S_{uu}} = \frac{U^*(f)V(f)}{U^*(f)U(f)} = \frac{V(f)}{U(f)} \tag{9.35a}$$

Here the superscript * denotes the complex conjugation operator. Alternately

$$H_0(f) = \frac{S_{vv}}{S_{vu}} = \frac{V^*(f)V(f)}{V^*(f)U(f)} = \frac{V(f)}{U(f)} \tag{9.35b}$$

For any measurement system contaminated by noise, it is obvious that it is impossible to measure the true FRF. Therefore, one needs to get estimates of the FRF, which are as close as possible, to the true FRFs.

9.4.3.1 H_1 Frequency Response Function

This is the conventional way of estimating the FRF. Here the cross-spectral density between the input and output is divided by the auto-spectral density of the input. The noise signals $n(t)$ and $m(t)$ in the input and output, respectively, are assumed to be uncorrelated with each other and with the true input $u(t)$ and true output $v(t)$.

$$H_1(f) = \frac{S_{xy}}{S_{xx}} \tag{9.36}$$

Here

$$S_{xy} = X^*(f)Y(f) = [U^*(f) + N^*(f)][V(f) + M(f)] \qquad (9.37)$$

Since uncorrelated noise $n(t)$ and $m(t)$ get cancelled out after averaging over a sufficiently large number of records, S_{xy} reduces to S_{uv}. This is expressed mathematically as

$$H_1(f) = \frac{X^*(f)Y(f)}{X^*(f)X(f)} = \frac{U^*(f)V(f)}{X^*(f)X(f)} = \frac{S_{uv}}{S_{xx}} \qquad (9.38)$$

However, by definition, the true FRF is as defined in Eq. (9.35). Dividing the numerator and denominator of Eq. (9.38) by S_{uu} and substituting for $H_0(f)$ from Eq. (9.35a), we obtain the following simplification.

$$H_1(f) = \frac{S_{uv}}{S_{xx}} = \frac{S_{uv}}{S_{uu} + S_{nn}} = \frac{(S_{uv}/S_{uu})}{1 + (S_{nn}/S_{uu})} = H_0(f)\left[\frac{1}{1 + (S_{nn}/S_{uu})}\right] \qquad (9.39)$$

The above equation relates the H_1 estimate to the true FRF H_0 and is seen to be sensitive to input noise S_{nn}. $H_1(f)$ is not good around resonance since S_{uu} itself becomes quite small in this region (of the order of S_{nn}). However, at antiresonance, it is to be noted that S_{nn}/S_{uu} becomes small and hence $H_1(f)$ is very good in this region.

9.4.3.2 H_2 Frequency Response Function

This method was proposed as an alternative to $H_1(f)$ which is contaminated by input measurement noise, as revealed by Eq. (9.39). This inverse method for calculating an FRF is given as

$$H_2(f) = \frac{S_{yy}}{S_{yx}} = \frac{S_{vu} + S_{mm}}{S_{vu}} \qquad (9.40)$$

Dividing numerator and denominator by S_{vv}, the relationship between $H_2(f)$ and the true FRF $H_0(f)$ is obtained as

$$H_2(f) = \left(1 + \frac{S_{mm}}{S_{vv}}\right)\frac{S_{vv}}{S_{vu}} = H_0(f)\left(1 + \frac{S_{mm}}{S_{vv}}\right) \qquad (9.41)$$

Most commercially available analyzers provide only one of these two forms of FRFs, and generally $H_1(f)$ is more common because it is slightly easier to compute. The $H_2(f)$ FRF estimate, unlike the conventional $H_1(f)$, is insensitive to input noise. Besides, since the input force signal is usually low relative to the measurement system noise at the resonant frequencies, i.e., since S_{nn}/S_{uu} is large, H_2 is a better estimator in this region. If H_1 is used in this situation, erroneous results will be obtained; on the other hand, H_2 remains unaffected. In addition, at resonance, the true output signal

is very large compared to the output noise, i.e., S_{mm}/S_{vv} is very small compared to errors in H_1 around resonance. Therefore, H_2 yields an almost perfect estimate of the true H_0 in the region of resonance. However, at antiresonance, the output power spectrum drops towards output noise S_{mm}, while the input power spectrum remains high compared to the input noise power spectrum. Under such conditions, H_2 does not yield acceptable results because of its sensitivity to output noise.

9.4.3.3 Coherence Function

The coherence function γ_{xy}^2 is used to detect errors in estimating an FRF. Coherence is defined as

$$\gamma_{xy}^2 = \frac{|S_{xy}|^2}{S_{xx}S_{yy}} \tag{9.42a}$$

The coherence function is always less than or equal to 1.0. One can compute the coherence function in an alternative way, as

$$\gamma_{xy}^2 = \frac{H_1(f)}{H_2(f)} = \frac{(S_{xy}/S_{xx})}{(S_{yy}/S_{yx})} = \frac{S_{xy}S_{xy}^*}{S_{xx}S_{yy}} = \frac{|S_{xy}|^2}{S_{xx}S_{yy}} \tag{9.42b}$$

If the measurements are made accurately, coherence must be very close to unity. If coherence is small, the FRF should be discarded and it must be determined why it is small. A small frequency resolution in the FFT analyzer may be a possible cause or there may be noise in either the excitation or the response signals, which are possibly degrading the measured spectra. One of the most likely sources of low coherence in lightly damped structures is poor frequency resolution in the analyzer. If the resolution is not fine enough to describe adequately the rapid amplitude changes around resonances and antiresonances, the resulting FRF data are poor. The coherence function also gives an indication if the measured response can be fully attributed to the measured inputs. A value of the coherence function (Fig. 9.23) less than 1 over a frequency range indicates that the response is only partially due to the measured inputs in this range and may be the result of unknown inputs. If the number of averages taken under such circumstances is increased, this situation does not change, and therefore the coherence function also does not change, but will only be less noisy. The only way to improve the quality of coherence functions is to take all possible precautions to eliminate unknown sources. This can be achieved by ensuring proper boundary conditions, minimizing shaker-structure interaction, etc. When the coherence is low in spite of all such precautions, a zoom measurement may be made to enhance the accuracy of the measurements around resonance, but at the cost of increased measurement time. It is a good practice to make several measurements and to accumulate a running average of the corresponding FRF estimates and the coherence.

While conducting MIMO studies, it has to be ensured that the excitation forces applied to the structure are random and uncorrelated with each other in order to satisfy the mathematical assumptions made in the analysis. Actual excitation forces
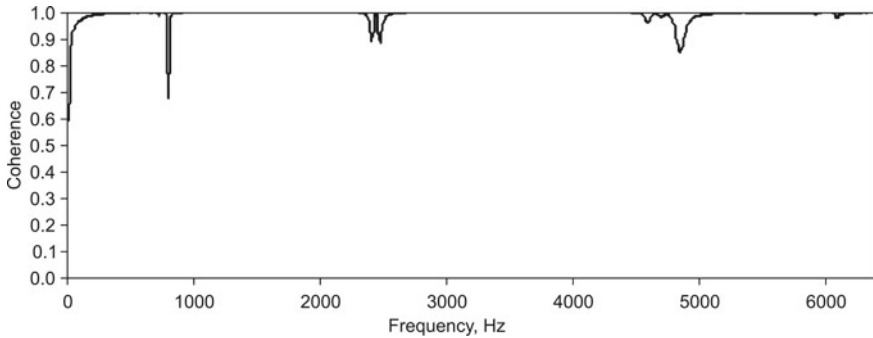
Fig. 9.23 Typical coherence plot

are in fact uncorrelated. However, at the natural frequencies, the structure loads the input drive systems, making the multiple excitation forces correlated and leading to degradation in the quality of FRFs and hence the coherence function. Constant current power amplifiers may be used to drive the exciters, instead of the usual constant voltage amplifiers to reduce these problems. Another problem is that the response measured at a point cannot be completely attributed to the force which is measured. This happens if the coupling between the shaker and the structure is too stiff and a lateral or rotational constraint is inadvertently applied to the test structure, giving rise to low coherence. System non-linearities can also give rise to low coherence.

9.4.3.4 H_3 Frequency Response Function

From previous discussions, it can be concluded that the H_1 estimator minimizes the error due to noise on the output, but is very sensitive to noise on the input, resulting in an underestimation of the true FRF, H_0, around resonance. The H_2 estimator, on the other hand, minimizes the error due to the noise on the input, but is very sensitive to noise on the output. This results in an overestimation of the true FRF H_0. Since H_1 is a lower bound estimator and H_2 an upper bound estimator, the true FRF H_0 must lie somewhere between the two estimates. This suggests a third method of estimating an FRF, called H_3 as the average or arithmetic mean of H_1 and H_2 as follows:

$$H_3(f) = \frac{H_1(f) + H_2(f)}{2} \tag{9.43}$$

This estimator has contamination from the input and output noises. However, the high and low estimators tend to cancel each other, thus yielding a better estimator of the FRF.
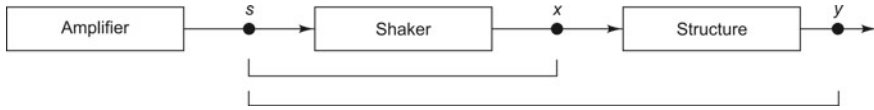
Fig. 9.24 Shaker-structure interaction

9.4.3.5 H_c Frequency Response Function

The H_c estimator is an unbiased FRF estimator. It attempts to eliminate the errors in FRF estimation caused by non-linearities and shaker-structure interaction during electrodynamic random shaker excitation as shown in Fig. 9.24.

Here

$s =$ driving or actuating signal
$x =$ input signal (reference)
$y =$ output signal (response)

The complex version of this frequency response estimator is given by the following formula:

$$H_c = \frac{S_{ys}}{S_{xs}} \tag{9.44}$$

where

$S_{ys} =$ cross-spectral density of response with actuating signal
$S_{xs} =$ cross-spectral density of reference signal with actuating signal

The subscript c in H_c is an indication that the estimator is obtained from the ratio of complex cross-spectral estimates. Both the cross-spectra are with reference to the driving or source signals. Thus, all the signals within the force applied to the structure and the structural responses which are uncorrelated with the input driving signal are averaged out in the cross-spectral averaging process, irrespective of whether or not one is in the vicinity of resonance or in the antiresonance regions of the FRF.

9.5 Signal Processing Problems and Solutions in Modal Analysis

DID YOU KNOW that the era of modal analysis began only after major advancements in signal processing software and hardware in the 1960s? Signals are processed at every stage of experimental modal analysis for the determination of modal parameters, starting from the generation of excitation signals and their

amplification, conditioning of the signals acquired from the transducers, estimation of FRFs from the measured noisy force and response signals, computation of coherence functions to the animation of mode shapes. Thus, though experimental modal analysis is considered a vibration engineer's domain, the signal processing toolbox used by him/her cannot be considered a black box.

There are many signal processing techniques which may be used for improving the quality and accuracy of measurements and for reducing measurement errors; they are autoranging, averaging, windowing and zooming, to name a few. Many of these aspects have already been dealt with in Chap. 8.

9.5.1 Autoranging

It is required to obtain a high SNR and to avoid signal overloads (to prevent clipping), especially with impact excitation, in order to get meaningful FRFs. For this, the dynamic range of the analogue-to-digital converter (ADC) of each channel may be optimized using a feature called autoranging which is available in most analyzers. If the input range of the ADC is set too low, clipping will occur; if, on the other hand, it is set too high, the sensitivity of the measurements will be poor. The ADC range of each input channel must, therefore, be set according to the maximum anticipated voltage levels on the channel. Ideally, the peak amplitude of the measured signal should remain within the upper half of the input range. This is easy to set with continuous signals, but not so with impact or other transient signals. Hence, the ADC range for impact testing may be set manually or the modal analysis system may be trained by repeatedly impacting the structure while the analysis system autoranges until no overloads are detected. If the impacts are consistent, then the autoranging may be successful with a smaller number of averages, if not a large number is required. This may be a time-consuming procedure, especially with a large number of input channels, but definitely yields good results. Another point to be noted while autoranging is that it should be done at the highest available sampling frequency, otherwise frequency components of the impact signal above the test frequency range could cause overloading. Often the range settings of the input channels may be used for several measurement locations if impacts are made in the same direction.

9.5.2 Removing Noise and Distortion from Measurements

In the process of obtaining transfer functions, power spectrum averaging is done as has been seen. This helps remove uncorrelated extraneous noise from the measurement. It also helps reduce the statistical variance of responses to random excitation functions and to reduce the effects of system non-linearities. Averaging measured spectra will decrease noise in FRFs and in the measured coherence function, but

will not change the shape of the latter. However, increasing the number of averages will increase statistical confidence in the measured FRFs; thus the results will be acceptable even with low coherence. Once enough records of data have been averaged together, the transfer function is computed by using H_1 or H_2 or any of the other estimates described earlier and unbiased estimates of the transfer function may be obtained in the presence of noise.

Distortion, associated with non-linear motion, is an unwanted contaminant of the measured vibration signals. While making measurements, we must ensure that the data obtained accurately describe the linear motion of the structure, and not any non-linear motion, since the dynamic model on which all modal testing theory is based is linear. Distortion can occur if the range of the ADC or amplifier is not set sufficiently high or if there is an unwanted vibration stop in the test setup. Distortions due to non-linearities in the structure show up as random noise in the spectrum and can be got rid of by a sufficiently large number of averages, like any other extraneous noise.

9.5.3 Windowing

Though FFT techniques are highly useful for fast computation of spectra, they have the limitation that they are based on periodicity of the time signals in the analysis window considered, resulting in leakage as seen in Sect. 8.8.3. In practice, due to leakage effects, the resulting FRF gets smeared, leading to a broadening of the resonance peaks. Therefore, in FRF measurements, leakage leads to underestimation of the magnitude at the FRF peaks and distortion of phase and drop in coherence at the peaks. This leads to errors in estimates of the modal parameters obtained from signals which are non-periodic in the time window, but does not cause a problem with transient signals such as an impact force or impulse response function which decays well before the end of the window. The periodic random (pseudo-random) signal contains the desirable properties of being random, yet repetitive in the time window, and removes distortions from the measurement. The only drawback of this excitation signal is that it is two to three times slower than the other random methods.

9.6 Modal Parameter Extraction Methods for SDOF and MDOF Systems

DID YOU KNOW that modal parameter extraction is actually a system identification procedure? This is a method for building a mathematical model of the vibratory system using the measured input force to the system and the measured vibratory response. The process of system identification involves measuring the input and output signals in the time or frequency domain to determine a math-

ematical relation between them without going into many of the finer details of what is going on inside the system. Actual parameter extraction is done from a set of impulse response functions (IRFs) in the time domain or a set of FRFs in the frequency domain, measured at salient locations on the structure (avoiding those locations for which the excitation or response measurement is in a nodal position). Modal parameter extraction may also be thought of as a curve-fitting procedure since it attempts to match an analytical expression (IRF or FRF) of the structure's dynamics with the measured data representing its dynamics, a process called curve fitting. During this process, the unknown parameters of the model are estimated; therefore, the process is also called parameter estimation. The purpose of all curve fitting procedures in modal analysis is to obtain the most accurate estimates of modal parameters. Good curve fits are necessary for accurate parameter estimates, but may not always be sufficient.

9.6.1 Preliminary Checks of FRF Data

Before embarking on a modal parameter extraction procedure, it has to be ascertained that the FRFs to be used for the same are good. The checks are typically made using log-log Bodé plots of the measured receptance, mobility or inertance FRFs. The points to be checked are listed below.

(i) The characteristics of the FRF plot at very low frequencies, below the first natural frequency, should be checked to see if the features corresponding to the test boundary conditions are reflected. A fixed support condition should result in a stiffness-line asymptote characteristic at low frequencies, the magnitude of this stiffness corresponding to the static stiffness of the structure at the clamp. A free support condition, on the other hand, should result in a mass-line asymptote in this low-frequency region, with its magnitude deducible from rigid body considerations. If these low-frequency characteristics do not match the expected behaviour, it can be concluded that measurements have not been made at sufficiently low frequencies or that the achieved support condition deviates from what was aimed at.

(ii) Another verification can be made near the upper end of the desired frequency range. If the point mobility measurements show the curve becoming asymptotic to a mass line or a stiffness line (which is more common), it indicates that the excitation has been applied to a point of very high mass or flexibility. Under such circumstances, it would be desirable to obtain the FRFs again using a different excitation point.

(iii) A third set of checks can be made to see if antiresonances and minima between resonances occur as expected. For this, of course, the FRFs should show clear resonance and antiresonance characteristics in the first place. A point mobility plot should reveal an antiresonance after every resonance; a transfer mobility

plot between two points far apart on the structure should show more minima than antiresonances. Besides, the resonance peaks and antiresonance dips should exhibit the same degree of sharpness on a log-log plot; if not, the implication is that the quality of measurements is poor because the measured vibration levels are too low or because the chosen frequency resolution on the FFT analyzer is not fine enough. Measurements may have to be repeated in such a case also.

(iv) If the FRF representations are in the form of Nyquist plots, then it is to be checked if each resonant region traces out a circular arc. For a system with well-separated modes, each of these arcs will correspond to a major portion of a circle, while for closely spaced modes, the arcs will subtend only 45° or 60° at the centres of the circles. If the arcs are not smooth, it implies that the measurements are poor.

9.6.2   SDOF Modal Analysis

There exist a number of modal analysis methods such as peak pick method, circle fit method, inverse method, etc. All these are frequency domain methods in the sense that they operate on the FRF to get the modal parameters and all of them assume that in the region of resonance, the natural mode under consideration contributes to most of the response, the contributions from other modes being negligible. The methods vary as to whether they assume all the responses being attributed to that single mode or whether the contribution of other modes is represented by a simple approximation. There are time domain methods also and these operate on the impulse response function, but these are not discussed in this book.

9.6.2.1   Peak Pick Method

The simplest of these methods is called the 'peak amplitude' or 'peak pick' method. The assumption behind this method is that the FRFs of the structures considered have certain characteristics: (i) they have well-separated modes, (ii) the modes are not too lightly damped as to make it difficult to obtain accurate measurements around resonance, and (iii) the modes are not so heavily damped that the response at resonance has contributions from more than one mode. Though these characteristics limit the applicability of the method, it can be used to obtain initial estimates to the modal parameters. Subsequently, for obtaining better estimates, other curve fitting procedures may be used. The peak pick procedure works on one mode after another and is as follows:

(i) Frequencies corresponding to peak values of response in the FRF curve (Fig. 9.25a) are identified as being the natural frequencies of various modes. Let us denote the frequency of maximum response as the natural frequency ω_r of a particular mode.
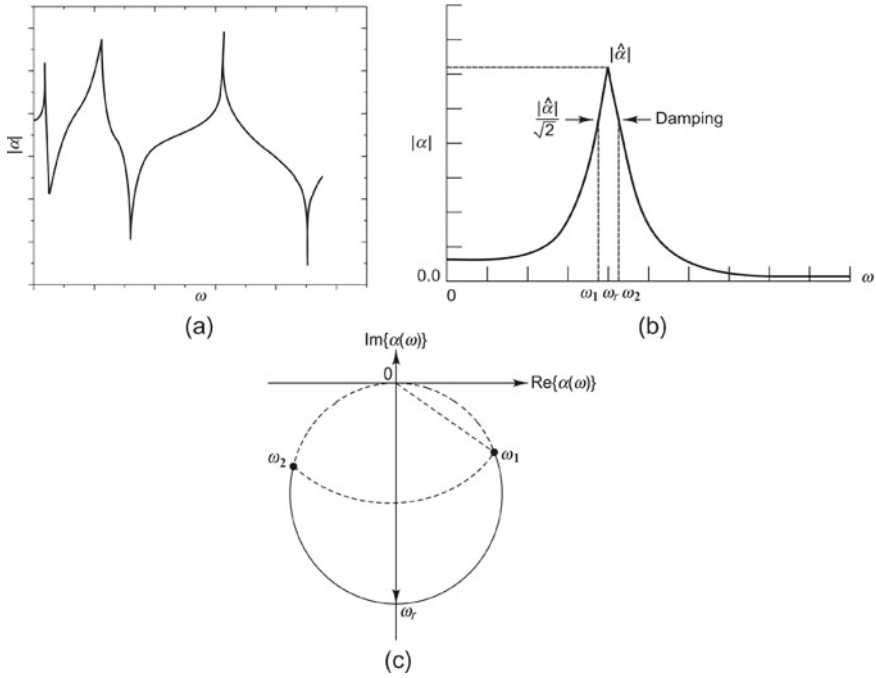
Fig. 9.25 Peak pick method of modal parameter extraction: **a** measured FRF plot, **b** determination of half-power frequencies, **c** Nyquist FRF plot

(ii) The peak value of the response corresponding to this mode is denoted as $|\hat{\alpha}|$. The half power or 3 dB frequencies ω_1 and ω_2 (Fig. 9.25b) with a response level of $|\hat{\alpha}|/\sqrt{2}$ are determined on either side of resonant frequency ω_r and the half-power bandwidth $\Delta\omega = \omega_2 - \omega_1$ is computed.

(iii) The damping of the mode under consideration can be got from the following formula:

$$\xi_r = \frac{(\omega_2^2 - \omega_1^2)}{\omega_r^2} \cong \frac{\Delta\omega}{\omega_r} \tag{9.45a}$$

$$\xi_r = 2\zeta_r \tag{9.45b}$$

Here ζ_r and ξ_r and the damping ratios corresponding to viscous damping and structural damping respectively of the rth mode.

(iv) An estimate of the frequency dependent modal constant of the mode being considered may be obtained based on the SDOF assumption that the total response in this frequency region is from only one term in the general FRF series representation.

$$\alpha_{jk}(\omega) = \sum_{r=1}^{N} \frac{(_r\psi_j)(_r\psi_k)}{\omega_r^2 - \omega^2 + i\xi_r\omega_r^2} = \sum_{r=1}^{N} \frac{_r A_{jk}}{\omega_r^2 - \omega^2 + i\xi_r\omega_r^2} \tag{9.46}$$

Here $_r A_{jk}$ is the modal constant corresponding to the response of the jth degree-of-freedom when the system is excited at the kth degree-of-freedom in the rth mode for a system with structural damping. Thus

$$|\hat{\alpha}| = \frac{A_r}{(\omega_r^2\xi_r)} \tag{9.47a}$$

or

$$A_r = |\hat{\alpha}|\omega_r^2\xi_r \tag{9.47b}$$

This method has certain shortcomings due to the assumptions made. The estimates of both damping and modal constant depend on the peak value of FRF level $|\hat{\alpha}|$. There is a lot of uncertainty in measurements around the resonance region, especially if measurements have been made with a poor frequency resolution and if the structure has lightly damped modes. This method ends up measuring the peak value based entirely on a single point in the FRF. Besides, this method gives estimates of real modal constants, implying that we are considering systems with real modes, or proportionally damped structures. In addition, the single-mode assumption is a major drawback leading to inaccuracy of the estimate. Even in systems with well-separated modes, it is often found that the neighbouring modes do contribute to the total response at the resonance under consideration. The circle fit method is a refinement of this technique and does away with these problems. Figure 9.25c shows a possible Nyquist plot which might give the modulus plot shown in Fig. 9.25b.

The above procedure may be applied to a plot of the real part of the receptance FRF instead of the modulus plot as shown in Fig. 9.26. This curve indicates that the positions and values of the maximum and minimum values yield good estimates of the half-power points. Besides, a more refined estimate of the natural frequency itself can be obtained as the point midway between the maximum and minimum on the imaginary plot as shown below.

$$A_r = (|MX| + |MN|)\omega_r^2\xi_r \tag{9.48}$$

9.6.2.2 SDOF Circle Fitting Method

We know that for a general SDOF system, a Nyquist plot of frequency response produces circle-like curves. We also know that specifically the receptance FRF gives a true circle with a structural damping model and the mobility FRF gives a circle with a viscous damping model (the circle is rotated anti-clockwise by 90° in the complex

Fig. 9.26 Real FRF plot

$Re\{\alpha(\omega)\}$
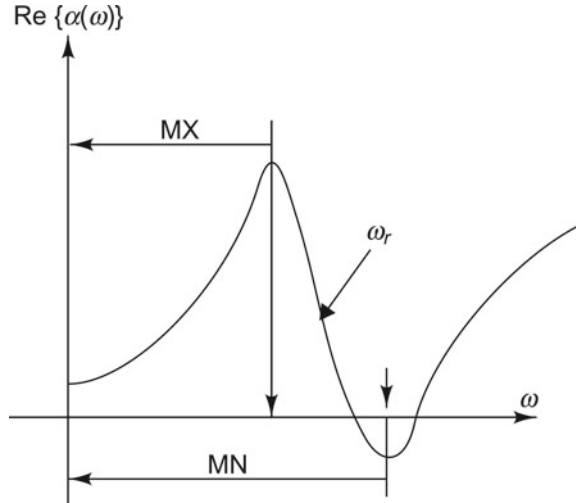
plane). Further, we have seen for MDOF systems that Nyquist plot of FRF data show sections of near-circular arcs in the regions around resonances. This feature provides the basis for the SDOF circle fitting method. The derivations in this section are based on a system with structural damping, and therefore, the receptance FRF data are used. Many of the modal analysis packages offer both types of damping and take the appropriate FRF representation for the circle fitting procedure.

As implied by the name, this method makes use of the fact that the response of a system around resonance is predominantly due to a single mode. This implies that the magnitude of the FRF is mainly due to one term in the series, i.e., due to the mode being analyzed. Equation (9.46) can be rewritten, without simplification, as

$$\alpha_{jk}(\omega) = \frac{rA_{jk}}{\omega_r^2 - \omega^2 + \mathrm{i}\xi_r\omega_r^2} + \sum_{s=1,\neq r}^{N} \frac{sA_{jk}}{\omega_s^2 - \omega^2 + \mathrm{i}\xi_s\omega_s^2} \qquad (9.49)$$

By virtue of making the SDOF assumption, it can be said that over a small range of frequencies around natural frequency ω_r of mode r, the frequency dependence of the second term in Eq. (9.49) is negligible and hence it simplifies to

$$\alpha_{jk}(\omega)_{\omega=\omega_r} = \frac{rA_{jk}}{\omega_r^2 - \omega^2 + i\xi_r\omega_r^2} +_r B_{jk} \qquad (9.50)$$

This is illustrated in Fig. 9.27. The figure shows each of the two terms in Eq. (9.50), as well as the total receptance over the same frequency range. It is seen that the first term (contribution of the mode being analyzed) varies considerably through the resonance region, sweeping out a circle, while the second term (contribution of all
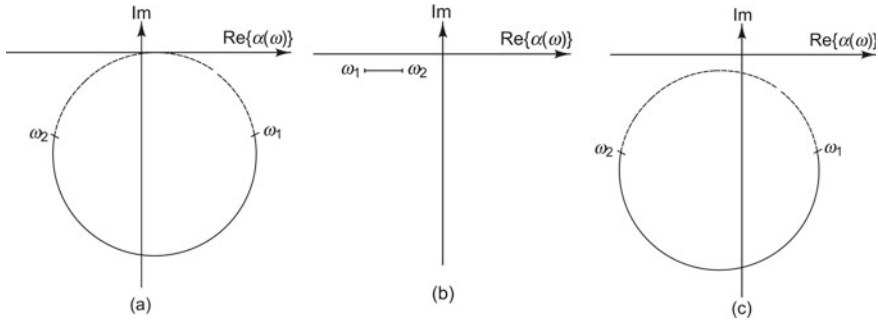
Fig. 9.27 Nyquist FRF plot

other modes) is a constant, independent of frequency. Hence, the total receptance is a circle with the same properties as that of the mode under consideration, but displaced from the origin of the Argand plane by the constant contributed to by all the other modes.

Properties of the Modal Circle with Structural Damping The relevant properties of the modal circle which are useful in extracting the required modal parameters are described in this section. The basic FRF is of the form

$$\alpha = \frac{1}{\omega_r^2[1 - (\omega/\omega_r)^2 + i\xi_r]} \tag{9.51}$$

$$\tan \gamma = \frac{\xi_r}{[1 - (\omega/\omega_r)^2]} \tag{9.52a}$$

$$\tan(90° - \gamma) = \tan\left(\frac{\theta}{2}\right) = \frac{[1 - (\omega/\omega_r)^2]}{\xi_r} \tag{9.52b}$$

If the modal constant $_rA_{jk}$ is included in the numerator of Eq. (9.51), it scales the size of the circle by $|_rA_{jk}|$ and rotates it by $\angle_rA_{jk}$ so that the principal diameter (the one passing through the natural frequency) is oriented at an angle $\angle_rA_{jk}$ to the negative imaginary axis. From Eq. (9.52b), we obtain

$$\omega^2 = \omega_r^2\left[1 - \xi_r \tan\left(\frac{\theta}{2}\right)\right] \tag{9.53}$$

Differentiating Eq. (9.53) with respect to θ, we get

$$\frac{d\omega^2}{d\theta} = \left(\frac{-\omega_r^2\xi_r}{2}\right)\left\{1 + \frac{[1 - (\omega/\omega_r)^2]^2}{\xi_r^2}\right\} \tag{9.54}$$
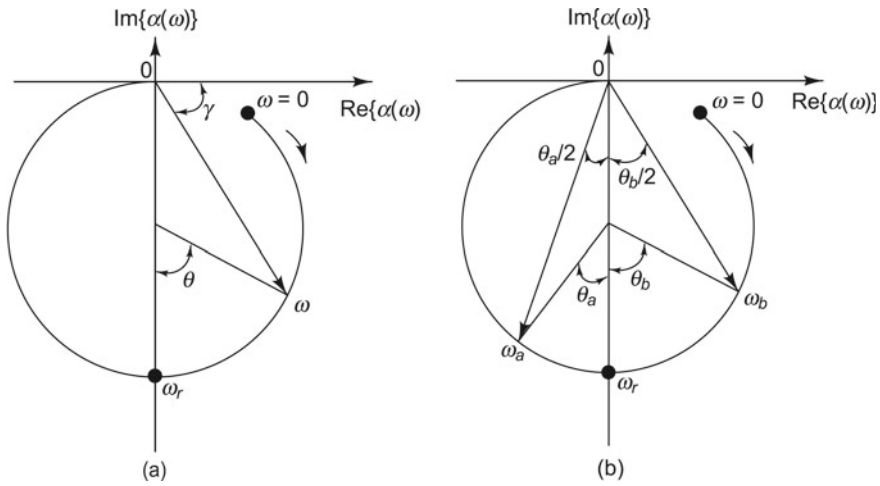
Fig. 9.28 Modal circle: **a** geometry, **b** details around resonance

The reciprocal of this quantity is a measure of the rate at which the locus sweeps around the circular arc; the maximum sweep rate occurs when $\omega = \omega_r$. Thus

$$\left(\frac{d}{d\omega}\right)\left(\frac{d\omega^2}{d\theta}\right) = 0 \quad \text{when} \quad \left(\omega_r^2 - \omega^2\right) = 0 \tag{9.55}$$

The sweep rate parameter provides an estimate of damping since

$$\left(\frac{d\theta}{d\omega^2}\right)_{\omega=\omega_r} = \frac{-2}{\left(\omega_r^2 \xi_r\right)} \tag{9.56}$$

This equation gives a hint as to where exactly the natural frequency is; the relative spacing of the measured data points around the circular arc near each resonance helps determine its value. In Fig. 9.28, let ω_b and ω_a correspond to two frequencies on the modal circle, ω_b being below the natural frequency and ω_a being above. Then

$$\tan\left(\frac{\theta_b}{2}\right) = \left[\frac{1 - (\omega_b/\omega_r)^2}{\xi_r}\right]; \quad \tan\left(\frac{\theta_a}{2}\right) = \left[\frac{(\omega_a/\omega_r)^2 - 1}{\xi_r}\right] \tag{9.57}$$

From these two equations, the exact equation for damping of the mode is

$$\xi_r = \frac{\left(\omega_a^2 - \omega_b^2\right)}{\{\omega_r^2[\tan(\theta_a/2) + \tan(\theta_b/2)]\}} \tag{9.58a}$$

For light damping (say, less than 2–3%), the expression becomes

$$\xi_r = \frac{2(\omega_a - \omega_b)}{\{\omega_r[\tan(\theta_a/2) + \tan(\theta_b/2)]\}} \tag{9.58b}$$

For the two half-power points for which $\theta_a = \theta_b = 90°$, we get

$$\xi_r = \frac{(\omega_2 - \omega_1)}{\omega_r} \tag{9.59a}$$

If the damping is not light, we get

$$\xi_r = \frac{(\omega_2^2 - \omega_1^2)}{\omega_r^2} \tag{9.59b}$$

Finally, the diameter of the circle is got as $1/(\omega_r{}^2\xi_r)$. If this equation is scaled by a modal constant in the numerator, the diameter of the circle is

$$_r D_{jk} = \frac{|_r A_{jk}|}{(\omega_r^2 \xi_r)} \tag{9.60}$$

Properties of the Modal Circle with Viscous Damping With viscous damping, mobility FRF gives a true circle and hence should be used instead of receptance. Thus

$$Y(\omega) = \frac{i\omega}{(k - \omega^2 m) + i\omega c} \tag{9.61}$$

From the real and imaginary parts of Eq. (9.61), we get

$$\tan\left(\frac{\theta}{2}\right) = \frac{\omega(k - \omega^2 m)}{\omega^2 c} = \frac{[1 - (\omega/\omega_r)^2]}{(2\zeta_r\omega/\omega_r)} \tag{9.62a}$$

$$\tan\left(\frac{\theta_b}{2}\right) = \frac{[1 - (\omega_b/\omega_r)^2]}{(2\zeta_r\omega_b/\omega_r)}; \quad \tan\left(\frac{\theta_a}{2}\right) = \frac{[(\omega_a/\omega_r)^2 - 1]}{(2\zeta_r\omega_a/\omega_r)} \tag{9.62b}$$

These equations give

$$\zeta_r = \frac{(\omega_a^2 - \omega_b^2)}{2\omega_r[\omega_a \tan(\theta_a/2) + \omega_b \tan(\theta_b/2)]} \tag{9.63a}$$

For lightly damped structures

$$\zeta_r \cong \frac{(\omega_a - \omega_b)}{\omega_r[\tan(\theta_a/2) + \tan(\theta_b/2)]} \tag{9.63b}$$

For the half-power frequencies $\theta_a = \theta_b = 90°$, and therefore, we have

$$\zeta_r = \frac{(\omega_2 - \omega_1)}{2\omega_r} \tag{9.63c}$$

This is an exact equation independent of the level of damping.

Circle Fitting Procedure The objective behind this analysis procedure is to extract the necessary coefficients in Eq. (9.46), and from them the modal parameters. The following five-step procedure is described for the case of structural damping. A similar approach is used for viscous damping also.

(i) *Selection of points to be used:* A fixed number of points (at least six) should be chosen on either side of any identified resonance in the FRF. If the system is user-interactive in nature, the analyst may use his/her discretion to separate true modes from spurious perturbations and to reject suspect data points. The selected points should not be influenced much by adjacent modes. It is preferable to have a circular arc of at least 180° and, if possible, more.

(ii) *Fitting a circle and calculating the quality of fit:* This is usually done using one of a host of curve fitting algorithms. It requires finding a circle with the minimum least-squares deviation for the selected points. The most accurate way of doing this is to minimize the deviations of the measured points from their actual locations on the circle. Another possible method involves minimizing the deviations of points from the nearest point on the circle and is not as accurate; nevertheless, it is more common since it is easier to implement. The centre and radius of the circle are obtained at the end of this exercise, along with an error which is nothing but the mean square deviation of the chosen points from the circle. These deviations are typically of the order of 1–2%.

(iii) *Locating natural frequencies and obtaining damping estimates:* One of the most common definitions used for the natural frequency in order to locate it is the frequency at which the sweep rate around the circle reaches a maximum. To determine this, radial lines are constructed from the centre of the circle to the points around resonance and the angles they subtend with each other are noted, using which the sweep rate through the region is estimated and thence the natural frequency. If, the frequency points used are spaced at regular intervals, then a finite difference method may be used. Other definitions of the natural frequency are also used; they are: (i) the frequency of maximum response, (ii) the frequency of maximum imaginary receptance, and (iii) the frequency of zero real receptance.

(iv) *Calculation of multiple damping estimates and scatter:* A number of damping estimates are generally calculated using all combinations of selected data points, two at a time with one on either side of resonance (Eq. 9.63a). It is a common practice to evaluate the mean, as well as the deviation of the estimates. A deviation of less than 4–5% indicates a successful analysis, but if the scatter is 20 or 30%, the values used should be re-examined. Random variations in damping estimates could be due to measurement errors; systematic variations, on the other hand, could be due to a poor experimental setup, contributions

from neighbouring modes, non-linear behaviour of the system, etc. Such errors will not get averaged out.

(v) *Determination of modal constant:* This involves finding the magnitude and phase of the modal constant from the diameter of the circle and from its orientation in the Argand plane. This calculation is straightforward once the natural frequency has been located and the damping estimates obtained.

Once the entire circle fitting analysis procedure is over, a theoretically regenerated FRF plot may be obtained and compared with the original measured data. At this juncture, it will be necessary to add the contribution of the other modes to the resonant mode under consideration. This contribution ($_r B_{jk}$ in Eq. 9.50) may be measured off as the distance from the top of the principal diameter to the origin. Then, using this equation and the modal parameters extracted, a curve based on the model obtained may be plotted.

9.6.2.3   SDOF Modal Analysis Using Inverse Method

The inverse method is based on the fact that a function, which generates a circle when plotted in the complex plane, will trace out a straight line when plotted as a reciprocal (Fig. 9.29). Thus, if we were to plot the reciprocal of the receptance of an SDOF system with structural damping, it would produce a straight line as can be seen from the expressions below.

$$\alpha(\omega) = \frac{(k - \omega^2 m) - ih}{(k - \omega^2 m)^2 + h^2} \tag{9.64a}$$

and
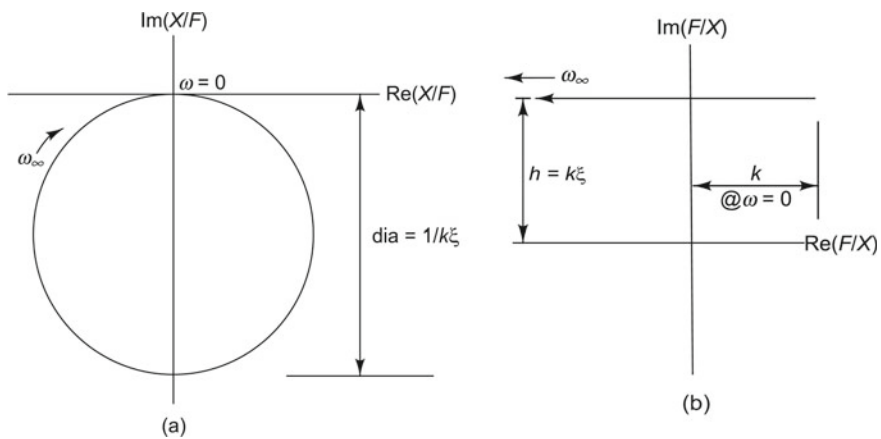
Fig. 9.29 Nyquist plots of receptance: **a** standard receptance, **b** inverse receptance

$$\frac{1}{\alpha(\omega)} = (k - \omega^2 m) + \mathrm{i}h \qquad\qquad (9.64b)$$

The procedure is as follows. The inverse FRF is first computed and a best-fitting straight line is constructed through the data points. The intercept of this line with the imaginary axis gives an estimate for the damping parameter. The deviations of the data points from the line itself give an indication of the reliability of the estimate. If the points are randomly scattered on either side of the line, it is possible that there are experimental errors, but if the deviations are systematic, then there is a source of bias in the data and this needs investigation. A least-squares operation indicating the deviation between the real part of the measured data points and that of the theoretical model gives estimates for the mass and stiffness parameters in the theoretical model. This method gives the best results for systems with real modes and relatively well-separated modes; corrections have to be applied if the FRF has contributions from more than one mode. A very attractive feature of this method is that the straight line can readily be obtained with data points far away from resonance and not necessarily exactly at the natural frequency (where the real part of the inverse receptance is zero). The method described here for structural damping can be extended to the case of viscous damping as before.

9.6.3 MDOF Curve Fitting Methods

The SDOF approach to modal analysis is inappropriate under certain circumstances such as when the system has closely coupled modes or extremely light damping; this makies measurements at resonance inaccurate and difficult. Under these conditions, alternative methods called MDOF curve fits may be used. By closely coupled modes are meant those modes for which the response at resonance is not due to just one mode (or term in the FRF series), because the natural frequencies are very closely spaced, or because they have relatively heavy damping, causing overlapping of modes.

A regenerated FRF obtained by repeated application of any of the SDOF curve fitting procedures around each of the resonances in the frequency range of interest reveals, on comparison with the FRF from original measurements, that the match is poor, especially in regions away from resonances. This is because the regenerated curve has taken into account the frequency range f_1 to f_2 in computing modes m_1 to m_2, whereas the actual measured curve has contributions from modes in the frequency ranges below f_1 and above f_2 also. Thus, the regenerated curve can be corrected by accounting for the effects of these modes also. This is done by including two simple extra terms called residuals.

Consider the case of an MDOF system with structural damping where the FRF is regenerated using a formula of the form

$$\alpha_{jk}(\omega) = \sum_{r=m_1}^{m_2} \frac{{}_r A_{jk}}{\omega_r^2 - \omega^2 + i\xi_r\omega_r^2} \tag{9.65}$$

Here modes from m_1 to m_2 alone have been included. But the fact remains that the modes outside this frequency range affect the measured FRFs. Hence, the equation most closely representing the measured data is

$$\alpha_{jk}(\omega) = \sum_{r=1}^{N} \frac{{}_r A_{jk}}{\omega_r^2 - \omega^2 + i\xi_r\omega_r^2} \tag{9.66a}$$

This may be rewritten as

$$\alpha_{jk}(\omega) = \sum_{r=1}^{m_1-1} + \sum_{r=m_1}^{m_2} + \sum_{r=m_2+1}^{N} \left(\frac{{}_r A_{jk}}{\omega_r^2 - \omega^2 + i\xi_r\omega_r^2} \right) \tag{9.66b}$$

The first term in this equation corresponds to the low-frequency modes and the third term to the high-frequency modes, both of which are not included in the regenerated FRF. It is only the second term which is included. Figure 9.30a–d show the contribution of all modes in the range 1 to N, modes from $m_2 + 1$ to N, modes from m_1 to m_2, and those from 1 to $m_1 - 1$, respectively.

From these figures, it is clearly seen that the contribution of modes 1 to $m_1 - 1$ is that of a mass-like behaviour, while the contribution of modes $m_2 + 1$ to N is seen as a stiffness-like behaviour and these effects may be accounted for in the regenerated FRF using residual mass and stiffness terms as shown below.

$$\alpha_{jk}(\omega) = -\frac{1}{\omega^2 {}_M R_{jk}} + \sum_{r=m_1}^{m_2} \left(\frac{{}_r A_{jk}}{\omega_r^2 - \omega^2 + i\xi_r\omega_r^2} \right) + \frac{1}{{}_K R_{jk}} \tag{9.67}$$

Here, ${}_M R_{jk}$ and ${}_K R_{jk}$ represent the residual mass and stiffness terms, respectively, for the frequency range considered. If the frequency range of analysis is changed, these terms also change. The regenerated curve is compared with the measured FRF at the lowest frequencies and from the difference between the two, ${}_M R_{jk}$ is obtained. In a similar fashion, the regenerated and actual curves are compared at the highest frequencies and the difference gives ${}_K R_{jk}$. This process is refined iteratively in case the addition of the stiffness residual upsets the effectiveness of the mass term and the addition of the mass term upsets that of the stiffness term.

There are many different methods available for MDOF modal parameter extraction. In this section, three frequency domain MDOF curve fitting methods are described. All of them are similar, in that, they perform a curve fit to the entire FRF measurement in one step. As for the SDOF cases, the structurally damped system is described and may be extended to the viscously damped case. The three approaches are as follows:

(i) an extension of the SDOF method outlined in Sect. 9.6.2.1,
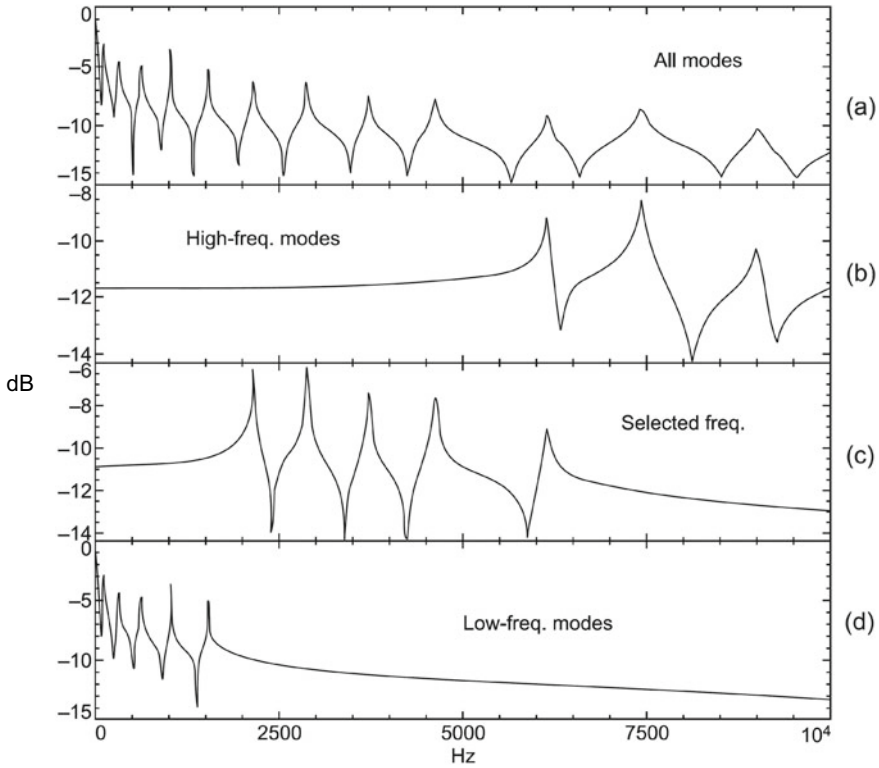
Fig. 9.30 Receptance FRF of MDOF system: **a** superposition of all modes, **b** modes corresponding to high-frequency range, **c** modes corresponding to identified frequency range, **d** modes corresponding to low-frequency range

(ii) a general approach to multimode curve-fitting, and
(iii) a method meant for very lightly damped structures.

9.6.3.1 Extension of SDOF Method

In the SDOF circle fitting modal analysis procedure, it was assumed that near the resonance under consideration, the effect of all the other modes could be represented by a constant. The present method is an extension of that assumption, enabling a more accurate analysis of the data. The receptance FRF in the frequency range of interest may be written taking into account the residual mass and stiffness terms as

$$\alpha_{jk}(\omega) = \sum_{s=m_1}^{m_2} \frac{{}_s A_{jk}}{\omega_s^2 - \omega^2 + i\xi_s\omega_s^2} + \frac{1}{{}_K R_{jk}} - \frac{1}{\omega^2 {}_M R_{jk}} \tag{9.68a}$$

This can be rearranged as

$$\alpha_{jk}(\omega) = \frac{_rA_{jk}}{\omega_r^2 - \omega^2 + i\xi_r\omega_r^2} + \sum_{s=m_1,\neq r}^{m_2} \frac{_sA_{jk}}{\omega_s^2 - \omega^2 + i\xi_s\omega_s^2} + \frac{1}{_KR_{jk}} - \frac{1}{\omega^2 {_M}R_{jk}}$$

(9.68b)

The first term gives the contribution of the rth mode, the second term gives the contribution of all modes from m_1 to m_2 save the rth mode, the third term gives the contribution of all high-frequency modes that have been left out of the FRF regeneration process, and the fourth term gives the contribution of all the left out low-frequency modes. The main difference between the SDOF circle fitting procedure and the present MDOF method is that while the second term was assumed to be a constant in the former curve fitting procedure, it is considered as frequency dependent in the present method. To simplify the curve fitting procedure, the magnitude of the coefficients in the second term in Eq. (9.68b) can be got from estimates which have already been obtained from an SDOF analysis. Let $\alpha_{jk}^m(\omega)$ represent a set of measured FRF data points around the resonance at ω_r. We can now regenerate FRFs at each frequency for which we have a measured FRF value. The difference between the two can be used as correction factors to get adjusted FRF data points which should then represent true SDOF behaviour as shown below.

$$\alpha_{jk}^m(\omega) - \left(\sum_{s=m_1}^{m_2} \frac{_sA_{jk}}{\omega_s^2 - \omega^2 + i\xi_s\omega_s^2} + \frac{1}{_KR_{jk}} - \frac{1}{\omega^2 {_M}R_{jk}} \right) = \frac{_rA_{jk}}{\omega_r^2 - \omega^2 + i\xi_r\omega_r^2}$$

(9.69)

This procedure should be repeated iteratively for all the modes in the frequency range of interest till convergence is reached. The method offers a drastic improvement in cases where the modes are strongly coupled, though this is not true for systems with relatively weakly coupled modes.

9.6.3.2 General MDOF Curve Fitting Procedure

The previous method was an extension of the SDOF analysis method. The general MDOF curve fitting methods requiring sophisticated computational methods are newer developments in the field. Let an individual FRF measured data be represented as

$$\alpha_{jk}^m(\Omega_n) = \alpha_n^m \tag{9.70}$$

The corresponding regenerated values are denoted by

$$\alpha_{jk}(\Omega_n) = \alpha_n = \sum_{s=m_1}^{m_2} \frac{_sA_{jk}}{\omega_s^2 - \Omega_n^2 + i\xi_s\omega_s^2} + \frac{1}{_KR_{jk}} - \frac{1}{\Omega_n^2 {_M}R_{jk}} \tag{9.71}$$

Here, the coefficients $_1A_{jk}, _2A_{jk}, \ldots, \omega_1, \omega_2, \ldots, \xi_1, \xi_2, \ldots, _KR_{jk}$, and $_MR_{jk}$ are to be found. Let ε_n denote an individual error as

$$\varepsilon_n = (\alpha_n^m - \alpha_n) \tag{9.72}$$

The total error can be expressed as

$$E_n = |\varepsilon_n^2| \tag{9.73}$$

A weightage w_n may be assigned to each frequency in the frequency range of interest. The curve fitting procedure has to be done such that the total error computed as shown below is minimized.

$$E = \sum_{n=1}^{P} w_n E_n \tag{9.74}$$

This may be done by differentiating Eq. (9.74) with respect to each unknown in turn, leading to as many equations as unknowns, many of which are non-linear in the ω_s and ξ_s parameters. Hence, there are many algorithms for solving these equations, each with its own simplifications and assumptions and each gives rise to a different procedure.

The theory and experimental techniques discussed in this chapter will come in handy while conducting the three experiments on modal analysis discussed in Chap. 10.

List of Symbols

Symbol	Meaning	SI unit
$a(t)$	acceleration	m/s^2
A	amplitude of acceleration a(t)	m/s^2
$_rA_{jk}$	modal constant $= (_r\psi_j)(_r\psi_k)$ of rth DOF	kg^{-1}
c	viscous damping coefficient	N.s/m
$f, f(t)$	force	N
f_k	force at kth DOF	N
f_r	natural frequency of rth mode	Hz
F	amplitude of force $f(t)$	N
h	complex stiffness	N/m
$H(\omega)$	FRF	
$H_1, H_2, H_3, H_1(f),$ $H_2(f), H_3(f)$	FRF estimators	

Symbol	Meaning	SI unit
H_c, H_v	FRF estimators	
H_{ij}	element in ith row and jth column of FRF matrix	
H_0, $H_0(f)$	true FRF	
i	$\sqrt{-1}$	
Im$\{\alpha\}$	imaginary part of receptance	m/N
k	stiffness	N/m
m	vibrating mass	kg
$_M R_{jk}$	residual mass	kg
$m(t)$	uncorrelated noise in output signal	
N	number of DOF	
$n(t)$	uncorrelated noise in input signal	
r	mode number	
$_K R_{jk}$	residual stiffness	N/m
Re$\{\alpha\}$	real part of receptance	m/N
s	driving signal	
$S_{\max}$	maximum allowable sweep rate (Hz/minute or octaves/minute)	
S_{nn}	auto-spectral density of random measurement noise in input	
S_{mm}	auto-spectral density of random measurement noise in output	
S_{xs}	cross-spectral density of reference signal with actuating signal	
$S_{xx}(\omega)$	auto-spectral density of excitation	
$S_{xy}(\omega)$	cross-spectral density between input and output signals	
S_{ys}	cross-spectral density between output and actuating signals	
$S_{yx}(\omega)$	cross-spectral density between output and input signals	
$S_{yy}(\omega)$	auto-spectral density of output	
S_{uv}	cross-spectral density between true input and true output signals	
S_{uu}	auto-spectral density of true input signal	
S_{vv}	auto-spectral density of true output signal	
$[S_{xx}(\omega)]$	$N \times N$ matrix of auto-spectral densities of different input DOFs	
$[S_{xy}(\omega)]$	$N \times N$ matrix of cross-spectral densities between inputs and outputs	
t	time	s
T	duration of time	s
$u(t)$	true input	
$U(f)$	Fourier transform of $u(t)$	
v_m	velocity of mth DOF	m/s
$v(t)$	velocity	m/s
$v(t)$	true output	
V	amplitude of velocity $v(t)$	m/s

Symbol	Meaning	SI unit
$V(f)$	Fourier transform of $v(t)$	
x	input signal	
x_j	displacement of jth DOF	m
$\{\dot{x}\}$	$N \times N$ vector of velocities	m/s
$x(t)$	displacement	m
$x(t)$	input signal measured by FFT analyzer	
X	amplitude of displacement $x(t)$	m
$X(f)$	Fourier transform of $x(t)$	
X_j	input spectrum at the jth DOF	
y	output signal	
$y(t)$	output signal measured by FFT analyzer	
Y_i	output spectrum at ith DOF	
$Y(f)$	Fourier transform of $y(t)$	
$Y_{jk}(\omega)$	element of mobility matrix for velocity of jth DOF due to force at kth DOF	m/s.N
$[Y(\omega)]$	$N \times N$ mobility matrix	m/s.N
$[Z(\omega)]$	$N \times N$ impedance matrix	N.s/m
α	constant	
$|\hat{\alpha}|$	peak value of response corresponding to a mode	m
α_{jk}	element of receptance matrix for displacement of jth DOF due to force at kth DOF	m/N
$\alpha_{jk}^m(\Omega_n)$	element of measured receptance matrix for displacement of jth DOF due to force at kth DOF at frequency Ω_n	m/N
$\alpha(\omega)$	receptance	m/N
β	constant	
γ	phase angle	rad
γ_{xy}^2	coherence function	
ζ	viscous damping ratio	
ζ_r	viscous damping ratio of rth mode	
θ_A	phase lag of accelerance	rad
θ_Y	phase lag of mobility	rad
θ_α	phase lag of receptance	rad
ξ	structural or hysteretic damping ratio	
ξ_r	structural or hysteretic damping ratio of rth mode	
ϕ	angle of lag	rad
$(_r\psi_j)$	mass-normalized displacement of jth DOF in rth mode	kg^{-1}
ω	circular frequency or excitation frequency	rad/s
ω_r	natural frequency of rth mode	rad/s
ω_1 and ω_2	lower and upper half-power frequencies	rad/s
ω_b and ω_a	two frequencies below and above a natural frequency	rad/s

Abbreviations

ADC	analogue-to-digital converter
DOF	degrees-of-freedom
FE	finite element
FFT	fast Fourier transform
FRF	frequency response function
IRF	impulse response function
ISO	International Organization for Standardization
LDV	laser Doppler vibrometer
MDOF	multi-degree-of-freedom
MIMO	multiple-input multiple-output
MRIT	multiple reference impact testing
PDF	probability density function
PRN	periodic-/ pseudo-random noise
PSD	power spectral density
RMS	root mean square
SDOF	single-degree-of-freedom
SIMO	single-input multiple output
SISO	single-input single-output

Questions

1. What is meant by modal mass and modal stiffness?
2. Who first observed the vibration of mode shapes by placing sand on vibrating plates?
3. What is meant by mode superposition technique? Where is it used to advantage?
4. What is a Bodé diagram?
5. How can you get decoupled equations of motion for an MDOF system?
6. What do you mean by orthogonality of modes?
7. Which are the mass controlled, stiffness controlled, and damping controlled regions of a vibration frequency response function?
8. What is the significance of the coherence function?
9. How can you find damping in an SDOF system?
10. What precautions should be taken in swept sine excitation?
11. What do you mean by operational modal analysis?
12. When do you use the 'zoom' option in modal testing?
13. What is the effect of inclusion of damping in the FRF data of MDOF systems?
14. How can you simulate a free boundary condition in the laboratory?
15. Give examples of non-intrusive exciters for modal testing.
16. What transducer can be used for the measurement of driving point FRF?

17. Which FRF estimator is contaminated by output noise. Why?
18. What are the problems associated with impact excitation?
19. What is the need for tips of different materials with impact excitation?
20. What are the advantages of impact excitation?
21. What is meant by modal parameter extraction?
22. Which modal parameter extraction methods are easier to implement, time domain or frequency domain methods? Why?
23. How can you obtain one row of the FRF matrix from the experiment?
24. What is meant by SDOF curve fitting procedure?
25. Outline a procedure for obtaining the mode shape corresponding to the second mode from FRFs.
26. How can you differentiate between point and transfer FRFs?
27. How can you find out the stiffness of a structure from the mobility plot?
28. Write the expression for one element of the receptance matrix of an MDOF system with structural damping.
29. What is the main problem in modal analysis of lightly damped structures?
30. Write the expression for one element of the receptance matrix of an MDOF system with viscous damping.
31. How can you obtain one column of the FRF matrix from experiments?
32. What decides whether an FRF will have minima or antiresonances?
33. How can you obtain damping from a Nyquist plot of an SDOF system with structural damping?
34. What are the differences between real and complex modes?
35. Which graphical representation of FRF data is applicable to both damped and undamped systems?
36. What are the problems associated with modal parameter extraction when there are heavily damped modes?
37. What is the major problem associated with modal analysis of structures with closely spaced modes?
38. Why are antiresonances seen in FRFs?
39. How many distinct natural frequencies can exist for an N-degree-of-freedom system?
40. What is the attractive feature of a Nyquist plot in modal analysis?
41. What is the major difference between a viscously damped and structurally damped system?
42. How can we find the response of a multi-degree-of-freedom system using the first few modes only?
43. When do you get repeated modes?
44. What is an impedance head? Where is it used?
45. When do we get complex eigenvalues?
46. How can you find out the stiffness of a structure from the receptance plot?
47. What is the advantage of a proportional damping model in the analysis of an MDOF system?
48. What do you mean by step relaxation technique?
49. What is a mode shape? How is it computed?

50. What are forward and inverse FRFs?
51. What is a pseudo-random signal? What is its advantage in modal testing?
52. Define these terms: proportional damping and modal damping ratio.
53. What is meant by modal participation factor?
54. What is meant by modal truncation?
55. What is meant by residuals in modal testing?
56. What is the advantage of narrow band random noise as compared to wide band random noise as an excitation signal?
57. How do you decide on exciter and pickup locations in modal testing?
58. What is a standard eigenvalue problem?
59. How do you ensure a good signal-to-noise ratio in modal testing?
60. What are the main differences between time and frequency domain modal parameter extraction methods?
61. What are receptance, mobility, and accelerance FRFs?
62. What is the difference between the theoretical and experimental routes to modal analysis?
63. What is meant by complex stiffness?
64. Why is damping said to be a model of convenience?
65. Why is the matrix inversion method of finding FRFs difficult?
66. What do you mean by dynamical matrix? What is its use?
67. What is a Frequency Response Function?
68. What is an Impulse Response Function?
69. How are IRF and FRF related?
70. Give some examples of dampers.
71. What is meant by the circle fit method?
72. What is the advantage of the inverse method of modal parameter extraction?
73. Describe three methods by which the natural frequency of a system may be identified.
74. What are auto- and cross-spectral densities?
75. How many elements of an NXN FRF matrix are required for modal parameter extraction?
76. What are the ambiguities in theoretical modal analysis?

Fill in the Blanks

77. The high-frequency modes not included in the experimental modal analysis may be accounted for using a _____ residual term.
78. Certain modes get missed out in an FRF. This implies that _____.
79. The ways in which FRF parameters can be graphically represented are _____.
80. To excite a mechanical system in the range 0–800 Hz, an impact hammer with a _____ tip is preferable.
81. The _____ FRF estimator is contaminated by input noise. Why?

82. For a system with the first two natural frequencies 3.5 and 4.2 Hz and corresponding damping ratios of 0.01 and 0.02, the Rayleigh damping constants are _____.
83. The _____ FRF estimator can be used to minimize input and output noise. Why?
84. To excite a mechanical system in the range 0–10000 Hz, an impact hammer with a _____ tip is preferable. Why?
85. The Nyquist plot of receptance of a system is symmetrical about the y-axis. This implies that _____.
86. To correlate analytical findings regarding the dynamic behaviour of a system with experimental findings, the preferred support conditions are _____. Why?
87. The low-frequency modes not included in the experimental modal analysis may be accounted for using a _____ residual term.

State whether TRUE or FALSE and Explain Why

88. It is experimentally easy to obtain both forward and inverse FRFs.
89. At resonance, the receptance FRF is imaginary.
90. It is always desirable to measure the point FRFs in addition to a row or column of the FRF matrix.
91. The best way of mathematically representing a typical element of the FRF matrix, from a modal analysis perspective, is as the ratio of polynomials.
92. Receptance term $\alpha_{jk}(\omega) = 1/K_{jk}(\omega)$ where $K_{jk}(\omega) =$ dynamic stiffness term.
93. The impedance FRF may very conveniently be found out from modal testing.
94. The H_2 FRF estimator overestimates the FRF.
95. To get the peak amplitude at resonance, the real part of the FRF should be used.
96. The mass, stiffness, and damping matrices of a multi degree-of-freedom system are always symmetric.

Bibliography

1. Agilent Technologies, Fundamentals of Modal Testing, Application Note 243-3.
2. Allemang, R. J., Rost, R. W., & Brown, D. L. (1983). Dual input estimation of frequency response functions for experimental modal analysis of aircraft structures. In *Proceedings of the 1st International Modal Analysis Conference*, Orlando, Florida, 333–340.
3. Avitabile, P. (2017). *Modal testing: A practitioner's guide*. Wiley.
4. Brincker, R., & Ventura, C. (2015). *Introduction to operational modal analysis*. Wiley.
5. Clough, R. W., & Penzien, J. (1993). *Dynamics of structures*. New York: McGraw Hill Book.
6. Conciauro, G., Guglielmi, M., & Sorrentino, R. (2000). *Advanced modal analysis*. New York: Wiley.
7. Craig, R. R., & Kurdila, A. J. (2006). *Fundamentals of structural dynamics*. Wiley.
8. de Silva, C. W. (2007). *Vibration damping, control, and design*. Boca Raton: CRC Press.
9. de Silva, C. W. (2007). *Vibration monitoring, testing, and instrumentation*. Boca Raton: CRC Press.

10. Døssing, O., Structural Testing. Part 1: Mechanical Mobility Measurements, Brüel & Kjær Theory and Application Booklet, BR 0458-12, Denmark.
11. Døssing, O., Structural Testing. Part 2: Modal Analysis and Simulation, Brüel & Kjær Theory and Application Booklet, BR 0507-11, Denmark.
12. de Silva, M., Júlio, M., & Maia, N. M. M. (1999). Modal analysis and testing. In *Proceedings of the NATO Advanced Study Institute*, Sesimbra, Portugal, 3–15 May, 1998, Series: NATO Science Series E (Vol. 363). Springer.
13. Elliott, K. B., & Mitchell, L. D. (1984). The improved frequency response function and its effect on modal circle fits. *ASME Journal of Applied Mechanics, 51*, 657–663.
14. Ewins, D. J. (1993). *Dynamic testing agency handbook on guidelines to best practice: Modal testing* (Vol. 3). Dynamic testing agency, United Kingdom.
15. Ewins, D. J. (2003). *Modal testing: Theory and practice*. England: Research Studies Press Ltd.
16. Fang, F. Z., & Xing, H. H. (2000). *Modal analysis theory and applications*. China Press.
17. Gade, S., Herlufsen, H., & Hansen, H. K., How to Determine the Modal Parameters of Simple Structures, Brüel & Kjær Application Note, Bo042, Brüel & Kjær, Denmark.
18. Ginsberg, J. H. (2001). *Mechanical and structural vibrations: Theory and applications*. New York: Wiley.
19. He, J., & Fu, Z. F. (2001). *Modal analysis*. Oxford: Butterworth-Heinemann.
20. Herlufsen, H. (1984). Dual Channel FFT Analysis. Parts 1 & 2, Brüel & Kjær Technical Reviews No. 1 & 2, BV0013-11 & BV 0014-11, Denmark.
21. Herlufsen, H. (1985). Modal Analysis using Multi-Reference and MIMO Techniques, Brüel & Kjær Application Note BT 0001-12, Denmark.
22. Inman, D. J. (2001). *Engineering vibration*. Upper Saddle River: Prentice Hall.
23. Inman, D. J. (2006). *Vibration with control*. Chichester: Wiley.
24. International Organization for Standardization. (1986). ISO 7626-1:1986, Vibration and Shock-Methods for the Experimental Determination of Mechanical Mobility.
25. Maia, N. M. M., & e Silva, M. (1997). *Theoretical and experimental modal analysis*. Somerset: Wiley.
26. McConnell, K. G., & Varoto, P. S. (1995). *Vibration testing: Theory and practice*. New York: Wiley-Interscience.
27. Mitchell, L. D. (1982). Improved methods for the fast fourier transform (FFT) calculation of the frequency response function. *ASME Journal of Mechanical Design*, *l04*, 277–279.
28. Patton, M. E., & Trethewey, M. W. (1987). A technique for non-intrusive modal analysis of very lightweight structures. In *Proceedings of the 5th International Modal Analysis Conference,* April 6–9, Imperial College of Science, London, England.
29. Patton, M. E., & Trethewey, M. W. (1987). A survey and assessment of non-intrusive-modal-testing techniques for ultralight weight structures. *International Journal of Analytical and Experimental Modal Analysis*, *2*(4), 163–173.
30. Rainieri, C. (2014). *Operational modal analysis of civil engineering structures*. Springer.
31. Rocklin, G. T., Crowley, J., & Vold, H. (1985). A comparison of H_1, H_2 and H_v frequency response functions. *Proceedings of IMAC III*, *l*, 272–278.
32. Rossing, T. D., & Fletcher, N. H. (2004). *Principles of vibration and sound*. New York: Springer.
33. Slater, J. C. (2007). *Vibration testing, with modal testing and health monitoring*. New York: Wiley.
34. Thomson, W. T. (1998). *Theory of vibration with applications*. USA: Prentice-Hall.
35. Zaveri, K., Modal Analysis of Large Structures-Multiple Exciter Systems, Brüel & Kjær, BT 0001-12, Denmark.
36. http://literature.agilent.com/litweb/pdf/5954-7957E.pdf. The Fundamentals of Modal Testing, Hewlett Packard Application Note 243-3, USA.

Chapter 10
Vibration Experiments

This chapter describes seventeen basic vibration experiments which can be conducted in the laboratory to understand the basic vibration theory described in Chap. 2. Suggestions are given for the choice of transducers or alternate experimental setups.

10.1 Free Vibration (Translation) Using Impact Excitation

INTERESTING FACTS: Free vibrations of skyscrapers were observed in Japan and in Taiwan during the earthquake in Tohoku, Japan, on 11th March 2011, with some buildings swaying for a duration of 10 to 30 min due to their low inherent damping. Vibration spectra recorded by seismic transducers in TAIPEI 101, the second tallest building in the world clearly showed the fundamental mode of vibration at 0.15 Hz, as well as higher mode vibrations. Magnification factors of ground accelerations for the fundamental modes of vibrations in the building to those at a nearby borehole station were found to be as high as 110 and 146, respectively, on the 74th and 90th floors above ground. The frequency content of the accelerograms recorded below the ground was found to be in a frequency band of 0.015–0.1 Hz, clearly proving it was free vibration and not forced excitation due to ground motion which caused high vibration in the TAIPEI 101 skyscraper.

Aim: To determine the first natural frequency and viscous damping ratio of a cantilever beam through a free vibration test.

Fig. 10.1 Single degree of
freedom damped system
response
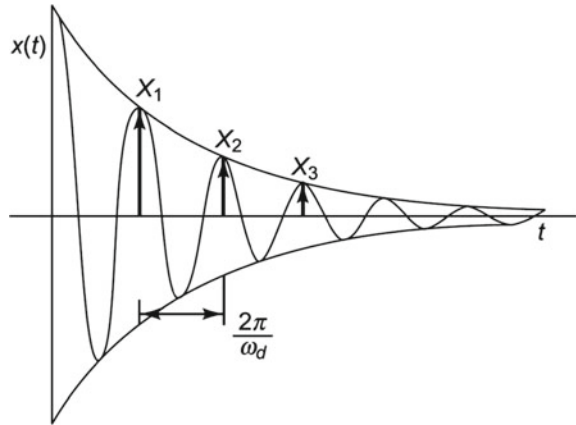
Equipment: Aluminium beam, non-contact electromagnetic velocity transducer
with ferromagnetic strip, storage oscilloscope/recorder, rigid clamp.

Theory: Free vibration takes place when an elastic system not acted upon by any
steady excitation force is disturbed from its mean equilibrium position. When a
system is allowed to vibrate freely, vibrations die out gradually over a few cycles of
motion because the energy imparted to the system initially is dissipated during the
motion. The damping in most mechanical systems is so small that it practically has no
influence on the natural frequency of the system. The theory related to free vibration
with viscous damping is described in Sects. 2.2.4 and 2.2.4.1 and the response of a
single degree of freedom (SDOF) damped system is as shown in Fig. 10.1.

The displacement from the mean equilibrium position at any point of time is given
by the equation

$$x(t) = Xe^{-\zeta \omega_n t} \sin\left(\sqrt{1 - \zeta^2}\omega_n t + \phi\right) \tag{10.1}$$

where

$x(t)$ = amplitude of vibration at any point of time
X = maximum amplitude of vibration
ζ = damping ratio = c/c_c
ω_n = natural frequency without damping
ω_d = natural frequency with damping = $\omega_n\sqrt{1 - \zeta^2}$
t = instant of time considered
ϕ = phase angle
c = damping coefficient
c_c = critical damping coefficient = $2m\omega_n$
m = mass of the system

The time period of vibration (with damping) is given by

$$T_d = \frac{2\pi}{\omega_d} \tag{10.2}$$

From this equation, the natural frequency can be calculated. For most mechanical systems, ζ lies in the range 0.001 to 0.05 and therefore $\omega_d \approx \omega_n$; hence while calculating the natural frequency, we can neglect damping. Considering two successive amplitudes X_1 and X_2 at times t_1 and $t_1 + T_d$, we have

$$X_1 = X e^{-\zeta \omega_n t_1} \sin\left(\sqrt{1 - \zeta^2} \omega_n t_1 + \phi\right) \tag{10.3}$$

$$X_2 = X e^{-\zeta \omega_n (t_1 + T_d)} \sin\left(\sqrt{1 - \zeta^2} \omega_n (t_1 + T_d) + \phi\right) \tag{10.4}$$

Since they are exponential functions, it is convenient to define a logarithmic decrement in two successive amplitudes as

$$\delta = \ln\left(\frac{X_1}{X_2}\right) = \zeta \omega_n T_d = \frac{2\pi \zeta}{\sqrt{1 - \zeta^2}} \tag{10.5}$$

$$\delta \approx 2\pi \zeta$$

The logarithmic decrement can also be defined as

$$\delta = \frac{1}{n} \ln\left[\frac{X_1}{X_{n+1}}\right] \tag{10.6}$$

Calculation of δ using Eq. (10.6) gives less error than that using Eq. (10.5) which involves two successive amplitudes.

Test Setup and Procedure: One end of the test beam is firmly clamped between two massive steel blocks and the other end is free as shown in Fig. 10.2. The electromagnetic pickup is fixed close to the free end of the beam. The rap test consists of simply tapping the beam by hand or by an impulse hammer and storing the response from the vibration pickup on the oscilloscope. One may also use a data acquisition system with a computer for recording the signal. From the decaying vibration response, the period of oscillation is noted and the natural frequency is obtained. The damping ratio may be obtained from Eqs. (10.5) and (10.6) by noting amplitudes X_1 and X_{n+1} which are n cycles apart.

Transducers Used: The principle of operation of an electromagnetic transducer is explained in Sect. 3.4.1. This particular transducer has been suggested for this experiment since it is a very inexpensive transducer of the non-contact type and
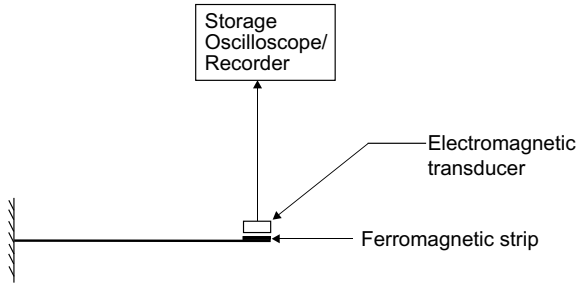
Fig. 10.2 Test setup

Table 10.1 Results of free vibration experiment

No.	Time for 10 oscillations (s)	Time for 1 oscillation, $T(s)$	$f = 1/T$ (Hz)	Calculated f_n (Hz)	n	X_1 (V)	X_{n+1} (V)	δ	ζ
1.									
2.									

does not need any additional signal conditioning amplifier. Its output voltage may be straightaway fed to a storage oscilloscope or recorder. Alternately, a non-contact capacitance transducer as described in Sect. 3.3.2 may be used with a displacement measuring unit. One could also use a miniature piezoelectric transducer with minimal mass loading along with a charge amplifier. If one is interested only in the natural frequency and damping ratio and not in the amplitude of vibration, a simple strain gauge pasted at the free end will serve the purpose. This will have to be used with a carrier frequency amplifier.

Analytical Validation: The experimental results obtained may be verified from the closed form expression for the natural frequencies of beams given (Sect. 2.5.3) by

$$\omega_n = (\beta_n l)^2 \sqrt{\frac{EI}{\rho l^4}} \tag{10.7}$$

Here, ρ is the mass per unit length of the beam, EI is the flexural rigidity, l is the length, n is the mode number and β_n depends on the boundary conditions of the beam. For a beam in cantilever configuration, $(\beta_1 l)^2 = 3.52$ and $(\beta_2 l)^2 = 22$. Results may be tabulated as shown in Table 10.1.

Questions

(1) Will the natural frequencies and damping ratios change for other boundary conditions?

(2) How would you get damping ratio in the case of structural damping?
(3) What would be the discrepancy in damping ratio got from the exact and approx-
     imate formulae in Eq. (10.5)?
(4) Why is the transducer kept near the free end?

10.2 Forced Vibration (Translation) Using Stepped Sinusoidal Excitation

DID YOU KNOW that the Broughton suspension bridge across the River
Irwell between Broughton and Pendelton near Manchester, England, suffered
a catastrophic collapse due to resonance? Five years into its use in 1831, as 4
columns of soldiers marched 'in time' across the bridge, their synchronized
footsteps induced resonance, causing the bridge to bounce increasingly with
their deliberate, playfully timed marching. When the soldiers reached the far
end of the bridge, a part near the other end collapsed, throwing around 40 men
into the thankfully shallow river. It seems the resonance they set up by their
synchronized marching caused high stresses in a major component, causing
it to break suddenly. Unfortunately, the troops realized that the bouncing
resonance due to their footsteps was fatal only after the structure broke and
collapsed. Consequent to the incident, the British Army issued an order that
troops should 'break step' when crossing a bridge.

Aim: To determine the first two natural frequencies and corresponding damping
ratios of a beam supported at both ends by pinned joints and subjected to transverse
vibration.

Theory: The theory behind the forced vibration response of an SDOF system with
viscous damping is discussed in Sect. 2.2.6.2. If an SDOF system having natural
frequency ω_n is excited by a harmonic force $x(t) = A \cos \omega t$, then the response is

$$y(t) = A \cos(\omega t - \phi) = A H(\omega) \cos \omega t \tag{10.8}$$

where $H(\omega)$ is the dynamic magnification factor with magnitude given by

$$|H(\omega)| = \frac{1}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \tag{10.9}$$

Here η = frequency ratio ω/ω_n and ζ = damping ratio.

At resonance $|H(\omega)| = Q = 1/2\zeta$ (for small values of ζ). The ratios η_1 and η_2,
for which the amplitude of $|H(\omega)|$ is $A_{\max}/\sqrt{2}$ as shown in Fig. 10.3 are called
half-power frequency ratios and it can be shown that
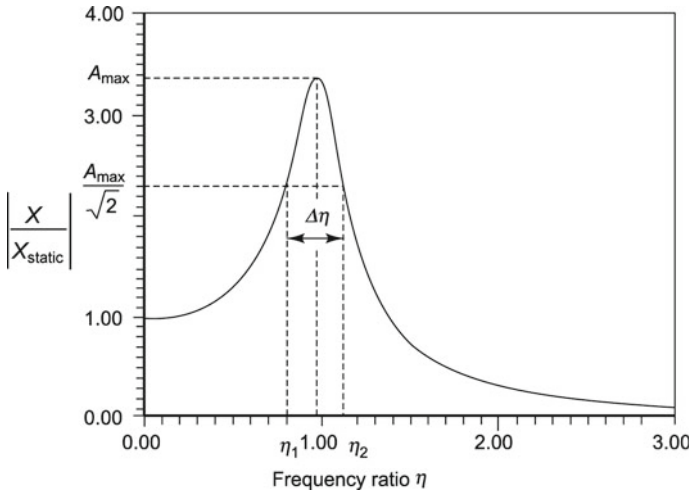
Fig. 10.3 Forced vibration response

$$Q = \frac{1}{2\zeta} = \frac{\omega_n}{\omega_2 - \omega_1} \tag{10.10}$$

where ω_2 and ω_1 are the frequencies corresponding to η_2 and η_1.

Experimental Setup and Test Procedure A schematic of the test setup is shown in Fig. 10.4. A discrete sinusoidal input with a fixed amplitude and frequency is given by the function generator to the electrodynamic exciter (please refer Sect. 3.10.1) to excite the beam at a specific location, ensuring that the vibration of the beam is not driven into the non-linear range. The exciter excites the beam through a slender rod called a stinger. If the output current supplied by the function generator is small, it may have to be amplified by a power amplifier before being fed into the exciter coil. Care should be taken to ensure that the coil is not overdriven. The electrodynamic exciter is the most common type of shaker. The excitation is called stepped sine since the command signal frequency is stepped from one discrete value to another in the entire frequency range such as to provide the necessary frequency resolution in the output or frequency response function (FRF) plot. The excitation frequency should not be changed abruptly; sufficient time should be given for the system to settle down to the steady-state condition. Especially near resonance, or if one is measuring lightly damped or closely spaced modes, care should be taken to avoid transient effects. The major advantage of stepped sine excitation is that it gives flexibility in choosing the frequency resolution. This method has the advantage that one can measure with a high frequency resolution around resonance and lower resolution away from resonance. This technique is useful for large, complex structures and for cases with high background noise.

1. Capacitance probe   2. Electrodynamic shaker   3. Beam support
4. Pinned-pinned beam   5. Displacement measuring unit 6. RMS voltmeter
7. Oscilloscope   8. Function generator   9. Pinned joint
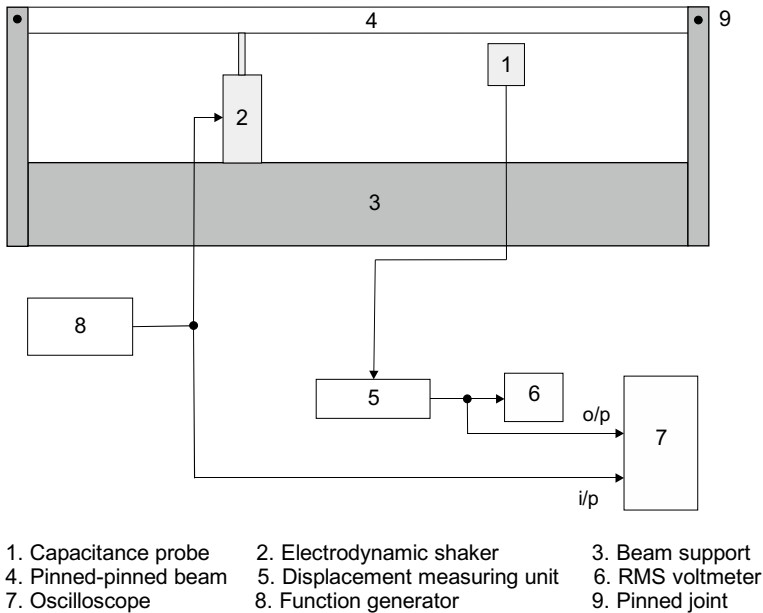
Fig. 10.4   Test setup

The vibration is sensed by a non-contact type capacitance pickup, nodal locations of the first two modes of vibration being avoided for excitation as well as sensing. One could also use a miniature piezoelectric accelerometer with built-in charge amplifier or a non-contact electromagnetic pickup (with a ferromagnetic strip pasted at the target footprint in the case of a non-magnetic beam). The change in vibration amplitude is converted into a proportional voltage by a displacement measuring unit and the voltage is measured using an RMS voltmeter. The exciter input and the capacitance pick up output are connected to a dual-channel oscilloscope.

The frequency of excitation is varied using the function generator and the output voltage for each frequency is noted. An RMS voltmeter may also be used since it gives readings in terms of dB. The voltage values at the half-power frequencies would be 3 dB below the value measured at the natural frequency. The frequency is varied in small steps near the resonance (with a fine frequency resolution of 0.1 Hz or so). A graph of voltage versus frequency is plotted. The half power frequencies are identified and the damping ratio is calculated. The procedure is repeated for the second mode. Tabulation may be done as shown in Table 10.2. The dimensions of the beam are to be noted for the analytical validation.

Analytical Validation: The experimental results obtained may be verified from the closed form expression for the natural frequencies of beams (Sect. 2.5.3) given by

Table 10.2 Results of forced vibration experiment

	Mode 1				Mode 2			
No.	f_n (Hz)	f_1 (Hz)	f_2 (Hz)	$\zeta = \frac{f_2-f_1}{2f_n}$	f_n (Hz)	f_1 (Hz)	f_2 (Hz)	$\zeta = \frac{f_2-f_1}{2f_n}$
1.								
2.								

$$\omega_n = (\beta_n l)^2 \sqrt{\frac{EI}{\rho l^4}} \tag{10.11}$$

Here, ρ is the mass per unit length of the beam, EI is the flexural rigidity, l is the length, n is the mode number, and β_n depends on the boundary conditions of the beam. For a simply supported beam, $(\beta_1 l)^2 = 2$ and $(\beta_2 l)^2 = 39.5$.

Questions

(1) What will happen if the stinger does not make contact with the beam?
(2) Will the natural frequencies and damping ratios change for other boundary conditions?
(3) How do you decide on the force input to the beam?
(4) How will the first natural frequency of an aluminium beam compare with that of a mild steel beam having the same dimensions? Why?
(5) What is the difficulty in measuring the damping ratio of a lightly damped mode?
(6) Why should you avoid nodal points for placement of the exciter or transducer?
(7) Why is there a discrepancy between the measured and computed values of natural frequencies?

10.3   Forced Vibration Experiment with Swept Sine Excitation

INTERESTING FACTS: A sine sweep signal is one in which the frequency of a sinusoidal wave increases or decreases with time, i.e., sweep up or sweep down. The term sine sweep is also used interchangeably with chirp (chirp burst typically having a smaller duration than a sweep), so we have up-chirp or down-chirp. It has wide applications in communications. The name chirp is used since the signal on being fed to a loudspeaker sounds like the chirping sound made by birds. In the time domain, a chirp signal evolves as a frequency modulated tone.

Aim: To determine the first two natural frequencies and corresponding damping ratios of a beam supported at both ends by pinned joints and subjected to transverse vibration.

Theory: The basic theory behind forced vibration using swept sine excitation is the same as that using stepped sine excitation as described in Sect. 10.2.

Experimental Setup and Procedure: The test setup is the same as shown in Fig. 10.4. The swept sinusoidal input with a fixed amplitude is fed by the sweep oscillator or function generator to the electrodynamic exciter as before. An automatic sweep command is given to sweep the frequency continuously from one value to another in the required frequency range. The sweep rate should not be high; sufficient time has to be given for the system to settle down to the steady-state conditions around each resonance. In order to identify the optimum sweep rate, Eq. (9.1) or (9.2) in Sect. 9.2.5.2 may be used. Alternately, a sweep-up command may be given initially, followed by a sweep-down command over the same frequency range. With the correct sweep rate, the peaks in the output response or FRF should occur at the same frequencies in both cases. If a peak is shifted to the right during a sweep up, it implies that the rate is higher than it should be as shown in Fig. 10.5.

If a large number of natural frequencies are to be identified, then initially a fast sweep may be done to roughly identify the modes. Subsequently, a slow sweep may be adopted around each mode. This method reduces the test time for structures where a number of modes have to be identified. The raw voltage output or the FRF may be captured on a storage oscilloscope or on a recorder. Figure 10.6 shows a typical output in the time domain and a magnitude FRF plot in the frequency domain. The damping ratio should be computed from the frequency domain plot using the half-power method as before. Results may be plotted as shown in Table 10.2.

Fig. 10.5 Effect of sweep rate on FRF
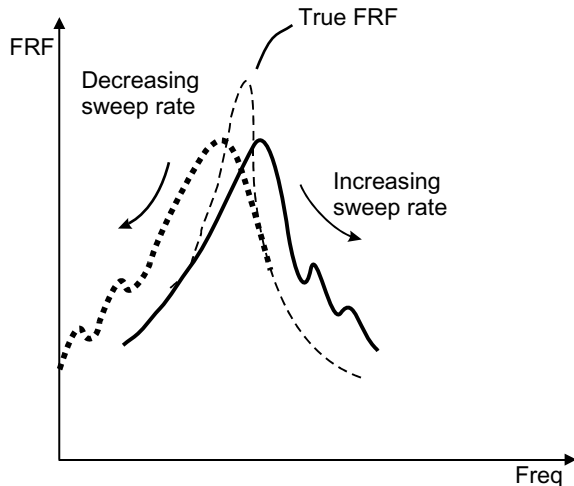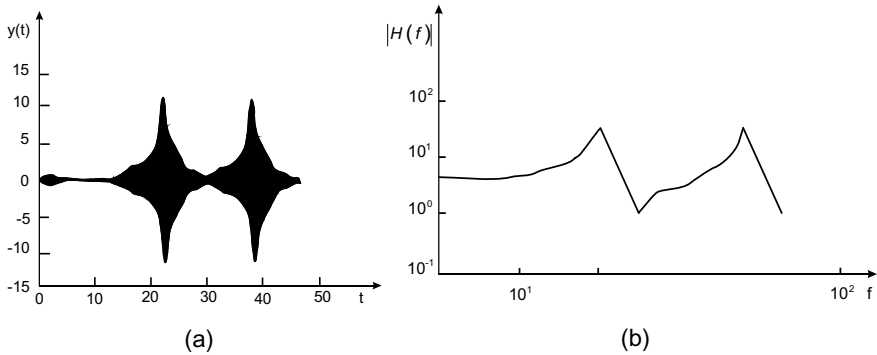
Fig. 10.6 Typical outputs with swept sine excitation: **a** time domain output, **b** magnitude FRF

Analytical Validation: The experimental results obtained may be verified from the closed form expression for the natural frequencies of beams given by Eq. 2.137 as before.

Questions

(1) Why is the expression for sweep rate dependent on damping?
(2) When would you use a linear sweep rate and when a logarithmic sweep rate?
(3) Do you think swept sine excitation has any drawback as compared to stepped sine excitation?

10.4 Forced Vibration Using Random Excitation

DID YOU KNOW that white noise is a random signal with equal intensity at all frequencies, and thus a constant power spectral density (PSD)? This signal is used in many fields of science and engineering, including physics, acoustics, telecommunications, statistical forecasting, etc. A white noise generator produces an acoustic signal that is random in nature. This sounds like a rushing waterfall or wind blowing through trees. Often these generators do not produce ideal white noise, which has a wide frequency range and a harsh sound, but pink noise, which has essentially low frequency content and whose power rolls off at higher frequencies. They may also produce noise of other colours.

Aim: To determine the first two natural frequencies and corresponding damping ratios of a beam supported at both ends by pinned joints and subjected to transverse vibration.

Theory: The basic theory behind forced vibration using random excitation is the same as that using stepped sine excitation as described in Sect. 10.2.

Experimental Setup and Test Procedure: The test setup is the same as shown in Fig. 10.4. True random signal is the most popular signal for shaker excitation since the measurement time is minimal. This signal may be obtained from a signal generator or random noise generator, which is commercially available or may be synthesized using a random number generation algorithm in a data acquisition system. Instead of white noise, band-limited random noise may be used since this allows the exciter energy to be concentrated in a small frequency range for improved SNR, especially when only a few modes are being studied. If required, the peak to RMS ratio (crest factor) and RMS value can be controlled. While testing, the true random signal is sent out to the shaker and the beam's response is measured simultaneously. During the next time period, a different random excitation record (with the same RMS level and crest factor) is sent out. Testing may be done with multiple random signals and linear averaging of the FRF may be done. Figure 10.7 shows a typical random input and output. The damping ratio should be computed using the half-power method as before. A table similar to 10.2 may be made for tabulation of results.

Analytical Validation: The experimental results obtained may be verified from the closed form expression for the natural frequencies of beams given by Eq. 2.137 as before.

Questions

(1) What decides the energy level in the random signal?
(2) How can you increase the frequency resolution of the signal?
(3) What is the need for averaging?
(4) How does the SNR of white noise compare with that of stepped sine excitation?
(5) When is pink noise preferred?
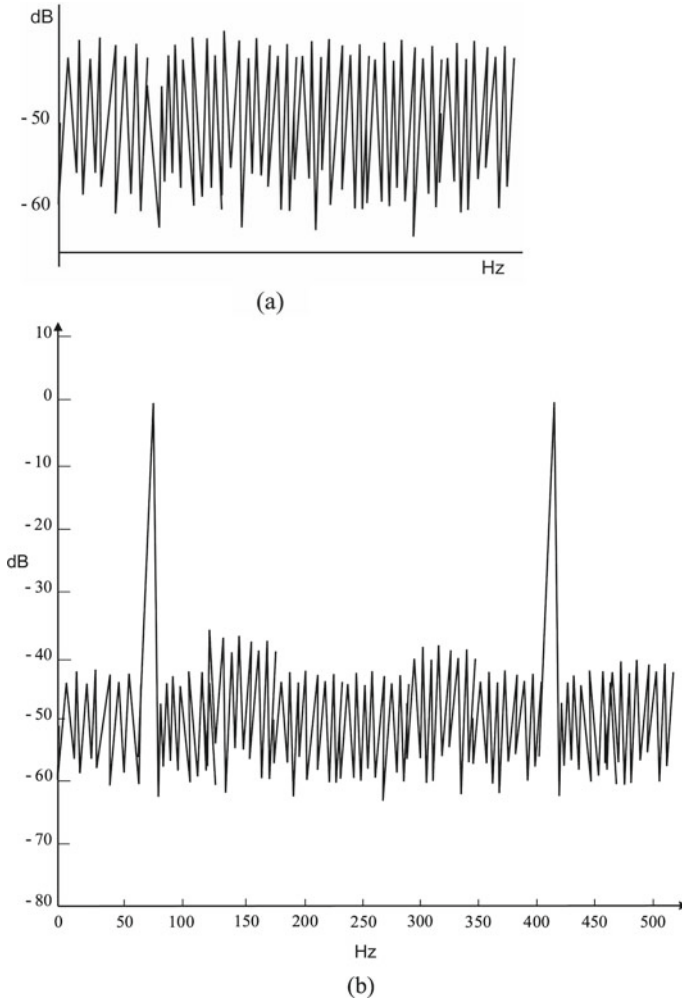
Fig. 10.7 Typical random input and output PSDs: **a** input, **b** output

10.5   Free Vibration (Torsion) and Determination of Mass Moment of Inertia-Bifilar Pendulum

INTERESTING FACTS: For two bodies with the same mass, the one where the mass is far away from the axis will have a larger moment of inertia as compared to the one where the mass is closer. Thus, a hollow cylinder has a larger moment of inertia than a solid cylinder of the same mass. We all have an intuitive feel for masses, but not so for mass moments of inertia. Did you know that the moments of inertia of a top, bicycle wheel and the earth about their spinning axes are 2.5×10^{-3} kg m^2, 0.09 to 0.15 kg m^2 and 8.04×10^{37} kg m^2?

Aim: To determine the moment of inertia of a rod by subjecting it to torsional oscillations.

Bifilar Pendulum Theory: A bifilar (2-filament) pendulum consists of a symmetric object (such as a uniform rod) suspended from two parallel filars or cords as shown in Fig. 10.8.

Such a bifilar suspension is used to determine the moment of inertia of the object about an axis of rotation parallel to the filars or cords. Suppose that the object has an
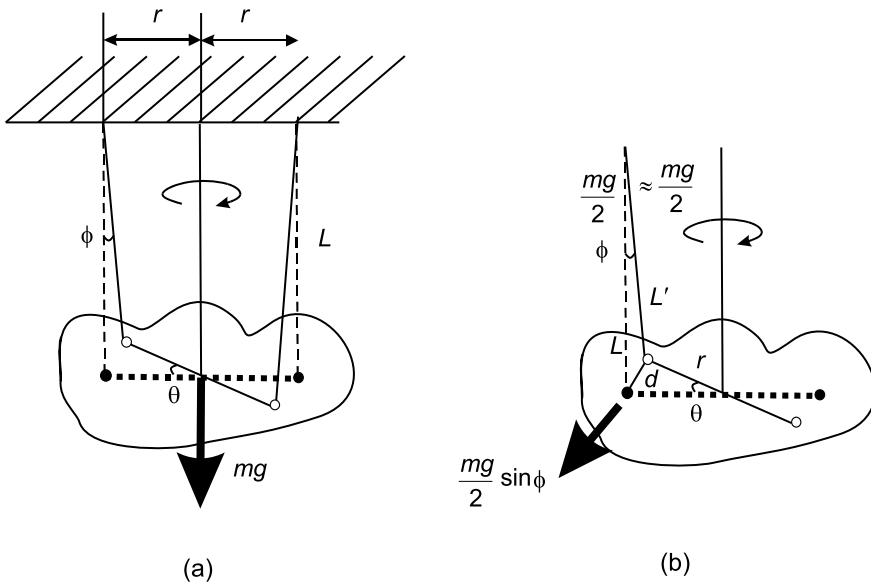
Fig. 10.8   Bifilar pendulum: **a** schematic drawing, **b** geometry

overall length $2r$ and is suspended by two filaments each of length L. It is desirable to have the cord length L to be much greater than $2r$, maybe 3 to 4 times $2r$. The object is to be set into torsional motion in the horizontal plane by giving it a small twist θ, such that there are no unbalanced vertical forces. The tension in each cord will then be

$$\frac{mg}{2} \cos \phi \approx \frac{mg}{2} \tag{10.12}$$

Here, ϕ is the small angle in the vertical plane due to the twist. From the geometry, it is clear that

$$L \sin \phi = r \sin \theta \tag{10.13}$$

The length d is the displacement in the horizontal plane and for a small twist, $\theta \approx d/r$. When the twisted pendulum is released, the restoring torque comes from the horizontal component of the tension $(mg/2) \sin \phi$ with a moment arm r about the axis of rotation. Therefore, the overall torque acting on the rod is

$$T = 2 \left(\frac{mg}{2} \right) r \sin \phi = mgr \sin \phi \tag{10.14}$$

Newton's second law for the angular motion is given by

$$T = J\alpha = J \frac{d^2\theta}{dt^2} \tag{10.15}$$

where J is the moment of inertia of the rod and $\alpha = d^2\theta/dt^2$ is the angular acceleration. From the geometry of the horizontal and vertical triangles, we have

$$d = L' \sin \phi \approx L \sin \phi \tag{10.16}$$

where L' is the twisted length. Since ϕ is small, $L' \approx L$. From the horizontal triangle

$$d = r \tan \theta \approx r \sin \theta \approx r\theta \tag{10.17}$$

From Eqs. (10.16) and (10.17), we have

$$\sin \phi \approx \left(\frac{r}{L} \right) \theta \tag{10.18}$$

Substituting in Eq. (10.14) and comparing with Eq. (10.15), we get

$$\frac{d^2\theta}{dt^2} + \left(\frac{mgr^2}{JL} \right) \theta = 0 \tag{10.19}$$

On comparison of Eq. (10.23) with that of simple harmonic motion, it is clear that the time period of the bifilar pendulum is given by
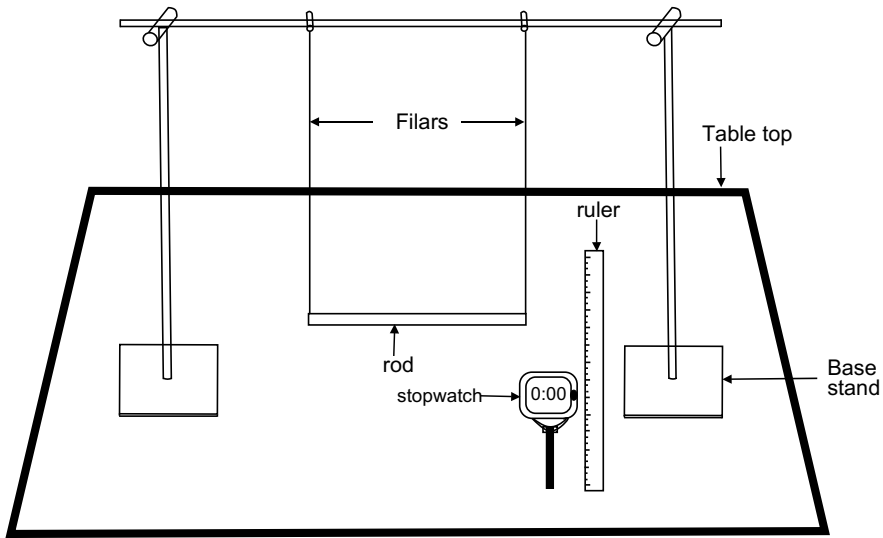
Fig. 10.9 Bifilar suspension setup

Table 10.3 Results of bifilar pendulum experiment

Trial no.	Time for 10 oscillations (s)	Time for 1 oscillation (s)	Measured moment of inertia J $(kg\, m^2)$	Calculated moment of inertia $(kg\, m^2)$
1.				
2.				

$$T = \frac{2\pi}{r} \sqrt{\frac{LJ}{mg}} \tag{10.20}$$

Test Setup and Procedure: The test setup is as shown in Fig. 10.9. Initially, the length and weight of the rod are to be measured. The rod is to be suspended so that its centre of gravity is symmetric with respect to the points of suspension. The rod is then to be given a gentle angular motion so that it oscillates in the horizontal plane. The time period of oscillation may be obtained by measuring the time taken for ten oscillations using a stopwatch. Once this is known, the moment of inertia may be calculated using Eq. (10.20). At least three such measurements may be made. Table 10.3 lists the mass and length measurements for the setup. Additionally, the moment of inertia is calculated by knowing the geometry of the rod. The experiment may be repeated with an object of irregular shape.

Questions

(1) Why should L be much greater than $2r$?

(2) What will the error be if the centre of gravity is not symmetric with respect to the points of suspension?
(3) What will the error be for an irregularly shaped object?
(4) When is it advantageous to make measurements with a trifilar pendulum?

10.6 Determination of Effective Radius of Gyration of a Body through Torsional Vibration of a Trifilar Pendulum

INTERESTING FACTS: Estimates of moments of inertia of a complicated structure like an aircraft may not be accurate since it is difficult to account for each and every small component. Experimental validation is therefore highly desirable. The techniques which were originally used were adaptations of classical laboratory test rigs, but with additional complexity due to handling issues arising from the size and weight of the aircraft. The setup conventionally used for the determination of the moment of inertia in pitch is to support the aircraft on knife edges at two rear jacking points, so that it can oscillate about a horizontal axis parallel to its pitching axis. The nose of the aircraft is made to rest in a wooden cradle, which is suspended on vertical coil springs. The arrangement for measuring the roll moment of inertia is very similar to that used for determining the pitch moment of inertia, but in the former there are no support points for the fuselage fore and aft centre line which can serve as a roll axis. The aircraft is hence mounted in a cradle with the fuselage supported on bulky wooden frames coinciding with the positions of load-bearing frames. The cradle, in turn, is supported on a pair of knife edges such that the entire assembly of cradle and aircraft is free to rotate about a longitudinal axis below the aircraft. Coil springs between the floor and each wing at about midspan position facilitate roll motion. In yaw inertia rigs the aircraft is pivoted about a knife-edge fulcrum or is suspended from a single torsionless cable, and is constrained by coil springs.

Introduction: A trifilar pendulum has a platform in the form of a circular or triangular disc, suspended by three inextensible strings of equal length as shown in Fig. 10.10. Both the top support and the platform are maintained horizontal. The wires are fixed to the platform 120° apart from each other. The object whose moment of inertia is to be determined is placed symmetrically on the platform. Let W_t be the total weight of the object and platform in N, W_p the weight of the platform, r the radius of the point of attachment of wires on the platform in m, l the length of each wire supporting the

Fig. 10.10 Trifilar
suspension setup
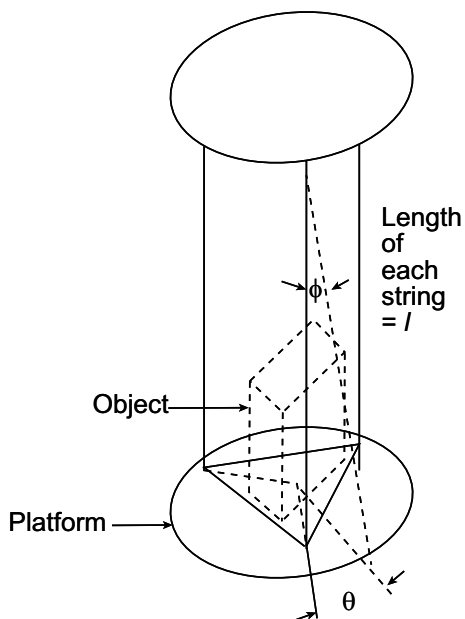
platform in m, J_t the combined moment of inertia of the object and the platform in
kg m^2, and J_p the moment of inertia of the platform in kg m^2.

When the disc is displaced through an angle θ in the horizontal plane, each string
will be displaced by an angle ϕ in the vertical plane. Equating the accelerating torque
to the restoring torque and assuming that there is no damping, the equation of motion
(for the combined platform and object) can be written as

$$J_t\ddot{\theta} + W_t r \sin\phi = 0$$

or

$$J_t\ddot{\theta} + W_t r\phi = 0 \tag{10.21}$$

for small angles of oscillation.

From the geometry of the trifilar suspension

$$\phi = \frac{r}{l}\theta \tag{10.22}$$

Substituting Eq. (10.22) into Eq. (10.21), we get

$$\ddot{\theta} + \frac{W_t}{J}\frac{r^2}{l}\theta = 0 \tag{10.23}$$

From this, the frequency of oscillation is obtained as

Table 10.4 Results of torsional vibration experiment

	Platform alone			
No.	Time for 10 oscillations (s)	Time for 1 oscillation, $T(s)$	$f = 1/T$ (Hz)	Moment of inertia, J_p kg m^2
1.				
2.				
	Platform and object			
No.	Time for 10 oscillations (s)	Time for 1 oscillation, $T(s)$	$f = 1/T$ (Hz)	Moment of inertia, J_t kg m^2
1.				
2.				

$$f_n = \frac{1}{2\pi}\sqrt{\frac{W_t r^2}{lJ}} \tag{10.24}$$

Test Procedure: The procedure is detailed as shown below:

1. Locate the centre of gravity of the platform.
2. Measure the length of the wires and the distance r from the centre of the platform as shown in Fig. 10.10.
3. Make the platform execute small amplitude vibrations in the horizontal plane.
4. Note the time taken for 10 oscillations.
5. Repeat the experiment 5 times and calculate the average time period and hence the natural frequency in torsion.
6. Compute the moment of inertia of the platform.
7. Place the given object on the platform so that the vertical axis through its centre of gravity (CG) coincides with that of the platform.
8. Repeat steps 3 to 5 with the object on the platform and obtain the combined moment of inertia.
9. Find the moment of inertia of the object by subtracting the moment of inertia of the platform from the total moment of inertia of the platform and object.
10. Verify results using analytical expressions for the regular object.
11. Find the moment of inertia of an irregular object using above test procedure.

The results may be tabulated as shown in Table 10.4.

Questions

(1) What happens if the centre of mass of the object does not coincide with the mass and geometric centre of the platform?
(2) What happens if the moments of inertia of the object and the platform are comparable?
(3) How much will an error in the weight of the platform or object cause?
(4) How much error will be caused if the cords stretch?

10.7 Measurement of Critical Speed of Shaft

> **DID YOU KNOW** that studies in rotor dynamics started in the last quarter of the nineteenth Century? It was W. J. M. Rankine, in 1869, who first performed an analysis on the whirling motions of a rotor, but his model was not adequate, in that, it did not take into account rotor imbalances. He, therefore, wrongly concluded that a rotating machine would never be able to operate above the first critical speed. In 1895, S. Dunkerley published the results of an experiment, describing supercritical speeds. Gustaf de Laval, a Swedish engineer, invented the elastically supported rotor, called de Laval rotor and ran it to supercritical speeds in 1889. In 1916, W. Kerr brought out a publication showing experimental evidence of a second critical speed. In 1919, Henry Jeffcott brought out his first paper where the theory of unbalanced rotors is described. Between the work of Jeffcott and the beginning of World War II, a lot of work was done in the area of instabilities and modelling techniques, which led to the work of N. O. Myklestad and M. A. Prohl on the transfer matrix method for analyzing rotors. Most of the present-day studies on rotor dynamics use the finite element method.

Introduction: Critical speed of a shaft is the rotational speed at which the unbalanced forces acting on it cause it to vibrate severely at its natural frequency with a large deflection. This resonant condition can cause the rotor to get damaged fast. Generally, it is a practice to pass through such rotational speeds quickly during coast up or coast down.

Theory: The critical frequency for a shaft may be obtained as the fundamental frequency of a beam performing transverse vibration as described in Sect. 2.2.7.1 and given by Eq. (2.137). The expression for constants $(\beta_n l)^2$ for different modes of vibration and for different boundary conditions may be obtained from Table 2.6.

Aim: To experimentally determine the critical speed of the given shaft for different boundary conditions and to validate the results theoretically.

Test Setup and Procedure: A whirling shaft apparatus may be fabricated in the laboratory or bought off the shelf. The schematic diagram of the test setup is shown in Fig. 10.11. The entire assembly is fixed to a solid metal frame. The horizontal test shaft is driven by a variable speed DC motor with a speed control unit through a kinematic coupling with a double universal joint. The shaft is held by one bearing at the driven end and by another bearing at the tail end of the shaft. The latter bearing slides in its housing to allow shaft movement in a longitudinal direction as it 'whirls'. Both bearings allow free angular movement of the shaft, much like a free-free condition. Two nylon bushes that can be moved along the length of the shaft help maintain its amplitude of vibration within allowable limits. The speed control unit
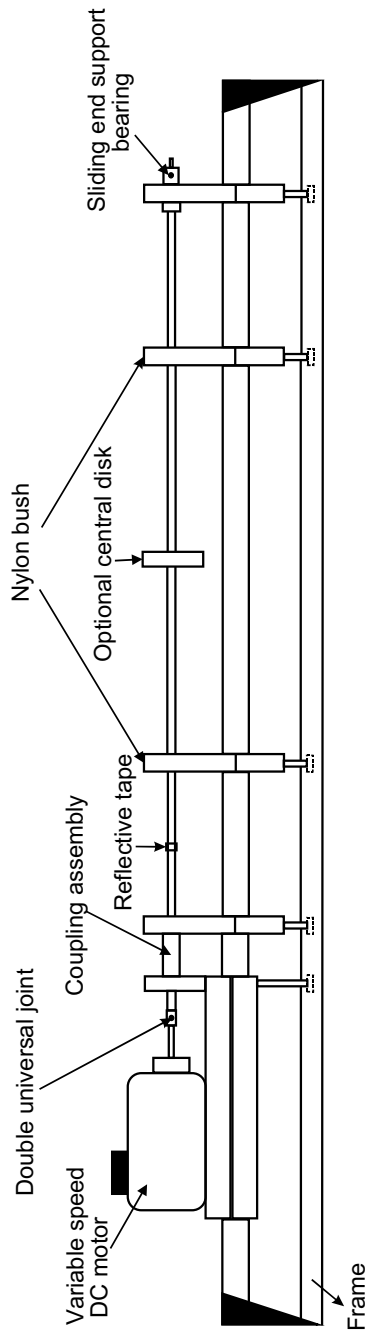
Fig. 10.11 Whirling of shaft setup

Mode 1

Mode 2
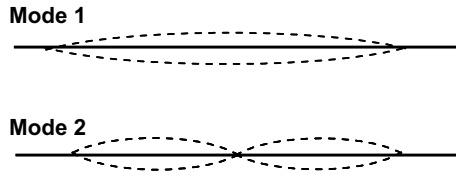
Fig. 10.12 Mode shapes of shaft

Table 10.5 Experimental results of critical frequencies and validation

No.	Shaft diameter (m)	Shaft length (m)	Simply supported f_n (Hz)		Supported fixed f_n (Hz)		Fixed-fixed f_n (Hz)	
			Th.	Exp.	Th.	Exp.	Th.	Exp.
1.								
2.								

Note: Th.-Theoretical; Exp.-Experimental

may have a display unit to show the shaft speed. Alternately, an optical tachometer triggered by a reflective tape on the shaft may be used.

The setup may be made such as to accommodate several shafts of various lengths and diameters by moving the sliding end bearing; this will enable a better understanding of how the shaft geometry affects whirling and its 'critical speed'. The tail end of the shaft may be supported using bearings that restrict its angular movement to give 'fixed ends' if necessary. A central circular disc may be fitted to the shaft to provide a load if required. It is necessary to have a removable safety guard surrounding the shaft.

Procedure: Part A: Whirling of Shaft Without Loading Fix a shaft of known diameter and length in the apparatus and ensure that the shaft has been properly fixed and simply supported at both ends. Before commencing the experiments, ensure that the safety guard has been fitted. Increase the speed slowly to the level where whirling is observed. Observe the loop appearing on the shaft with a single bow and note down the speed using the tachometer. Increase the speed slowly till the second mode is observed with two bows. Now reduce the speed slowly and bring the shaft back to rest. The mode shapes will appear as shown in Fig. 10.12. The experiment may be repeated with fixed-supported and fixed-fixed boundary conditions and also with shafts of various diameters and lengths. Readings may be tabulated as shown in Table 10.5.

Procedure: Part B: Whirling of Shaft with Disc loading The experiments may also be repeated with a central disc as shown in Fig. 10.13. The procedure is the same as before, except that the experiment is performed with the central disc attached as shown in Fig. 10.13. The whirling frequency f of the shaft-disc system is to be determined. The critical frequency for the disc alone is calculated from Eq. (10.25).
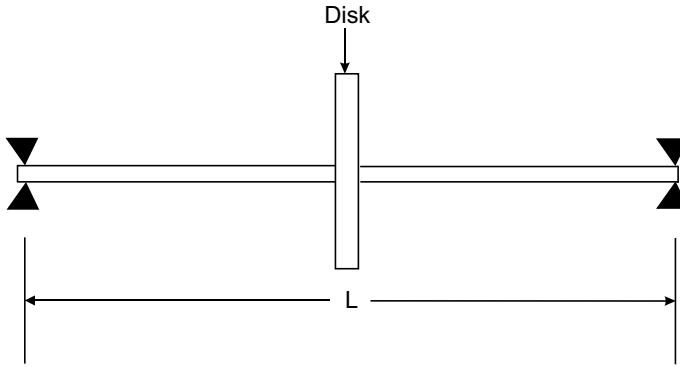
Fig. 10.13   Shaft loaded with a central disc

Table 10.6   Experimental results of critical speed of shaft with central disc

No.	Loading	Measured system critical frequency, f (Hz)	Shaft critical frequency f_s (Hz)	Disc frequency f_1 (Hz)
1.	Disc 1			
2.	Disc 2			

$$\frac{1}{f^2} = \frac{1}{f_s^2} + \frac{1}{f_1^2} \tag{10.25}$$

where f = critical speed of the whole system, f_s = critical speed of shaft alone, and f_1 = critical speed with disc alone. The results for critical frequency with the disc may be tabulated as shown in Table 10.6.

Questions

(1) What happens if the shaft is not straight?
(2) What are the factors which can cause errors in the computed critical speed?
(3) What are the difficulties in measuring the critical speed accurately?
(4) Try to prove Eq. (10.25).

10.8 Determination of Force Transmissibility

> **INTERESTING FACTS:** Machinery and processes which generate vibratory forces contribute to unwanted vibration exposure. A very effective way of reducing vibration is to interrupt the propagation path between the vibrating source and the receiver (quite often the human being or other machinery or equipment in the vicinity). In fact, open trenches with depths up to a few 10s of metres are used for cutting the path of vibration between residential buildings and nearby industries with vibrating machinery. Historically, people have used passive vibration isolation with steel in the form of springs, pads and cables, rubber, felt, cork, elastomers, foam, fibre glass boards, etc. for machinery. However, active vibration isolation systems using electric power, sensors, actuators, and control systems are being increasingly used.

Introduction: Vibration isolation is achieved by mounting the source on the test bed or receiver through rubber pads. It is required to understand how effective these isolators are in vibration isolation. Hence, it is required to measure force transmissibility.

Theory: The theory behind force transmissibility has been described in Sect. 2.2.8.1. The expression for force transmissibility is given by the expression

$$TR = \left| \frac{F_T}{F} \right| = \frac{\sqrt{1 + (2\zeta\eta)^2}}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \tag{10.26}$$

Figure 10.16 shows the dependence of transmissibility on frequency ratio.

The following are clear from the figure:

   (i) For good vibration isolation, the natural frequency of a vibration isolator system should be small, i.e., the isolator springs should be soft.
  (ii) The operating frequency of the system should be much higher than its natural frequency, at least 1.414 times.
 (iii) In this frequency range, an undamped spring is better than a damped spring.

Test Setup and Procedure: For the measurement of force transmissibility, a specimen may be custom made in the form of a three-layer sandwich structure with a rubber pad sandwiched between two aluminium plates. An electrodynamic shaker fixed horizontally is given harmonic excitation from a function generator (Fig. 10.14a) to one face of the specimen through a stinger and force transducer. The force response is sensed on the other face of the specimen using a force transducer. It is desirable to have both the force transducers with built-in charge amplifiers of the Integral Circuit Piezoelectric (ICP) type; otherwise, external charge amplifiers have to be used for
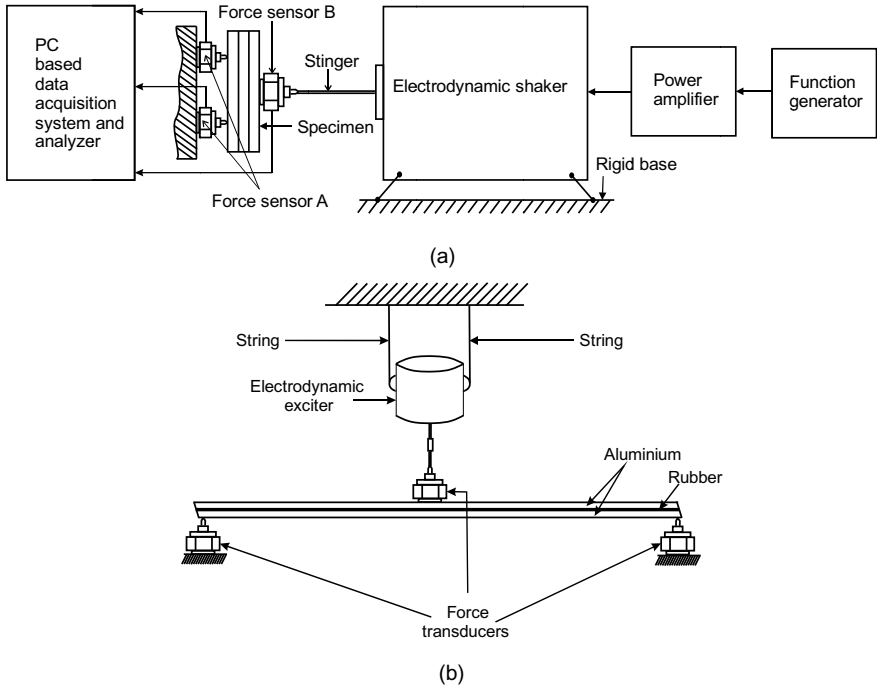
Fig. 10.14 Test setup for force transmissibility studies: **a** horizontal arrangement, **b** vertical arrangement

signal conditioning. Stepped sine or swept sine excitation with appropriate sweep rate and frequency range has to be used. If required, a power amplifier has to be used to boost the force amplitude. The input and response signals are fed to a PC-based data acquisition system or a spectrum analyzer and the results are obtained in terms of input force F_{inp} and response force F_{out} spectra. Figure 10.14a shows the layout of the experimental setup with the electrodynamic exciter fixed to impart horizontal motion. An alternate arrangement is shown in Fig. 10.14b where the exciter is constrained to vibrate vertically.

The measured transmissibility is

$$TR = \frac{F_{out}}{F_{inp}} \tag{10.27}$$

The results may be tabulated as shown in Table 10.7.

Questions

(1) What purpose does the stinger serve?

Table 10.7 Results of force transmissibility

Sl. no.	Frequency (Hz)	F_{out} (N)	F_{inp} (N)	TR
1.				
2.				
3.				

(2) Does the force transducer have to be sandwiched between the base and the specimen?

(3) Explain why the exciter has been supported in the condition shown in the setup.

(4) What is the phase difference between input and output at resonance?

10.9   Measurement of Displacement Transmissibility

> **DID YOU KNOW** the Tomb of Cyrus in Iran is the oldest structure in the world with vibration base isolation, making it resilient to earthquake excitation? The tomb, made of limestone has two foundations. The lower one was made of stones cemented together with a special mortar and polished, resulting in a smooth surface. The upper foundation was made of wide blocks of polished stones, attached by metal bars and clips, creating a large plate, which was not mortared to the base. In the event of an earthquake, the base foundation moves, and the upper part slides freely on its base. Later, vibration isolation systems were designed such that buildings would rest on resilient load-bearing members (stiffness) and energy dissipating mechanisms (dampers) fixed to the foundation, typically with steel plates. Common springs used are pneumatic, steel coils or pads, rubber (elastomeric) pads, fibre glass boards, etc. When an earthquake occurs, these isolators being more compliant horizontally, allow the building to rock gently back and forth.

Theory: The theory behind displacement transmissibility is discussed in Sect. 2.2.8.2 and is very similar to that for force transmissibility.

The expression for displacement transmissibility is

$$TR = \sqrt{\frac{1 + (2\zeta\eta)^2}{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \tag{10.28}$$

It may be shown that acceleration transmissibility has the same expression as displacement transmissibility and is given by

$$TR = \left| \frac{\ddot{x}}{\ddot{y}} \right| = \frac{\sqrt{1 + (2\zeta\eta)^2}}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} = \left| \frac{X}{Y} \right| \qquad (10.29)$$

In the test setup described for the measurement of displacement transmissibility, accelerometers have been used since they are easier to fix on the specimen than LVDTs.

Experimental Setup and Test Procedure: The experimental setup used for the measurement of displacement transmissibility may be configured as described below. It consists of an impact hammer, a piezoelectric accelerometer on the excitation side and one on the output side, both with in-built charge amplifiers (of the ICP type). The input and output acceleration signals are connected to a spectrum analyzer or a PC-based data acquisition system. The specimen in the form of a multilayer sandwich structure can be suspended by two nylon cords from a fixed support. The specimen may be made in the form of a three-element sandwich plate with a rubber layer sandwiched between two metallic plates. The impact hammer with an appropriate tip, rubber or plastic or metal, depending on the frequency range of interest, is used to give a transverse impact to one side of the plate and the accelerometer measures the input signal. The resulting vibration wave is measured on the other side of the plate by another accelerometer. Both input and output have to be measured transverse to the plate. These signals are fed to the data acquisition system and the results are obtained in terms of spectra using software installed in the analyzer or the computer. The analyzer may be made to compute the ratio of output acceleration to input acceleration. An averaged signal using a reasonable number of averages may be used to get the transmissibility. Figure 10.15 shows the layout of the experimental setup. Results may be plotted as transmissibility versus frequency in Hz as shown in Fig. 10.16a in the form of a linear plot as a function of frequency or as a logarithmic plot of frequency in Fig. 10.16b to boost the height of the smaller peaks.
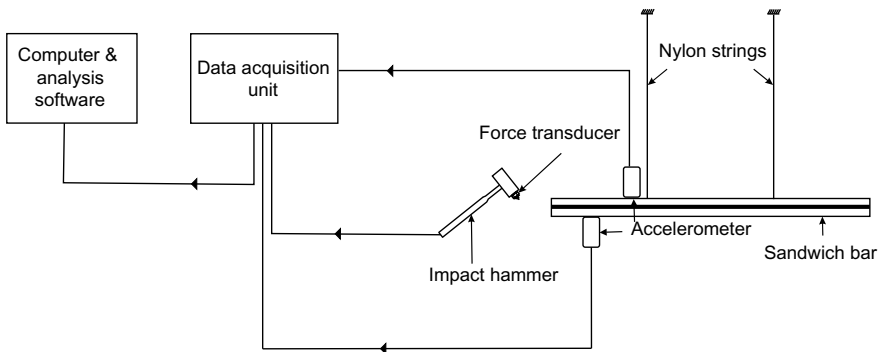
Fig. 10.15 Experimental set up for measurement of displacement transmissibility
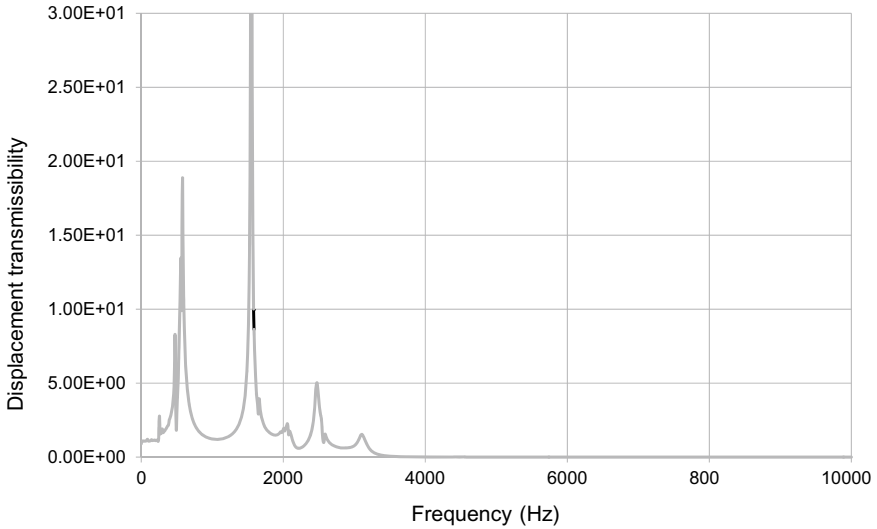
(a)
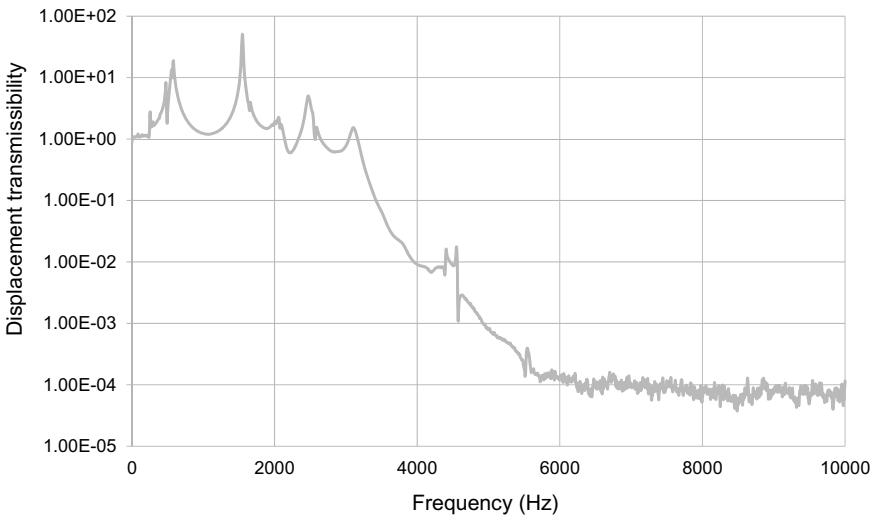
(b)

Fig. 10.16 Displacement transmissibility plots: **a** linear plot, **b** logarithmic plot

Questions

(1) How do you design isolators for a machine running at constant speed?
(2) How do you design isolators for a machine running at variable speed?
(3) When do you use a linear plot and when a logarithmic plot?
(4) Why is the orientation of the impact hammer important?

10.10   Determination of Stiffness of Isolators

INTERESTING FACTS: Metallic springs constitute one of the earliest isolators. Rubber in the form of rubber pads is the second most popular vibration isolator. Pads or sheets of flexible materials like elastomers, rubber, cork, dense foam, and laminate materials are also being extensively used for the isolation of vibration from industrial equipment such as pumps, motors, heating, ventilation, and air conditioning systems, as well as in vehicles and under common household items such as washing machines and even under audio systems. Oil, foam, and polyurethane are also used to isolate vibrations since they serve as both isolators and dampers. The most common vibration isolators are pneumatic or air isolators for large vehicles and industrial equipment and mechanical springs as heavy-duty isolators for building systems and industry. Moulded or bonded rubber and elastomeric isolators are often used as machinery mounts or in vehicles. Wire rope isolators which are very durable are often used in military applications. Workers who regularly use power tools in industries ranging from construction and maintenance to mining and forestry use anti-vibration gloves. Suspended seats are being used to isolate passengers from vehicular vibration with the intention of reducing whole body vibration (WBV).

Introduction: A vibrating source, say an engine, directly bolted to a test bed transmits forces to the foundation. Vibration isolation is achieved by mounting the engine on the test bed through rubber pads. In a test rig, often any four identical cork pads or isolators which are available on the shop floor are used without doing any evaluation of stiffness. For any vibration isolation problem, computation or measurement of the stiffness of the pads is essential. If pads of wrong stiffness are used, the effect is often adverse, the vibration transmissibility increasing instead of decreasing. Hence, it is required to understand how to measure the stiffness of isolator pads.

Aim of the Experiment: To evaluate the stiffness of isolators in the vertical and two mutually perpendicular horizontal directions and arrive at desirable stiffness of the isolators to be used for effective vibration isolation.

Test Setup and Procedure: The four given identical vibration mounts, the stiffness of which are to be evaluated are mounted on the floor at the four corners beneath a double mass reaction exciter (of known mass *m*) as shown in Fig. 10.17. The exciter is used instead of an engine. The procedure is as described below.

(i)   The exciter is constrained to vibrate in the vertical direction and the vibration velocity in this direction is measured using an electrodynamic velocity transducer. The speed of the exciter is swept from 100 to around 3000 RPM. At each speed, the vibration of the exciter is sensed using a velocity pickup and fed to a
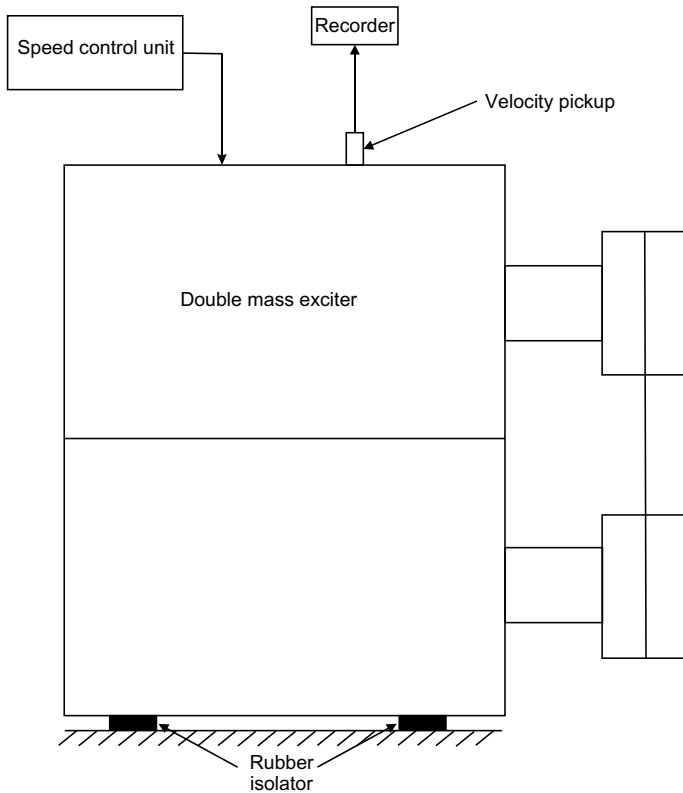
Fig. 10.17 Typical setup for transmissibility experiment

recorder to display vibration velocity. Measurements can be made for two other configurations to evaluate the stiffness in two mutually perpendicular horizontal directions.

(ii) The exciter is constrained to vibrate in the horizontal direction with the shear direction perpendicular to the direction of exciting force; the vibration velocities in the direction of exciting force and perpendicular to it are measured.

(iii) The exciter is constrained to vibrate in the horizontal direction with the shear direction along the direction of exciting force; the vibration velocities in the direction of exciting force and perpendicular to it are measured as before.

Ideally, a test like this should be carried out using an exciter which gives a constant excitation force. Since the reaction exciter gives a force output proportional to the square of the running speed (being an unbalanced mass exciter), the measured velocity values have to be normalized to simulate a constant force in the forced vibration experiment. The normalization may be done using a reference speed, say, 1000 RPM. The normalized velocity is calculated as

Fig. 10.18 Normalized
velocity versus speed
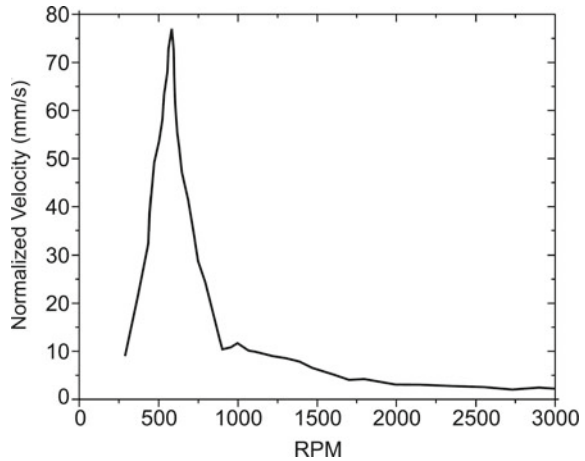
$$Velocity_{\text{normalized}} = Velocity_{\text{measured}} \times \left(\frac{\omega_{\text{ref}}}{\omega}\right)^2 \qquad (10.30)$$

Here ω and ω_{ref} are the excitation and reference frequencies.

The normalized velocities are plotted as a function of running speed (typical plot shown in Fig. 10.18) and the resonant frequency of the spring-mass system (constituted by the exciter and the four rubber isolators) is found in the three directions mentioned above.

Assume that a machine with the same mass as that of the exciter and with a predominant first-order vibration frequency is to be mounted on the four mounts and that the machine is expected to run at speeds from 800 to 3500 RPM; are the isolators suitable? If not, choose isolators which will not result in a system resonance in the operating frequency range.

Conduct the same test in all three directions with the newly designed isolators and repeat the test in the frequency range 100–3500 RPM in steps of 100 RPM as before. Plot the normalized velocities as before and prove that the isolators are suitable.

Questions

(1) Why is it required to have an exciter which produces a constant force in the frequency range of interest?
(2) Suppose the machine has a predominant 2× rpm component, then how would the design of the isolators change?
(3) If you were to use an accelerometer instead of a velocity transducer, would there be any change in the procedure?
(4) If you were to use a displacement pickup instead of a velocity transducer, would there be any change in the procedure?
(5) If an electrodynamic exciter is used instead of a double mass reaction exciter, how would the test setup change?

10.11 Determination of Natural Frequencies and Damping Ratios of a Torsional Vibration Damper

> **INTERESTING FACTS:** Besides translational motion, we often come across angular oscillations, e.g. a section of a shaft moving clockwise or anticlockwise in relation to the other cross sections. Such a problem is said to be one of the torsional vibration. A torsional vibration damper pulley is used in vehicles for reducing potentially damaging torsional vibrations in the crankshafts of internal combustion engines. This type of damper reduces the vibration when a torsional excitation frequency equals the first natural frequency of the crankshaft, but not at other speeds. If such vibration is not controlled, it can lead to failure of the crankshaft or other accessories.

Aim of the Experiment: The aim of the experiment is to determine the first two natural frequencies and damping ratios of a torsional vibration damper pulley which may be bought in an automobile scrap shop. If a torsional vibration damper is not available in the laboratory, then a solid circular disc may be used as the test specimen.

Test Setup and Procedure: The procedure involves carrying out a forced vibration test, not in translation, but in torsion. Two lugs are welded to the outer periphery of the damper (or disc), diametrically opposite to each other. The damper is fixed to a rigid support by holding the shaft keyed to the inner hub as shown in Fig. 10.19.

It is excited in the torsional mode through a stinger by an electrodynamic exciter, the excitation point being one of the lugs. Such an arrangement may be used in the absence of a torsional exciter. Stepped sinusoidal excitation is imparted to the electrodynamic shaker using a function generator. The acceleration signals at both lugs are simultaneously picked up by piezoelectric accelerometers with built-in charge amplifiers. The outputs of these accelerometers are fed to a storage oscilloscope or recorder or data acquisition system. The natural frequency of torsional oscillation f_n is found out corresponding to the maximum measured output voltage. The signal recorded by the storage oscilloscope is used to determine if the vibration response corresponds to a bending mode or torsional mode. The phase shift between the acceleration signals at the two lugs is used for this. This phase shift is 0° in the case of bending vibration and 180° in the case of torsional vibration. The damping ratio is determined by the half-power method using an RMS voltmeter or by observing the amplitude of the accelerometer outputs on the oscilloscope. Having got the first natural frequency and damping ratio from the half-power frequencies f_1 and f_2, the same process is repeated to get the second natural frequency and damping ratio.
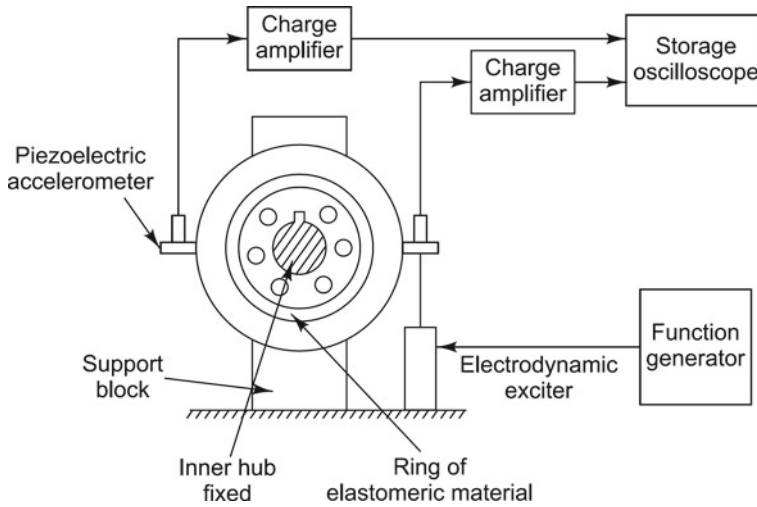
Fig. 10.19 Setup for testing torsional damper

Table 10.8 Torsional natural frequency and damping ratio results

Mode	Natural frequency, f_n (Hz)	Lower half-power freq., f_1 (Hz)	Upper half-power freq., f_2 (Hz)	Damping ratio, ζ
Mode 1				
Mode 2				

The natural frequencies and damping ratios of the damper/disc may be tabulated as shown in Table 10.8. The values reported are the average values taken from three sets of measurements.

Questions

(1) What is the difficulty in measuring damping ratio if it is very small?
(2) Without looking at the waveforms, would you be able to distinguish between torsional and translational modes of vibration?
(3) Comment on the vibration force required to excite different modes of vibration.
(4) How would you verify the torsional natural frequency analytically?

10.12   **Calibration of Vibration Transducer**

> **DID YOU KNOW** that one of the most common vibration frequencies of calibrators is 159.2 Hz, which is equivalent to 1000 rad/s? The reason for this is that at 1000 rad/s, the numerical values of the amplitudes for all three vibration quantities are the same, i.e., 10 μm of displacement, 10 mm/s of velocity, and 10 m/s^2 of acceleration. Vibration calibrators are often used for calibrating vibration transducers at the site and are, hence, usually portable and battery operated. Some calibrators allow calibration at several frequencies or allow fine-tuning to a specific frequency range. The most common frequencies of calibrators in the market are 160 Hz and 10 kHz. A few commercially available vibration calibrators also contain signal conditioning amplifiers for use with various types of transducers and may also display the sensitivity of the pickup.

Introduction: The vibration engineer is often faced with the task of calibrating a transducer for which calibration data have been lost, or for which it is a long time since calibration has been done. A simple calibration procedure for an accelerometer involves placing it on a horizontal surface with the sensing element vertical and then turning it by 90° to make it horizontal. The resulting voltage change corresponds to 1 g and the sensitivity in mV/g may be obtained. For this, the accelerometer should be capable of recording a static (DC) change in acceleration or in other words it should be connected in DC coupling mode. If a hand-held calibrator is available, the accelerometer may be mounted on it. A typical calibrator produces 1 g at 160 Hz. The voltage output corresponding to this change in acceleration by 1 g may be recorded to get the sensitivity

Aim of the Experiment: The aim of the experiment is to calibrate a transducer, the sensitivity of which is not known, using a reference pickup with known sensitivity.

Test Setup and Procedure: For a general calibration procedure, a shaker table with a known output displacement over a range of frequencies may be used. Alternately, the pickup to be calibrated and a reference pickup, the sensitivity of which is known and is constant over the required frequency range, may be fixed on the same shaker. The sensitivity of the pickup to be calibrated is determined, knowing that of the reference transducer and the vibration amplitude of the table. Figure 10.20 shows the test setup.

   The accelerometer to be calibrated and the reference accelerometer (with available calibration sheet) are both mounted on a mechanical shaker of the cam follower type. This type of shaker has the advantage that a constant displacement may be obtained over a range of frequency values; however, such a shaker may not be able to operate above 6000 RPM (100 Hz). For calibration in higher frequency ranges, an electrodynamic exciter can be used. This, however, has the drawback that a con-
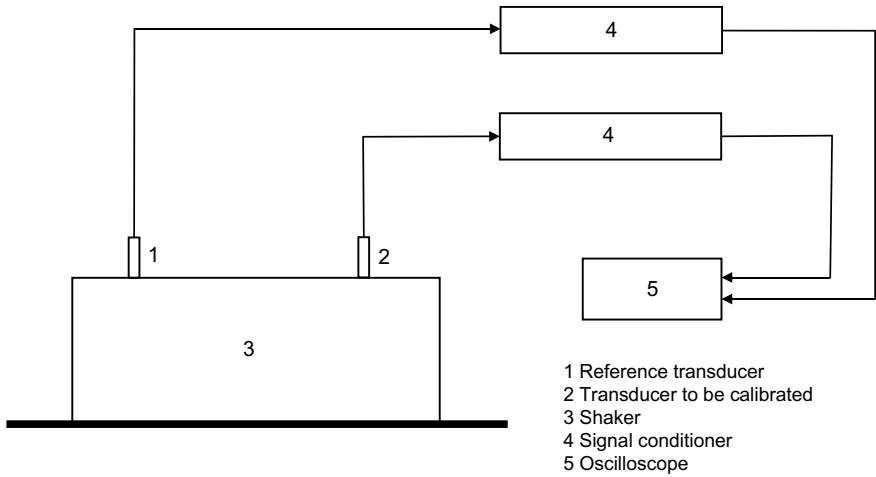
Fig. 10.20 Test setup for calibration of transducer

1 Reference transducer
2 Transducer to be calibrated
3 Shaker
4 Signal conditioner
5 Oscilloscope

Fig. 10.21 Typical calibration curve of accelerometer
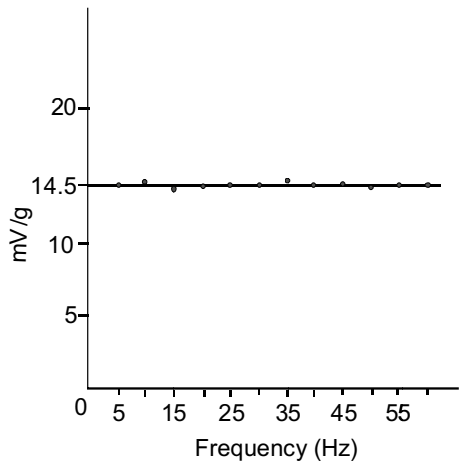
stant displacement amplitude cannot be obtained. The outputs of both transducers are provided with appropriate signal conditioning and the amplified voltages are fed to an oscilloscope or analyzer for known shaker acceleration. The table is driven over the required frequency range and the outputs of both transducers are read on an analyzer at fine frequency intervals. The sensitivity of the test accelerometer is calculated in terms of mV/g and is plotted as a function of frequency. Figure 10.21 shows a sample plot.

Questions

(1) Discuss/compare the results found from the three different calibration procedures used on the same accelerometer.

(2) Discuss the limitation(s) of each calibration method.
(3) How would you calibrate with a shaker which does not produce a constant vibration amplitude over a range of frequencies?
(4) What do you mean by frequency range of a transducer?

10.13   Measurement of Complex Modulus of Elasticity

INTERESTING FACTS: We are all familiar with Young's modulus of elasticity or elastic modulus named after Thomas Young (1773–1829), an English physician and physicist. This quantity measures an object's resistance to being deformed elastically (non-permanently) when a tensile stress is applied to it. For viscoelastic materials, stress–strain plots form 'hysteresis' loops, the area within each loop representing the energy lost. The viscoelastic response of such materials is typically analyzed in terms of their complex modulus of elasticity. The most common viscoelastic or damping materials are plastics, composites, wood, asphalt, rubber, cork, coir, viscoelastic paints, skin tissue, polymer foams, etc. The discs in the human spine are viscoelastic and they undergo creep, i.e., they get shorter over time under normal body weight. Lying down allows the discs to recover and that is why most people are taller in the morning than at night. Astronauts are known to gain up to 5 cm in height under near-zero gravity conditions. A newborn baby's head is viscoelastic and if a baby lies in one specific position for a long time, its head becomes misshaped due to creep deformation.

Introduction: Young's modulus is basically the slope of the tensile (unidirectional) stress–tensile strain or longitudinal load/longitudinal deformation line within the elastic range for ideal, elastic bodies. For such bodies, stress and strain are (i) in phase and (ii) invariant with respect to time. For viscoelastic materials on the other hand, stress–strain plots form 'hysteresis' loops, the area within each loop representing the energy lost. Viscoelastic materials have two different properties combined, viscosity and elasticity. The term 'viscous' implies that when they are exposed to an external force, they deform slowly with a permanent rearrangement of the fluid molecules. When a viscoelastic material is exposed to a dynamic stress, the resulting strain lags the applied stress and this delay can be described by a phase angle δ. This is the reason why viscoelastic stress–strain plots form 'hysteresis' loops. This energy absorption by viscoelastic materials is the reason why they are good shock absorbers. Besides, their mechanical properties depend on the rate at which they are deformed, i.e., the stiffness of the material increases with the loading rate. Hence, a whole family of stress/strain curves is required to represent the deformation behaviour at different

deformation rates. While designing/redesigning a structure for reduction of vibration and noise, in straightforward cases involving only a few natural frequencies, a suitable modification of the stiffness to mass ratio can move the natural frequencies outside the range of forcing frequencies. However, if the range of these frequencies is large, as in variable speed machines, then 'tuning out' of resonances becomes very difficult and the only solution is to reduce the vibration amplitude using these viscoelastic damping materials or 'deadeners'. Many rubber materials are nowadays being used as deadeners for automotive applications. The chemical compositions of these rubbers are tweaked, so as to result in large values of loss factors.

Theory: Under static conditions, modulus of elasticity is defined as

$$E = \frac{\sigma}{\varepsilon} \tag{10.31a}$$

Under dynamic conditions the phase shift between the stress and strain gives rise to a complex modulus of elasticity given by

$$E^* = E(1 + i\xi) \tag{10.31b}$$

where ξ is the loss factor and is given by

$$\xi = \tan \delta \tag{10.31c}$$

Here δ is the phase shift caused by loss of energy.

Aim of the Experiment: The aim of this exercise is to determine the loss factors of rubber materials.

Test Setup and Procedure: Typically, evaluation of loss factors is done very conveniently using what is called a complex modulus apparatus, which is nothing but a setup that facilitates forced vibration tests to be conducted on a bar-like specimen for the measurement of dynamic (complex) modulus of elasticity and the measurement of internal damping/loss factors. Alternately, the experiment can be done using any forced vibration experiment setup.

The complex modulus apparatus or Oberst apparatus (Type 3930, Brüel & Kjær, Denmark) can be used for the purpose. Figure 10.22 shows a close-up of the Complex Modulus Apparatus along with the associated instrumentation.

This special apparatus allows the test specimen to be fixed either in a clamped-free or clamped-clamped condition and is so designed as to enable firm clamping of the samples with very low parasitic losses, and precise mounting of the transducers with respect to the sample. It has a guide pillar with adjustable supports for two identical non-contact electromagnetic transducers (Type MM0002, Brüel & Kjær, Denmark). One transducer, typically the one near the free end (if the test is being conducted in the cantilever mode), is used as the non-contact vibration exciter, since
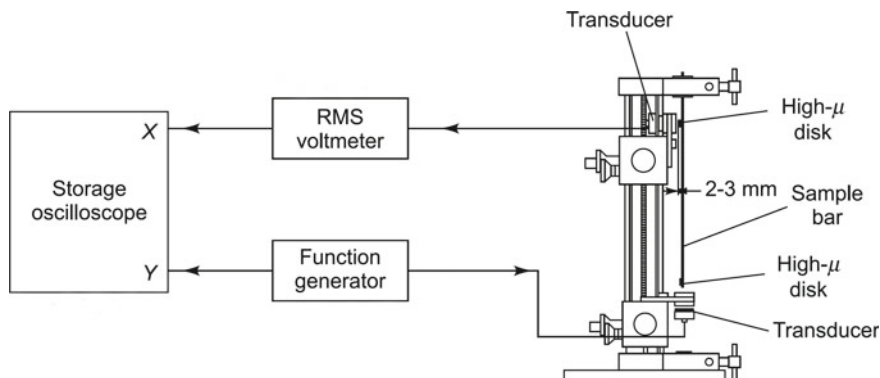
Fig. 10.22 Setup for measurement of complex modulus

larger vibration amplitudes can be imparted here than near the fixed end. The other transducer is used as the vibration sensitive pickup.

Rubber specimens, the loss factors of which are to be evaluated, can be made in the form of sheets of thickness around 2 mm. They have to be cut and bonded onto mild steel bars (having negligible damping) of length 75 to 220 mm. The beam is then set into forced harmonic vibration; for this, the exciter is fed with a variable frequency sinusoidal excitation signal from a function/arbitrary waveform generator.

When the forcing frequency is changed, the vibration response changes; it shoots up at the resonant frequencies, the amplitude at resonance being limited only by the mechanical damping present in the bar. Natural frequencies of the first few modes and the corresponding half-power frequencies can be captured using a digital storage oscilloscope or recorder or data acquisition system. It is good to see if the response of the strip is sinusoidal to ensure that the readings are taken in the linear vibratory range of the specimen; this will not be possible if the voltage value is simply read off using a voltmeter. Subsequently, an RMS voltmeter may be used to read off the response voltage directly in dB to find out the half-power frequencies (frequencies where the voltage is 3 dB below that at resonance). From the measured values, the structural loss factor can be found by the half-power method using the formula:

$$\xi = \frac{f_2 - f_1}{f_n} \tag{10.32}$$

where f_n is the natural frequency and f_1 and f_2 are the lower and upper half-power frequencies, respectively. The loss factors measured in the first few modes of vibration may be averaged, if required, for use in analytical models of damping. Results may be tabulated as shown in Table 10.9.

Table 10.9 Loss factor measurements

Mode	Natural frequency, f_n (Hz)	Lower half-power freq., f_1 (Hz)	Upper half-power freq., f_2 (Hz)	Loss factor, ξ
Mode 1				
Mode 2				

Questions

(1) How is loss factor ξ related to damping ratio ζ?
(2) Roughly what is the ratio of damping ratio of rubber to that of mild steel?
(3) How can you conduct the experiment using random excitation?
(4) What will happen if a contact type of transducer is used?

10.14   Study of Dynamic Vibration Absorber

INTERESTING FACTS: Many of the world's tall skyscrapers have a secret device, a tuned mass damper (TMD) protecting the building and the people inside from strong motion due to wind and earthquakes. The TMD is essentially a vibration absorber. The world's first TMDs were added to the 60-storey John Hancock Tower in Boston in 1975 to reduce the response to wind gust loading. But the most impressive of all TMDs was the one built in 2004 in the Taipei 101 Tower in Taiwan. It rises to 509 m and consists of exactly 101 floors and is just 201 m from a major fault line. The engineers had to design a structure that could withstand winds up to 216 km/h and the strongest earthquakes. Typically skyscrapers must be flexible in strong gales, yet remain rigid enough to prevent large sideways movement. The damper consists of a steel sphere 5.49 m across and weighing 660 tonnes, suspended above Floor 87 by cables anchored at Floor 92. Eight steel cables form a sling to support the ball, while eight viscous dampers linking the ball and Floor 87 act like shock absorbers when the sphere shifts. Acting like a giant pendulum, the massive steel ball sways to counteract the building's movement caused by strong gusts of wind. The ball can move 1.52 m in any direction and reduce sways by 40%. The engineers were so proud of their creation that they made the damper visible to the public from an indoor observatory located inside the tower, as against TMDs in most other buildings which are hidden from viewers.

Introduction: If a lightly damped structure is excited at its natural frequency, it will vibrate with a large amplitude leading to resonance. Such large displacements will

generate stresses large enough to create damage and over a long period of time cause ultimate failure due to fatigue. Hence, while designing a structure, it is required to know both the natural frequencies of the structure and the frequencies at which it will get excited, so that they can be kept far apart. Quite often, one does not have a say on the excitation frequency; however the natural frequency of the structure, depending on its mass and stiffness, can be altered to avoid resonance. If one has to operate close to resonance, then the damping in the structure should be increased through the use of deadeners. A third way of controlling vibrations is by attaching a vibration absorber tuned to the natural frequency of the system, so that it will extract energy at the resonant frequency and reduce the amplitude of vibration of the main structure.

Aim: The aim of this experiment is to investigate the effectiveness of tuned dynamic vibration absorbers by comparing the vibration levels of a structure with and without an absorber.

Theory: The theory behind the vibration absorber has already been described in Sect. 2.3.3.

Test Setup and Procedure: The test setup consists of a specially fabricated small 3D single storey building-like structure. The building frame can be made of moment-resisting flexible steel rods of diameter around 3 to 5 mm and the height of the storey may be around 450 mm. The base and first floor may be in the form of square plates of dimensions 300 mm × 300 mm made of rigid steel or aluminium alloy (though shown as transparent in Fig. 10.23 for visualizing the setup) with a mass of around 3 kg so as to result in a low natural frequency less 10 Hz.

The bottom of each rod is fixed to the base floor by steel bolts. The base floor itself is a concentrated mass in the form of a plate connected to a moving platform through the rods which are rigidly fixed to the moving platform. The structure has one degree of freedom, being the lateral motion of the first floor. The complete moving platform is mounted through anti-friction ball bearings on rails fixed to a base frame to ensure that the first floor mass has only one DOF. The electrodynamic shaker is mounted horizontally; another slender rod is attached to the free end of the stinger of the shaker through a coupling and excites the moving base. The shaker is provided with a sinusoidal frequency sweep from 0 to 60 Hz using a power amplifier. The structure has a vibration absorber placed on the first floor. This absorber is designed to reduce the first mode of vibration of the system. Figure 10.24 shows block diagrams of three types of passive vibration control devices which may serve as the absorber: (i) tuned mass damper (TMD), (ii) tuned liquid damper (TLD) or (iii) tuned liquid column damper (TLCD). A light material such as fibre glass may be used to fabricate the TLD and TCLD, while the TMD may be made of steel as shown in Fig. 10.24.

Irrespective of which damper is used, its mass may be kept as 3–5% of that of the main structure. The TMD is in the form of a mass which can be changed and can also be locked in any position along a flexible steel rod and fixed to the first floor. In the case of the TLD and TCLD, the containers are placed on the first floor and the water inside the container is open to air and can move freely when the structure is
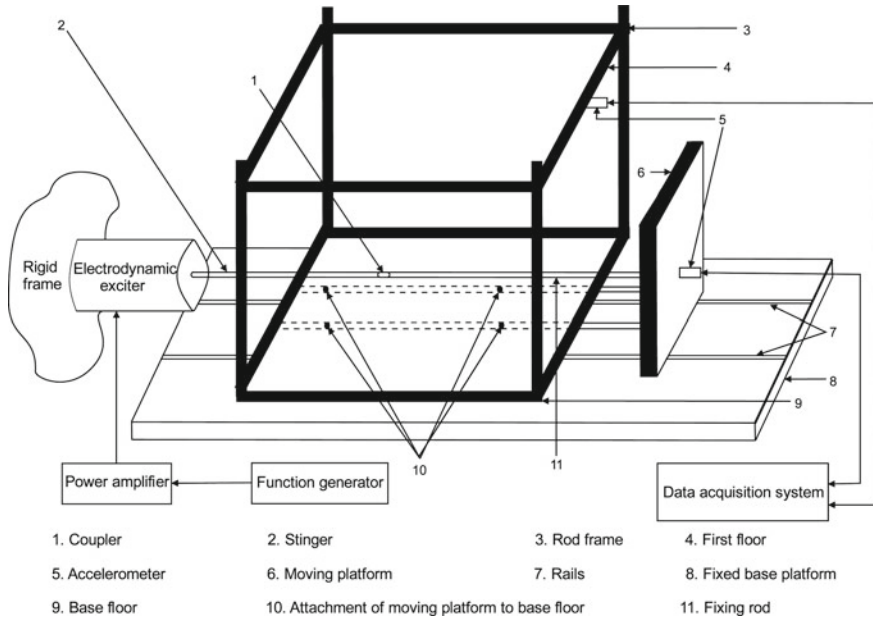
Fig. 10.23 Test setup for evaluation of vibration absorber

excited. The acceleration responses on the ground and first floors are measured by identical low frequency uniaxial accelerometers fixed horizontally along the direction of motion on each each floor; their outputs may be passed through a low pass filter before being connected to an oscilloscope or a PC-based data acquisition system or recorder or signal analyzer. Experimental results with and without damper may be compared by viewing them on the oscilloscope and plotting them as shown in Fig. 10.25 and tabulating them as shown in Table 10.10. The experiments may be repeated by changing the mass and type of the absorber.

Questions

(1) What is the effect of the absorber mass on the vibration of the first floor?
(2) How do you tune the absorber in the case of TMD, TLD, and TCLD?
(3) Instead of conducting a forced vibration experiment, can you determine the effectiveness of the dampers through a free vibration experiment?
(4) Can you redo the experiment with a pendulum type of absorber?
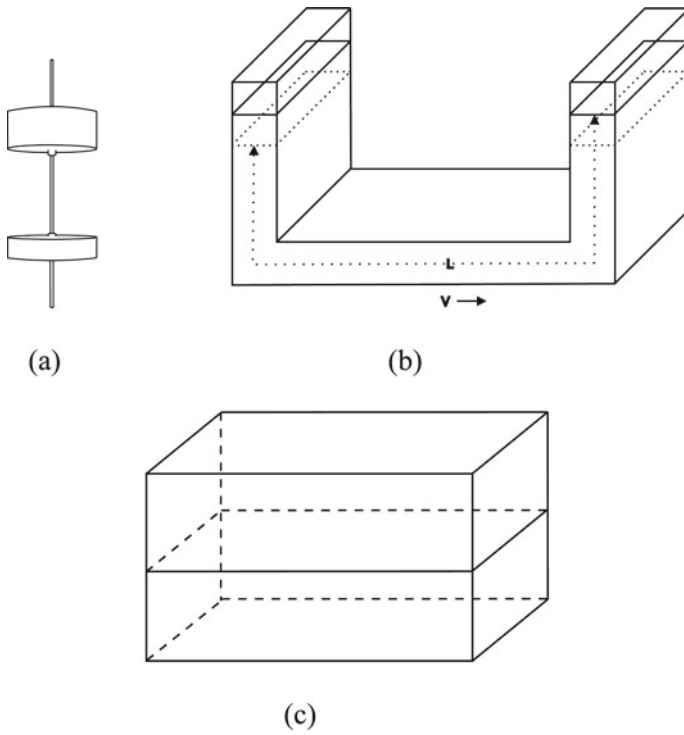
Fig. 10.24 Vibration absorbers: **a** TMD, **b** TLD and **c** TLCD
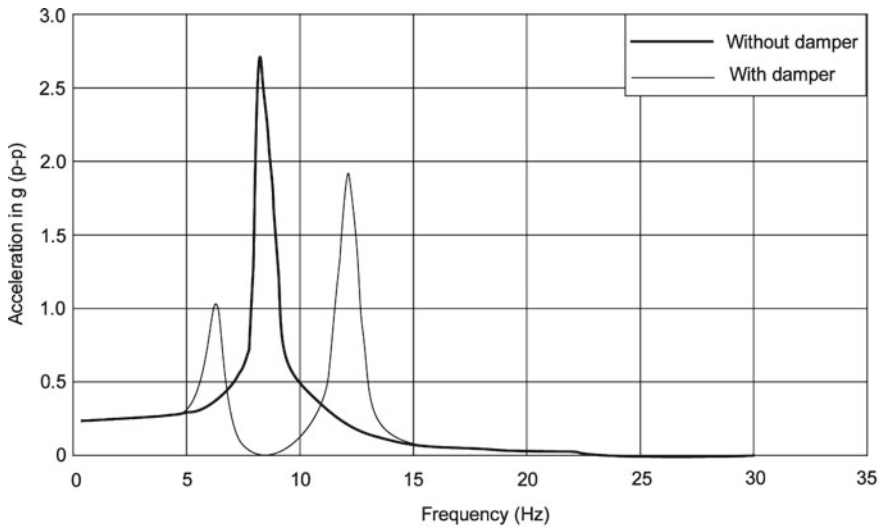
Fig. 10.25 Vibration levels with and without damper

Table 10.10 Comparison of vibration levels with and without damper

No.	Natural frequency without damper (Hz)	Acceleration of base floor, g (p-p)	Acceleration of 1st floor, g (p-p) without damper	Damper mass (kg)	Acceleration of 1st floor, g (p-p) with damper	Reduction in vibration of 1st floor, %
1.						
2.						
3.						

10.15   Modal Analysis of Plates: Sand Patterns of Mode Shapes

> **DID YOU KNOW** that on July 8, 1680, Robert Hooke conducted one of the earliest experiments in modal analysis? He sprinkled flour on a glass plate and caused it to vibrate by running a violin bow along the edge, noting the mode shapes that formed as the particles rearranged themselves. Later, Ernst Chladni, in 1787, published in his book 'Entdeckungen über die Theorie des Klanges' ('Discoveries in the Theory of Sound') details of his experiments in which he drew a bow over a metal plate lightly covered with sand. When the plate reached resonance, the sand grains moved and concentrated along the nodal lines, visually displaying the vibration/acoustic mode shapes. Chladni then proceeded to the next level where he systematically investigated the sound patterns of square, rectangular, and circular plates. Such patterns are now called Chladni figures. Blessed with a musician's ear, he could discern slight changes in frequency and he methodically recorded the different distinct patterns associated with different frequencies. Because the Chladni patterns showed exactly the modal patterns in the back plates of musical instruments, his technique became a useful tool for the makers of musical instruments and is widely used even today.

Introduction: Experimental modal analysis involves finding out the natural frequencies, mode shapes, and damping ratios of a structure and is expensive in terms of hardware, software, and time taken for analysis. Prior to obtaining the mode shapes through experimental modal analysis, a simple test may be conducted to visualize the mode shapes in the case of plate or bar-like specimens or specimens with a uniform surface, and this has been described in this experiment.

Aim of the Experiment: The aim is to visualize modal patterns of a plate by generating Chladni figures.
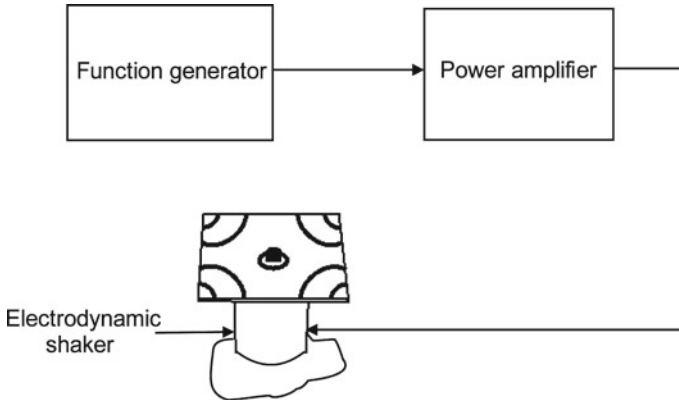
Fig. 10.26 Test setup for generating Chladni patterns

Test Setup and Procedure: An aluminium or mild steel square plate of dimensions, say, 400 mm × 400 mm × 2 mm may be chosen. Alternately, a circular plate of similar size may be chosen. The plate may be subjected to sinusoidal point excitation using an electrodynamic exciter fixed vertically on a rigid base. Figure 10.26 shows the experimental setup.

The point of excitation is generally chosen such that it does not coincide with a nodal point of the first few modes of vibration and may be taken as the geometric centre of the plate for ease of fixity. The exciter is given harmonic excitation through a function generator and power amplifier. Alternately, the plate may be fixed to a mechanical exciter of the cam follower type with a speed control unit. Before conducting the experiment, the location of nodal lines may be obtained from the literature in the case of square or circular plates. For more complicated geometries, it can be obtained only by trial and error or through finite element analysis. To visualize the mode shapes, fine sand or salt is sprinkled over the surface of the plate. The plate is excited sequentially at each of the first few natural frequencies and the patterns formed on the surface are recorded as shown in Fig. 10.27.

Questions

(1) How will the boundary conditions of the plate affect the mode shapes?
(2) Why do you hear some modes?
(3) Why do you get repeated modes for circular and square specimens?
(4) Can you relate the transverse vibration of a beam to that of a rectangular plate?
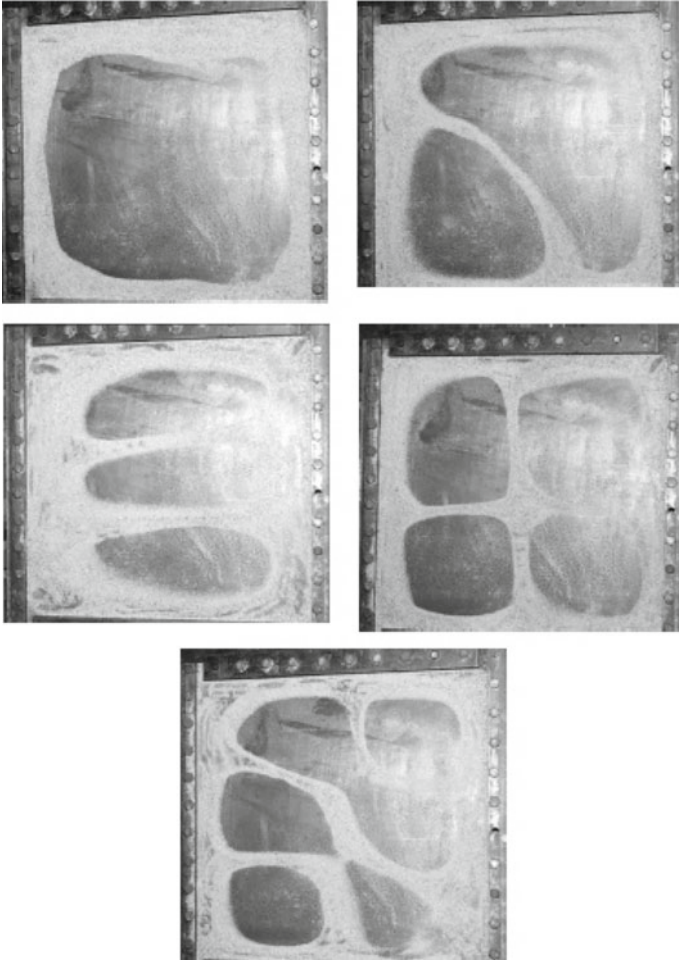
Fig. 10.27   First few mode shapes of a square plate

10.16   Modal Analysis: Roving Pickup and Fixed Shaker

> **DID YOU KNOW** that it was only in the 1960s that the era of experimental modal analysis began? This was due to the fact that many technologies that were developed earlier in the 1900s came of age and got integrated by the 1960s. The theoretical background for modern test methods was developed in the 1930s, 1940s, and 1950s and was well established in literature by the 1960s. Besides, experimental modal analysis could not take off until equipment was readily available for measuring both magnitude and phase or the real and imaginary parts of harmonic signals. With the development of the commercially available dynamic analyzers or signal analyzers, tracking filters and mechanical impedance or transfer function analyzers, commercial equipment was finally available for experimental modal analysis. Also, by the 1960s, sufficiently stable and accurate sensors were available off the shelf to measure force and acceleration, thus becoming popular in experimental modal analysis. Today, we have annual international conferences dedicated exclusively to modal analysis!

Introduction: An experimental modal analysis procedure is often conducted in order to validate the results of a finite element simulation. In practice, people choose either to rove the hammer, keeping the response transducer fixed or rove the accelerometer, keeping the shaker fixed. There is not much difference between these two choices. The idea is to obtain either the row or the column in the FRF matrix from the first or second choice.

Theory: The theory behind this has already been explained in Sect. 9.4.

Aim: To do experimental modal analysis on a bar or rod-like specimen and to determine its natural frequencies and mode shapes for the first four modes of vibration.

Test Setup and Procedure: Clamped-clamped or clamped-free boundary condition may be used for a typical laboratory test. For realizing clamped-clamped end conditions, the bar is tightened well between the clamps and the effectiveness of the end fixity is ensured by observing that the clamped ends do not show any motion in the animated mode shapes obtained from experimental modal analysis. In order to ensure minimal disturbance from the surrounding environment, the ceiling fan is kept off and it should be ensured that no other disturbance passes to the test structure through the floor.

The measurement of an FRF involves the measurement of the input force and output response as a function of frequency with all the phase information intact. Typically, in cases where the specimen does not get excited using a hammer due to the flexibility and large damping associated with the former, an electrodynamic shaker is used to excite the specimen. FRF measurements are made at a sufficiently

large number of DOFs (based on the discretization) to get accurate mode shapes. With shaker excitation, the excitation DOF is fixed, serving as the reference DOF. We then have roving response measurements with the response transducer being moved around the structure. The transducer could be a piezoelectric accelerometer fixed with a magnetic base. This will cause a varying mass loading of the structure and corrective action for this must be taken while extracting modal parameters when using contact type of transducers. But if a non-contact laser vibration transducer is used, the mass loading at each DOF is the same in all the FRFs, providing better consistency of results. Therefore, a laser vibration transducer can be used with the laser beam focussed normally on the measurement location, the transducer itself kept fixed by a stand.

Measurement of the input force is one of the most crucial aspects of a successful modal test; a force transducer (which is part of an impedance head) can be used for measuring the forces to get the FRFs. The shaker should transmit forces in line with the main axis of the force transducer. To avoid rotations about the other two axes, the exciter is connected to the force transducer through a stinger. A small pre-tension is imparted to the stinger to ensure proper contact with the test specimen.

The test beam is to be discretized into a number of elements for modal testing. Care should be taken in locating and orienting the transducer; it should not be mounted at or close to a node of any structural mode of interest. For point FRF measurements, the force and the response are required to be measured at the same point, while they are to be measured at different points while measuring transfer FRFs. It is important to note that the direction of the force and response are the same. Once the excitation and transduction systems that best suit the application have been decided on, the next step in the measurement process involves selecting an excitation function such as stepped sine, sine sweep, random, transient, etc. Though stepped sine testing gives the best signal-to-noise ratio, it is extremely slow for broadband measurements. Swept sine excitation takes less time for testing; as in the case of stepped sine excitation, it is imperative in this case also, to ensure that the frequency sweep is done sufficiently slowly to enable the system to settle down to steady-state conditions before measurements are made. If the sweep rate is too high, then the FRF gets distorted with the true resonance peak getting shifted to the right (higher frequency) with increasing sweep rate. True random signal is the most popular signal for shaker excitation since measurement time is minimal. This signal is obtained from a signal generator and is the most suitable signal to estimate the best linear model of a non-linear system, and hence it may be used as the excitation signal. Broadband white noise with equal power per unit bandwidth with a flat power spectral density across the frequency range of interest may be used. In practical measurements involving white noise, ideally, an infinite number of samples are required for achieving a flat PSD. Hence, the power spectra should be averaged over a reasonably long duration to get good FRFs.

Figure 10.28 shows a schematic drawing of the typical modal analysis test setup used with a fixed shaker and roving transducer. Modal parameters may subsequently be extracted from the FRFs by using the SDOF peak pick method which is the simplest method. This method has been described in Sect. 9.6.2.1. It has to be ensured that driving-point FRFs (with the excitation and response measurements made at
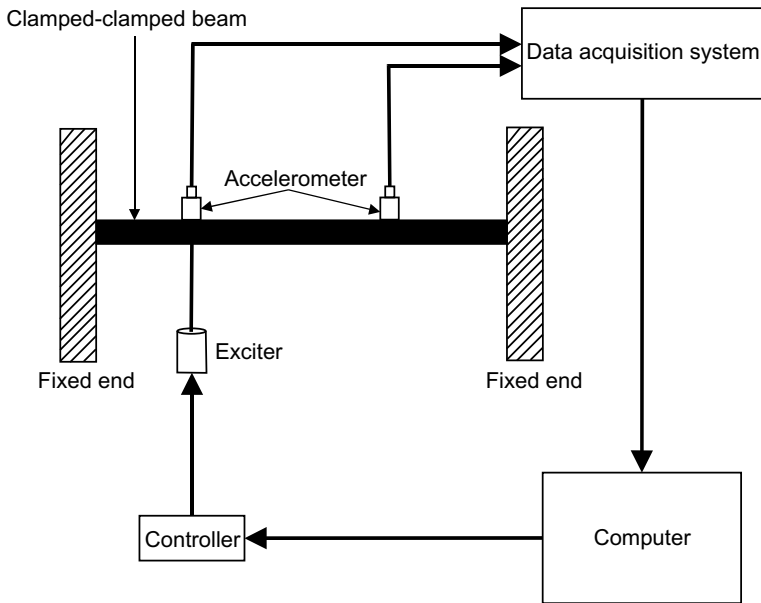
Fig. 10.28   Modal analysis test setup
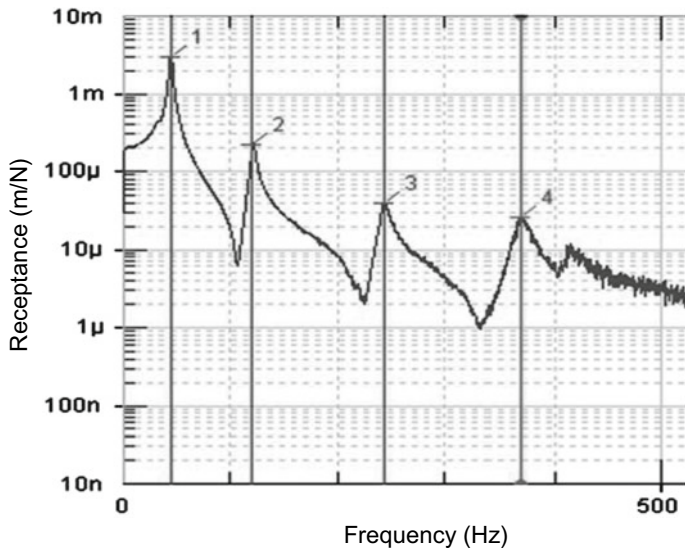
Fig. 10.29   Typical measured FRF
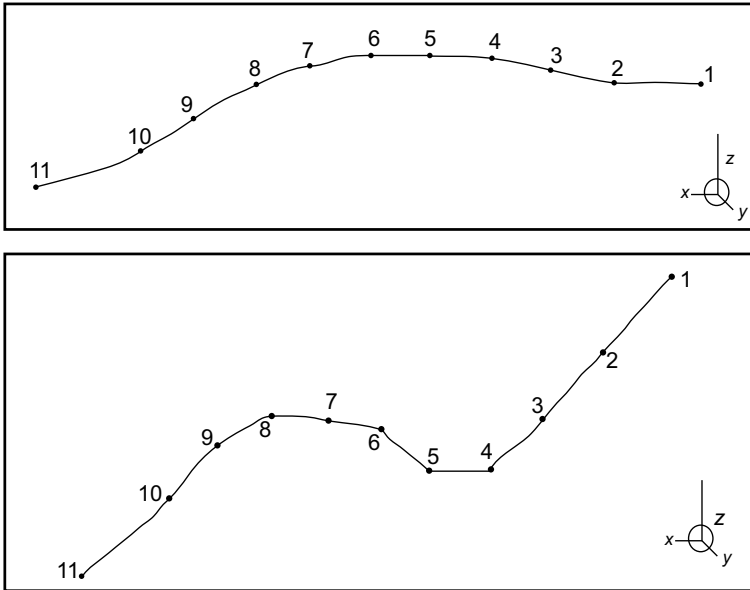
Fig. 10.30 First two measured mode shapes

the same DOF) are included, in order to calibrate (scale) the modal model. The assumption behind this method is that the structure: (i) has well-separated modes, (ii) has modes which are not too lightly damped to obtain accurate measurements around resonance, and (iii) the modes are not so heavily damped that the response at a resonance has contributions from more than one mode. A typical FRF is shown in Fig. 10.29. Figure 10.30 shows the first two measured mode shapes of the beam.

Questions

(1) Why is it difficult to obtain a perfectly clamped condition?
(2) How will the mode shapes be different for a simply supported boundary condition?
(3) Why are the natural frequencies from experiments slightly different from those obtained from the analysis?
(4) How do you distinguish between true and spurious modes in an FRF?
(5) What precautions should you take if the mode is lightly damped?

10.17 Modal Analysis: Roving Impact and Fixed Transducer

DID YOU KNOW that impact testing methods have now been utilized for over forty-five years? Impulse hammer testing is very convenient for modal analysis since the instrumented hammer, accelerometers, cables, and accessories can all be carried to the test site in a small suitcase. The impact hammer is ideal for small light weight structures; however, testing of large civil engineering structures with a hammer results in a poor signal-to-noise ratio. These hammers are available in different sizes: miniature hammers of length around 10 cm and effective head mass of 2 gm for ultra lightweight specimens, medium hammers of length 22 cm and mass 150 gm for general purpose applications, and instrumented sledge hammers of length 90 cm and mass 6 kg for large structures such as heavy foundations, ships, buildings, etc.

Introduction: Impact testing is probably the most popular excitation technique used in experimental modal analysis since the test setup is less complicated than for testing with a shaker. For lightweight or rotating structures, the hammer is more suitable than the shaker as the excitation source since it is almost nonintrusive. A hammer test is also called an impact test. As in modal analysis using a fixed shaker with a roving transducer, here also it is to be ensured that the excitation or response should not lie on the nodal line of any mode as the mode will not get excited.

Aim of the Test: The aim of the test is to find the modal parameters, i.e., natural frequencies, damping ratios, and mode shapes for the first four modes of vibration of a beam.

Theory: The theory behind the modal analysis method using roving impact and fixed transducer has already been discussed in Sect. 9.4.1.

Test Setup and Procedure: In this experiment, modal analysis of a beam is to be done. The testing can be done by using free-free boundary conditions or fixed-free boundary conditions. To achieve free-free conditions, the beam can be suspended from the ceiling at one end using a small chuck, a light spring and thin nylon rope; however, it has to be borne in mind that the flexible modes of vibration will be seen, along with the rigid body modes with this boundary condition. Free-free condition can also be achieved by supporting the beam on a soft sponge. Alternately, the test may be conducted with fixed-free boundary conditions. For this, the beam is to be fixed to a cement block at one end to arrest movement as shown in Fig. 10.31. The beam is discretized into a fixed number of points along the centre line if a one-dimensional model is required, or nodes may be created along both edges if a two-dimensional model is required (Fig. 10.32). It is then excited by an impact
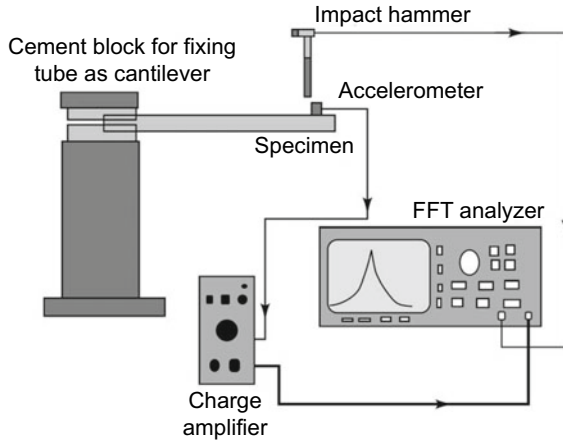
Fig. 10.31 Test setup for modal analysis
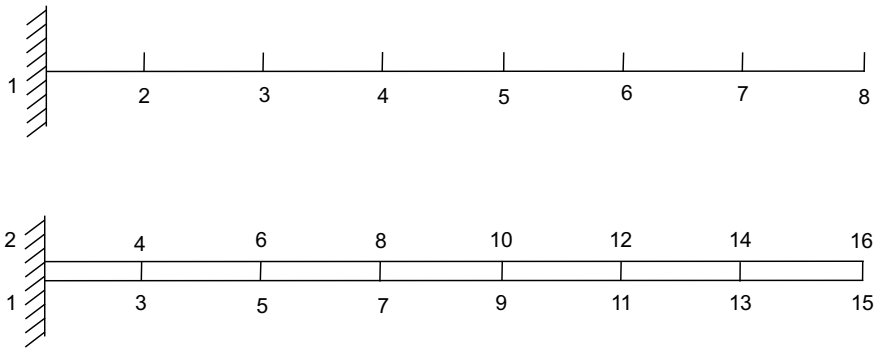
Fig. 10.32 Discretization of beam for modal analysis

hammer with plastic/steel tip depending on whether the expected natural frequencies are within 2 or 10 kHz. A piezoelectric accelerometer may be fixed using bee's wax or a magnetic base at a nodal location near the free end of the beam. Repositioning an accelerometer could be a time-consuming process if it is attached with wax and hence a fixed transducer location makes sense. The output of the accelerometer is conditioned using a charge amplifier. For obtaining the transverse modes of vibration, the direction of hit, as well as the orientation of the transducer have to be transverse to the beam.

FRFs can be obtained at all nodal locations if the number of nodes is small. However, if the number of nodes is large, then only one row or one column of the FRF matrix, as well as the diagonal elements of the FRF can be obtained; this will be enough to accurately represent the dynamic behaviour of the entire beam. A dual-channel dynamic signal analyser is used to read signals both from the accelerometer and the force transducer of the impact hammer and provide FRFs. At every nodal
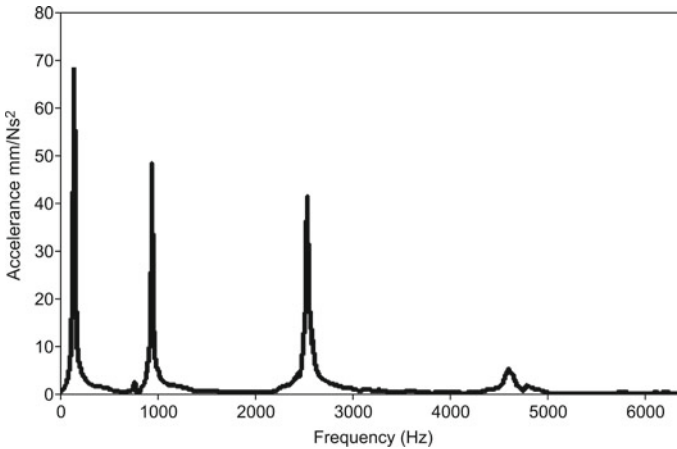
Fig. 10.33 Typical accelerance function

location, a minimum of three hits is given and the averaged FRF is obtained. Once all the required FRFs are obtained, the natural frequencies, damping ratios, and mode shapes corresponding to the first four modes of vibration can be extracted. Figure 10.33 shows a typical accelerance function. Damping ratio corresponding to each mode can be obtained using the zoom feature in the analyser and the half-power method for computation.

Questions

(1) What is the need for averaging the FRFs?
(2) Under what circumstances is it not advisable to place the pickup near the free end?
(3) What are the problems associated with modal analysis of highly flexible structures?
(4) How do you choose the frequency resolution of the FFT analyzer?

Bibliography

1. Beckwith, T. G., Marangoni, R. D., & Lienhard, J. H. (1993). *Mechanical measurements*. US: Pearson Education.
2. Brandt, A. (2011). *Noise and vibration analysis: Signal analysis and experimental procedures*. Wiley.
3. Buzdugan, Gh., Mihăilescu, E., & Rades, M. (1986). *Vibration measurement (mechanics: Dynamical systems)*. Netherland: Springer.
4. Collacott, R. A. (1979). *Vibration monitoring and diagnosis*. London: George Godwin Ltd.
5. Crocker, M. J. (2007). *Handbook of noise and vibration control*. New York: McGraw-Hill.

6. Figliola, R. S., & Beasley, D. E. (2014). *Theory and design for mechanical measurements*. Wiley.
7. Goldman, S. (1999). *Vibration spectrum analysis: A practical approach*. Industrial Press Inc.
8. Inman, D. J. (1989). *Vibration: With control, measurement, and stability*. Pearson College Div.
9. McConnell, K. G., & Varoto, P. S. (1995). *Vibration testing: Theory and practice*. New York: Wiley-IEEE.
10. Nakra, B. C., & Chaudhry, K. K. (2016). *Instrumentation, measurement and analysis*. McGraw-Hill Education India Private Limited.
11. Piersol, A. G. (2002). *Harris' shock and vibration handbook*. McGraw-Hill Professional, eBook.
12. Piersol, A., & Paez, T. (2009). *Harris' shock and vibration handbook* (McGraw-Hill Handbooks), McGraw-Hill Education.
13. Scheffer, C., & Girdhar, P. (2004). *Practical machinery vibration analysis and predictive maintenance (practical professional)*. Newnes.
14. Sinha, J. K. (2014). *Vibration analysis, instruments, and signal processing*. CRC Press.
15. Smith, J. D. (1989). *Vibration measurement and analysis*. London: Butterworth.
16. Sujatha, C. (2010). *Vibration and acoustics: Measurement and signal analysis*. Tata McGraw-Hill Education.
17. Wowk, V. (1991). *Machinery vibration: Measurement and analysis*. McGraw-Hill Education.
18. Venkateshan, S. P. (2008). *Mechanical measurements*. ANE Books India, and CRC Press.

Chapter 11
Experiments in Acoustics

INTERESTING FACTS: Anyone who has worked with acoustic or electronic devices would surely have heard of the deciBel (abbreviated dB). The deciBel began as the Transmission Unit defined by researchers at Bell Telephone Company in 1924 to replace the 'mile of standard cable', a unit of power ratio used in telephone engineering to measure loss in telephone and telegraph cables. The transmission unit was renamed the Bel (in honour of Alexander Graham Bell) in 1928. But the Bel was found too large for practical applications, so the deciBel (dB) began to be used. The apparatus and practices that were developed for hearing tests in both otology and telephone engineering were found to be convenient for city noise measurement. Engineers started quantifying noise in terms of loudness and the unit of the deciBel found prevalence. The deciBel is a measure of a power ratio and is a logarithmic quantity; it is, therefore, useful for denoting quantities like antenna or amplifier gain. It is also used for measuring sound and light, the reason being that human ears and eyes have a logarithmic response to those quantities, with a very large dynamic range. It is a unit widely used in electrical engineering and acoustics to denote the ratio between two values with the same dimensions. Over time, the deciBel began to be used to express ratios between two current levels, two voltage levels, or levels for any other physical quantities.

11.1   Measurement of Sound Pressure Level

Introduction: There is often a need for measuring the sound pressure level (SPL) or L_p emanating from a machine and also to find its directivity index. SPL measurements are also required for finding out vehicle pass-by noise, machinery noise in workshops and factories, as well as construction noise. Sound level measurement is also required to evaluate environmental noise, such as traffic noise and noise near an airport or in a commercial area. Other specialized applications include measurement and analysis of transient signals such as sonic booms, explosions, and gunshots.

Aim of the Experiment: To measure the SPLs of a source in the Linear and A scales, and also to determine the directivity pattern of the source. It is also required to do a spectrum analysis in the 1/1 and 1/3 octave bands.

Theory: The theory behind sound level measurement is described in Sect. 4.1.1. Details regarding the sound level meter (SLM) with linear and A scales, as well as 1/1 and 1/3 octave filters, have been described in Sects. 5.4.1, 5.4.2, and 5.4.3.

Test Setup and Procedure: At the outset, an appropriate noise source available in the laboratory must be identified for measurement. Then a suitable microphone of diameter half inch or 1 in. should be chosen depending on the frequency range of the sound produced by this machine. The microphone should then be connected at the front end of a sound level meter and calibrated using a piston phone producing a known sound pressure level at a known frequency. While calibrating, it is to be ensured that the microphone sits snugly in the piston phone cavity. The calibration will give the confidence that the measurements made subsequently are correct. The output of the SLM is connected to a central processing unit (CPU). The next stage is to make measurements of the background noise using the sound level meter with the calibrated microphone after switching off all fans, airconditioners, and noise-producing sources. It is to be ensured that people are not talking. Next, the sound-producing machine is switched on and the overall sound pressure level produced by the source is measured at a distance of 1 m from the source as shown in Fig. 11.1. If the background noise is less than the sound being measured by more than 10 dB, then it may be considered to be negligible; otherwise, it has to be subtracted from the measured noise. This has to be done for 1/1 and 1/3 octave analysis as well.

In order to find the directional characteristics of the source, it is customary to take sound pressure level readings 360° around the source. Readings are generally taken on dB Linear (LIN) and dB (A) scales. Next, in order to understand the spectral characteristics of the source, frequency analysis should be done in the 1/1 and 1/3 octave scales with centre frequencies from 31.5 Hz and 12.5 Hz, respectively, up to 16 kHz. The use of the A-weighting filter is to mimic the response of the human ear, while the LIN weighting gives an overall value as would be sensed by a microphone having a flat frequency response in the audio-acoustic frequency range. SPL in 1/1 octave bands may be plotted as shown in Fig. 11.2a. SPL versus angle in any 1/1 or
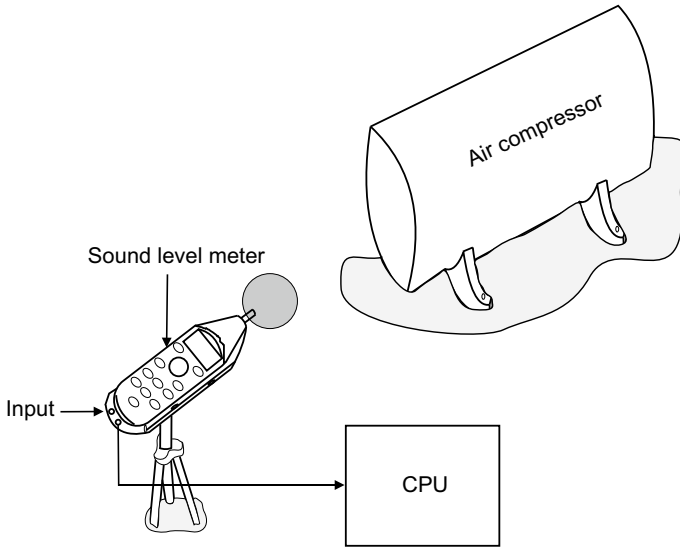
Fig. 11.1 Sound level measurement setup

1/3 octave band may be plotted as shown in Fig. 11.2b. The directional properties of the source may also be tabulated as shown in Table 11.1 as a function of 1/1 octave bands and as a function of 1/3 octave bands as shown in Table 11.2. SPL as a function of the distance from the source may be plotted as shown in Table 11.3.

Repeat the measurements at double the distance and four times the distance from the source and tabulate the results. Plot the SPL in dB (LIN) got above as a function of the logarithm of the distance from the source. Comment on your observation. Using the SPL values measured in the 1/1 octave bands, the overall dB (LIN) as given in Eq. 4.81 has to be calculated. It is to be noted that L_{p1}, L_{p2}, etc. in this equaiton now refer to the SPLs in the various frequency bands. Also, SPLs may be measured using a different source or may be even busy traffic noise.

Questions

(1) Comment on the differences in SPL readings you obtained with the LIN and A scales.
(2) Why is the background noise considered to be negligible if it is less than that of the source by 10 dB?
(3) Compare the overall dB (LIN) value calculated using Eq. (4.81) from the SPL values measured in the 1/1 octave bands with the measured overall dB (LIN) value.
(4) What is 0 dB plus 0 dB?
(5) Why are 1/1 and 1/3 octave bands preferred in acoustics over constant-frequency bandwidth analysers?
(6) Which frequency band is the human ear most sensitive to?
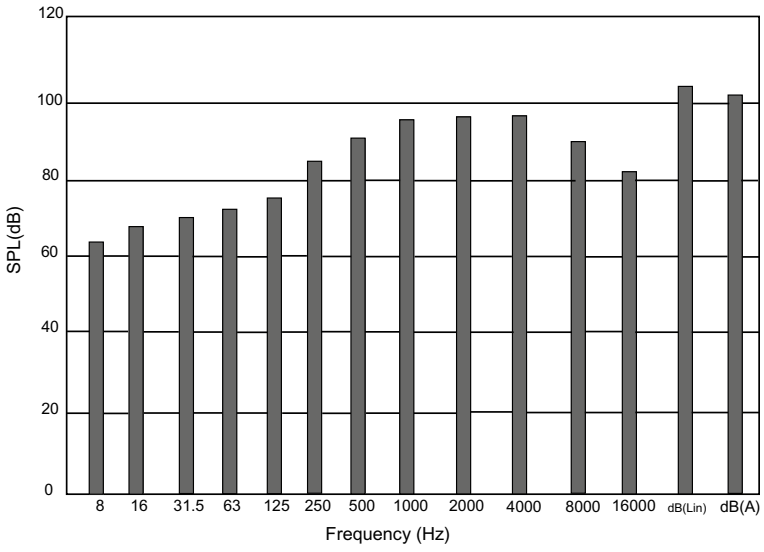
(a)
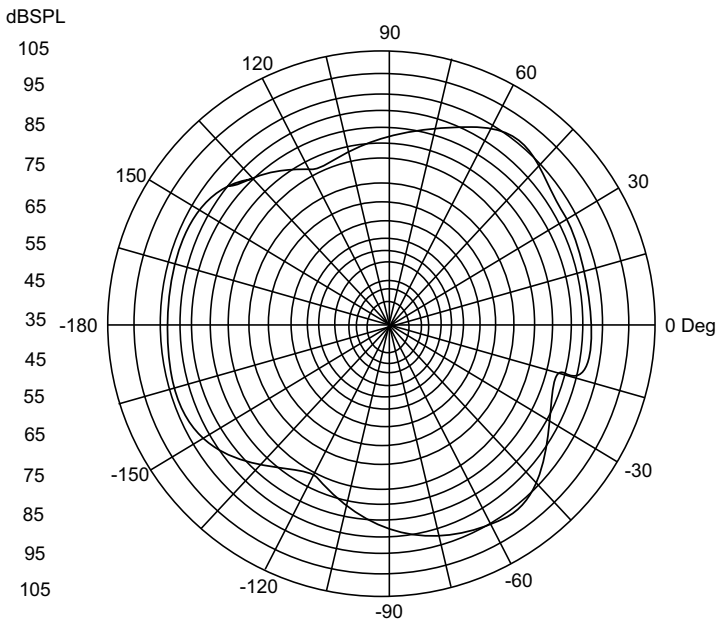
(b)

Fig. 11.2 Measured SPL: **a** in 1/1 octave scale, **b** versus angle

Table 11.1 SPL around the source as a function of 1/1 octave bands

Pos.	Hz										Overall SPL dB (LIN)	Overall SPL dB (A)
	31.5	63	125	250	500	1k	2k	4k	8k	16k		
0°												
45°												
90°												
135°												
180°												
225°												
270°												

Note: Pos. denotes angular position around source

Table 11.2 SPL around the source as a function of 1/3 octave bands

Pos.	Hz																	
	12.5	16	20	25	31.5	40	50	63	80	125	160	200	250	315	400	500	630	800
0°																		
45°																		
90°																		
135°																		
180°																		
225°																		
270°																		

Table 11.2 Continued.

Pos.	Hz															Overall dB (LIN)	Overall dB (A)
	1k	1.25k	1.6k	2k	2.5k	3.15k	4k	5k	6.3k	8k	10k	12.5k	16k	20k			
0°																	
45°																	
90°																	
135°																	
180°																	
225°																	
270°																	

Note: Pos. denotes angular position around source

Table 11.3 SPL versus distance from source

Sl. No.	Distance (m)	SPL dB (LIN)
1.		
2.		
3.		
4.		

11.2 Measurement of Sound Power from Sound Pressure Levels: ISO 3744 and 3745

DID YOU KNOW that for an omnidirectional source in free space, sound power level in dB is equal to the sound pressure level in dB (with reference to 20 microPascal) at a distance of 0.2821 m from the source? The sound powers that we come across in our day-to-day lives vary from around 10^{-11} W (10 dB) to around 100 MW (200 dB) with a reference power of 10^{-12} W (0 dB).

Introduction: The strength of a noise source is completely described by two quantities: the sound power level and its directivity. The sound power level is an indication of the total sound power radiated by a source in all directions. It can be measured as an overall quantity or in 1/1 or 1/3 octave bands like sound pressure level. Directivity gives an indication of the radiation with direction around the source and is also measured in each frequency band. For machines with pronounced directivity patterns such as a jet engine, a considerable amount of data is required. Some sources are, however, non-directional, radiating uniformly in all directions (omnidirectional). These sources are small compared to the wavelength of the sound they radiate. The International Standards Organization has published six standards: 3741–3746 which outline different methods for measurement of sound power. The first three deal with measurements in a diffuse field which is difficult to simulate in the laboratory if a reverberation chamber is not available. The last three deal with measurements in a free-field environment which can be obtained in an anechoic chamber or in a semi-anechoic chamber or in a vast open field with a reflecting ground.

Aim of the Experiment: To measure the sound power from sound pressure levels using ISO 3744 and 3745.

Theory: This method assumes that the sound power level of a source can be computed using sound pressure levels measured on a hemispherical surface. The sound power level of the source can be calculated for each frequency band from Eq. (11.1).

$$L_w = L_p + 10 \log_{10} \left(\frac{2\pi r^2}{S_{\text{ref}}} \right) - K - B \tag{11.1}$$

where L_w is the band sound power level in dB, L_p is the mean surface sound pressure level over the test hemisphere in dB, r is the radius of the hemisphere, S_{ref} is the reference area ($1\,\text{m}^2$), K is the band environmental correction factor in dB, and B is the pressure and temperature correction term used when the atmospheric temperature and pressure differ significantly from $20\,°C$ and 10^5 Pa, respectively. The factor K may be taken as 0.5 dB.

The pressure and temperature correction terms may be calculated as per Eq. (11.2) for air.

$$B = 10 \log_{10} \left\{ \frac{p_x}{10^5} \right\} + 5 \log_{10} \left\{ \frac{306}{273 + t_x} \right\} \tag{11.2}$$

Here, p_x and t_x are the pressure and temperature at the location in Pa and $°C$, respectively. B works out to 0.06. For normal conditions, B may be taken as zero for air.

Associated with each microphone position is a surface area, S_i, over which the average sound pressure level is assumed to be equal to that measured at the microphone. The sound power level in each frequency band of interest can be calculated from the following equation.

$$L_w = 10 \log_{10} \left[\sum_{i=1}^{N} \left(10^{(0.1 L_{pi})} \times S_i \right) \right] - K - B \tag{11.3}$$

where L_{pi} is the band sound pressure level at the ith measuring position in dB and S_i is the surface area associated with the ith measuring position in m^2.

The overall sound power level can be calculated from the 1/1 or 1/3 octave band sound power levels for each band as per the expression

$$L_{w,\text{overall}} = 10 \log_{10} \sum_{j} 10^{0.1 L_{w,f_j}} \tag{11.4}$$

where L_{w, f_j} is the sound power level in the jth 1/1 or 1/3 octave band. It is possible to determine the directivity index (DI) when the mean surface sound pressure level has been determined. For measurements over a hemisphere, the directivity index in dB in the ith direction can be calculated from

$$\text{DI} = L_{pi} - L_p + 3 \tag{11.5}$$

Test Procedure as per ISO 3744 and 3745: ISO 3745 is a precise method for determination of sound power levels of a noise source from anechoic and semi-anechoic rooms, while ISO 3744 gives recommendations on measurement surfaces. The location of the source under test, the measurement surfaces, and the microphone
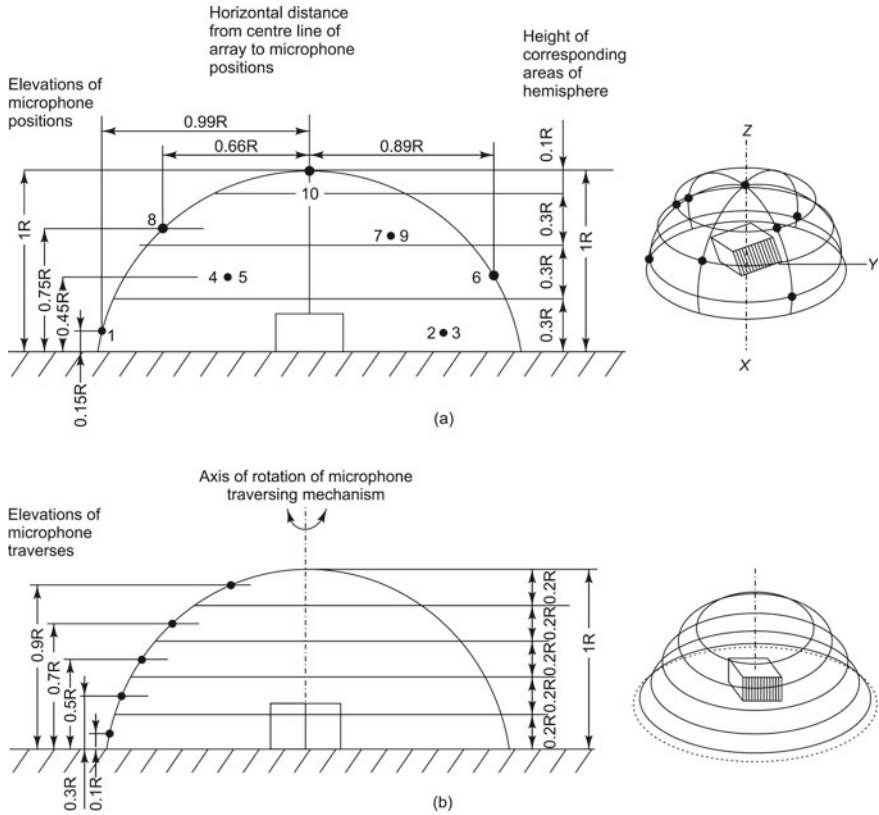
Fig. 11.3 Microphone positions: **a** on equal areas on the surface of a hemisphere, **b** with traversing mechanism

positions are defined by a coordinate system. The most accurate sound power level values can be obtained by making measurements in an anechoic chamber with the source suspended from the ceiling at the centre of the room. However, such a chamber is expensive and is not always available in all laboratories; besides, some sources are too heavy to be suspended, or are associated with a reflecting plane. In such cases, measurements can be carried out in a semi-anechoic chamber, where the equipment is placed on a hard reflecting surface with an absorption coefficient less than 0.06. The type of measurement surface may be chosen based on the shape and size of the noise source, with the distance from each microphone to the noise source approximately the same. In order to obtain the average value of the mean square pressure on the measurement surface, an array of fixed microphone positions should be used, the locations of which may be defined by a hypothetical reference box. Besides, the measured power levels tend to be most accurate when most of the sound energy passes through the measurement surface at normal incidence. Based on this, ISO 3744 suggests three types of measurement surfaces:
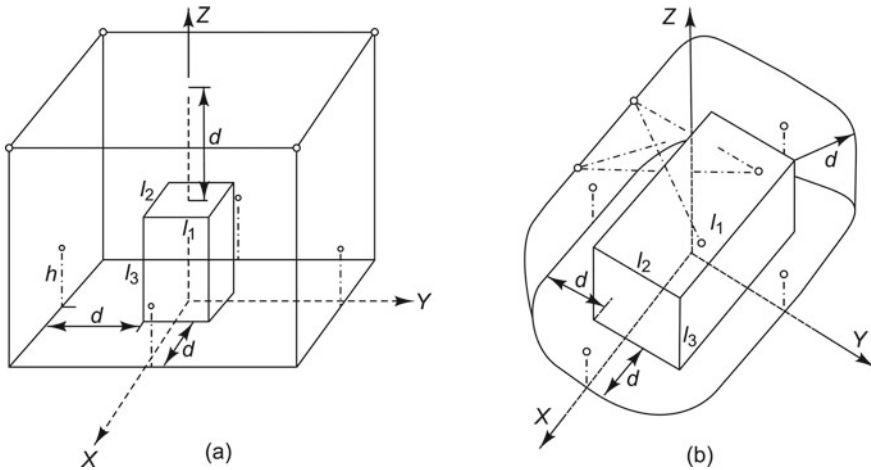
Fig. 11.4 Microphone locations: **a** on parallelepiped, **b** on conformal surface

 (i) For a relatively small source, a hemispherical surface of radius r may be used with 10 microphone positions as shown in Fig. 11.3. A single microphone may be moved from positions 1 to 10, or an array of fixed microphones may be used and their outputs sampled sequentially.

 (ii) For a long box-like source, a rectangular parallelepiped is preferred, each side being spaced a distance d (the measurement distance) from the corresponding side of the reference box. The sides of the parallelepiped should be parallel to those of the reference box. The number of microphone locations could be 10, 20, or 40 as shown in Fig. 11.4.

(iii) For a tall, but not wide or long source, a cylinder, half-cylinder, or quarter-cylinder of diameter $2R$ and height h may be the preferred measurement surface as shown in Fig. 11.5.

(iv) A combination of two segments, each being hemispherical, rectangular, or cylindrical in form, may also be used.

Instrumentation and Measurement The sound source may be any noise-producing machine available in the laboratory. Since different measurement surfaces have different requirements for the minimum distance between the microphones and the source, one has to consider other aspects like the amount of background noise in the test room and the size of the source relative to the usable measurement space of the test environment while deciding on the appropriate measurement surface. The measurement point locations may be found as shown in Figs. 11.3 or 11.4 or 11.5. The points in the X–Y-plane may be marked on the ground with a suitable marker. The heights in the Z-direction may be marked on a marker pole, the pole being slim enough so as not to cause any appreciable disturbance in the free-field environment. Measurements may be made with a sound level meter or sound analyser with a $\frac{1}{2}$ inch condenser microphone at the front end in the case of machinery frequencies up to 40

A : $4.5h_s/5$ B : $3.5h_s/5$ C : $2.5h_s/5$ D : $1.5h_s/5$

E : $0.5h_s/5$ F : $h_s/5$ G : h_s

A : $3.5R/4$ B : $2.5R/4$ C : $1.5R/4$

D : $0.5R/4$ E : $R/4$ F : R
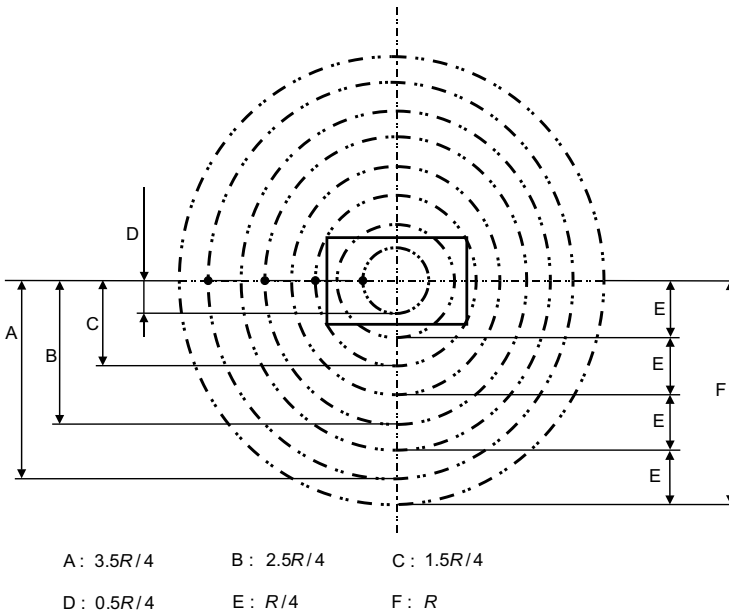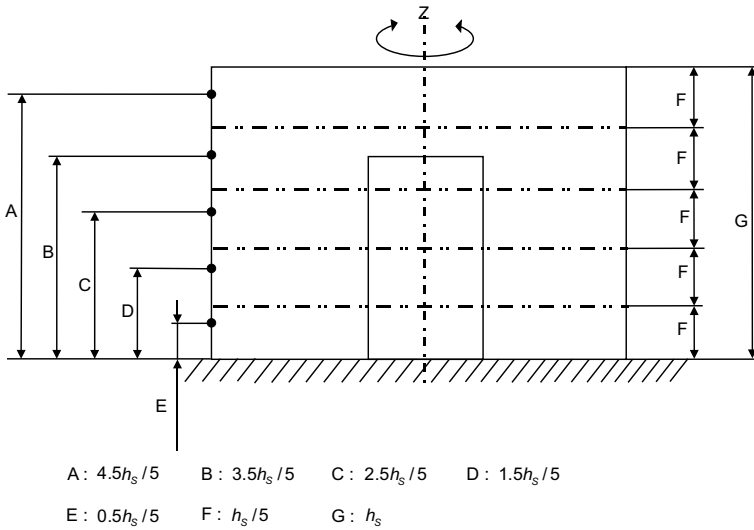
Fig. 11.5 Microphone array with five side microphones and top four microphone paths on a cylindrical surface. *Note:* •: microphone positions, h_s: height of measurement surface, R: measurement surface radius, and Z: axis of rotation of microphone traversing mechanism

kHz or a 1 in. microphone for frequencies up to 15 kHz. Sound pressure levels in 1/3 octave bands, on 'A-weighted scale, from 3.15 Hz to 10 kHz should be measured at the 10 points on the hemisphere. The measurement time for each reading may be set to 30 s to ensure effective averaging. Readings may also be taken in the LIN scale to provide actual sound pressure levels produced by the machine. The microphone should be oriented to face the source so that the angle of incidence of the sound waves on the microphone is that for normal incidence. Three methods may be used for measurement:

(i) In the first method, a single microphone is moved from positions 1 to 10 manually, as shown in Fig. 11.3a, or an array of fixed microphones (if available) is used and their outputs are sampled sequentially.

(ii) In the second method, a single microphone is scanned at constant speed along horizontal circular paths successively. A minimum of five circular paths should be used, as shown in Fig. 11.3b, where the annular area of the hemisphere associated with each path is equal. Alternately, the microphone could be placed at the respective heights, while the sound source is rotated at a constant speed using a turntable (in the case of a small source).

(iii) The third method requires scanning a single microphone along a meridional arc, i.e., the microphone is traversed along a quarter of a circular arc about a horizontal axis to the centre of the source. At least eight microphone traverses at equal azimuthal angles around the source should be carried out and may be achieved by rotating the source through 45° before each traverse.

A representative plot of sound pressure levels for the 1/3 octave band spectrum in 'A-weighted scale from 3.15 Hz to 10 kHz is shown in Fig. 11.6. Table 11.4 shows how the readings may be tabulated.
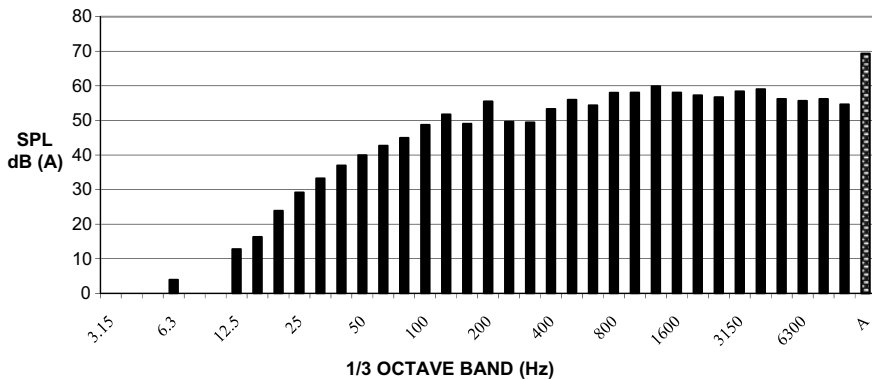
Fig. 11.6 Typical sound spectrum

Table 11.4 Readings for computation of sound power level-ISO 3744 and 3745

$i \longrightarrow$		1	2	3	4	5	6	7	8	9 10	L_{w,f_j} (dB)	DI_{f_j} (dB)
$S_i\,(m^2)$ of the i th microphone location $\longrightarrow$												
jth freq. band	1/1 Octave band centre freq. (Hz)	L_{pi} (dB) at ith location										
1	31.5											
2	63											
3	125											
4	250											
5	500											
6	1 k											
7	2k											
8	4k											
9	8k											
10	16k											
$L_{w,\text{overall}}$ (dB)												

Questions

(1) Calculate the overall SPL at each location.
(2) Draw the directivity pattern of the sound power level.
(3) Why is it better to characterize a sound source in terms of its sound power level and not sound pressure level?
(4) Knowing the overall sound power level, calculate the sound power.
(5) What are the main sources of error in this measurement?

11.3   Measurement of Sound Power from Sound Pressure Levels: ISO 3746 Engineering Method

DID YOU KNOW that sound has a comparatively low level of energy when compared to other forms of energy and hence is not used for generating electricity? Measurement of sound power level is often done because legislation demands this, but sometimes it is also done to understand the effect of noise on human beings. Quite often it is done for estimating sound power levels of identical machines/products to compare their performances and, if necessary, to redesign the equipment to make them quieter. For this comparison, it is imperative that

measurements are made in accordance with an international standard. This helps if consumer organizations have to be brought into the picture later.

Introduction: This standard describes methods of accuracy grade 2 or engineering grade as defined in ISO 12001. Here, the measurements are made in an environment that is approximately an acoustical free field over a reflecting plane. Such a space can be found in specially designed rooms or inside industrial buildings. The sound source may also be mounted on a hard reflecting plane in a large open space. For sources fixed on the floor of machine rooms, there are usually undesired reflections from nearby objects, walls, and the ceiling resulting in background noises, and corrections are defined to account for these.

Aim of the Experiment: To measure the sound power of a source from sound pressure levels as per ISO 3746 engineering method.

Theory: The theory is the same as described in Sect. 11.2.

Measurement Procedure: Initially, the microphone, along with the sound level meter, is calibrated using a piston phone. Then the background noise level is measured. If the SPL of the background noise level is less than that of the sound source by more than 10 dB, then the effect of the former may be neglected. For sources usually mounted in and/or to be measured in large open areas under satisfactory acoustical conditions, a large measurement distance is to be chosen, and in this case, the hemispherical measurement surface is preferred. Measurements on large machines (having a largest dimension of 15 m) can be carried out at the site where the contributions of the reverberant field to the sound pressures on the measurement surface are small in comparison with those of the direct field of the source. Since the test objects may be of various sizes and shapes, the measurement surfaces may be as before:

 (i) A hemispherical surface (Fig. 11.3) with a radius more than twice the major source dimension or four times the average distance of the source from the reflecting plane, whichever is larger, and in any case not less than 1 m: four microphone positions may be used in this case.
(ii) A rectangular parallelepiped or a conformal surface (Fig. 11.4), which is the same as a rectangular parallelepiped, except that the corners are rounded and formed by portions of cylinders and spheres. Though the conformal surface is expected to give more accurate results, greater effort is required for positioning the microphone, and therefore, the use of the hemispherical surface is generally preferred.

If measurements are made at the site, the ratio of the room absorption *a* (defined in Eqs. 4.157 and 4.161) to the measurement surface area should be greater than or equal to 6; the higher the value of this ratio, the better. The total absorption of the room (m^3) can be determined in each octave band using the formula

$$a = 0.16(V/T) \qquad\qquad (11.6)$$

where V is the volume of the room in m^3 and T is the reverberation time of the room in seconds.

Five points of measurement may be used. The procedure is the same as discussed in Sect. 11.2. Sound power can be found by adding the products of the areas times the acoustic intensities for the areas on any hypothetical surface containing the source as defined in Eqs. 11.3–11.5 as before. Figure 11.6 and Table 11.4 may be used for these measurements.

Questions

(1) When are sound pressure levels and sound intensity levels almost the same?
(2) What are the main sources of experimental error in this experiment?
(3) Why is it recommended that the ratio of the room absorption a to the measurement surface area be greater than or equal to 6?
(4) If a hemispherical surface is used, why should its radius be more than twice the major source dimension?
(5) How do you find out if the environment in which measurements are being made is approximately an acoustical free field over a reflecting plane?

11.4 Measurement of Sound Power from Sound Pressure Levels in Diffuse Field Environment: ISO 3741–3743

DID YOU KNOW that in a diffuse/random/random-incidence sound field (i) the sound pressure level is uniform at every location within the field at any instant in time and (ii) all angles of incidence have an equal probability at any instant in time? A purely diffuse sound field is different from a reverberant sound field, which is also diffuse to some extent, in that acoustic energy should not persist in a diffuse sound field beyond the time it is created as it does in a reverberant sound field. For practically creating a truly diffuse sound field with random angles of incidence in a bounded space, the surfaces of the objects and boundaries should be irregularly shaped so as to create reflections and diffraction of sound waves. Such a field is used in audiometry to ascertain hearing loss.

Introduction: ISO 3741 to 3743 deal with measurements in a diffuse field, which is a little more difficult to obtain than a semi-anechoic condition. The diffuse field is another type of well-defined environment for evaluation of sound power and can be obtained in a highly reverberant enclosure, which is relatively cheaper to construct than an anechoic chamber. The accuracy of this method depends on the diffuseness

of the field, and this is not always easy to achieve, especially at very low frequencies and when the source radiates pure tones. The lowest frequency of interest decides the minimum volume of the room. A volume of $200\,\mathrm{m}^3$ is required for the $125\,\mathrm{Hz}$ octave band. A volume larger than that required may bring about non-uniformity of the reverberant fields in the higher frequency bands due to air absorption. For very precise measurements, the guidelines given in ISO 3741 may be followed. When very high precision is not required and when cost and labour involved in measurements are of prime concern, the method described in ISO 3743 may be followed. This method can be used for testing small machines in a special test room with a minimum room volume of $70\,\mathrm{m}^3$, the maximum size of the source being $0.7\,\mathrm{m}^3$.

Aim of the Experiment: To measure the sound power of a source from sound pressure levels as per ISO 3741 to ISO 3743.

Theory: In a diffuse field, the steady-state sound energy is equal to the sound energy transmitted by the source minus that absorbed by the room boundaries. Therefore, L_w, the sound power of the equipment under test, can be calculated as

$$L_\mathrm{w} = L_p - 10\log_{10}\left(\frac{T}{T_0}\right) + 10\log_{10}\left(\frac{V}{V_0}\right) + 10\log_{10}\left(1 + \frac{S\lambda}{8V}\right)$$
$$+ 10\log_{10}\left(\frac{B}{B_0}\right) - 14\ \mathrm{dB} \tag{11.7}$$

where L_p is the mean band sound pressure level, T is reverberation time (RT) in seconds, $T_0 = 1$ s, V is the volume of room in m^3 , $V_0 = 1\ \mathrm{m}^3$, S is the total surface area of the room in m^2, λ is the wavelength at the centre frequency of the band in m, B is the barometric pressure in millibar for which logarithm is taken, and B_0 is the reference pressure of 1000 millibar.

Creation of a Diffuse Field: A room with hard reflecting surfaces such as concrete walls may be used. The prescribed value of RT for such rooms is met at higher frequencies, though it is usually high in the low and mid-frequency range. For reduction of RT at low frequencies, membrane absorbers may be used, and for middle and high frequencies, perforated panels with mineral wool interiors may be used. Initially, the RT of the room should be measured in 1/3 octave bands and should be normalized with respect to the RT at $1000\,\mathrm{Hz}$. If the normalized curve falls within the prescribed limiting curves, the enclosure can be used; else, it should be adapted to get the required changes. The following guidelines may be used for creating a diffuse field in the measurement chamber.

 (i) Multiple loudspeakers which are at random spatial orientations relative to the measurement position may be used.
(ii) Signals from the individual multiple loudspeakers must be random, i.e., the acoustic energy from the different sources should be instantaneously uncorrelated in amplitude, phase, and frequency.

Table 11.5 Readings for computation of sound power level

$i \longrightarrow$		1	2	3	4	5	6	7	8	9	10	L_p (dB)	λ (m)	T (s)	L_w (dB)
$S_i\,(m^2)$ of the ith microphone location $\longrightarrow$															
	1/1 Octave band	L_{pi} (dB) at ith location													
jth freq. band	**centre freq. (Hz)**														
1	31.5														
2	63														
3	125														
4	250														
5	500														
6	1 k														
7	2k														
8	4k														
9	8k														
10	16k														
$L_{w,overall}$ (dB)															

(iii) The physical dimensions of the loudspeaker baffle and cabinetry should be small so that they do not influence high-frequency diffraction.

(iv) The loudspeakers should be positioned far enough away from the centre of the measurement position so that this position is in the far field (for practical purposes, a spacing of 1 m serves the purpose).

(v) Equidistant placement of the loudspeakers from the measurement location is acceptable because uncorrelated signals are being used.

(vi) The excitation signal used to create a diffuse field is important. Multiple periodic signals are preferred over a single harmonic signal, but many of these will still cancel or reinforce each other. Hence, broadband random noise is the most ideal signal, as it will generate the maximum interference of reflected waves.

Test Equipment and Procedure: If multiple sources are not available, a single appropriate sound source should be chosen and placed near a corner of the test room with diffuse field or reverberation chamber. A calibrated $\frac{1}{2}$ inch microphone connected to a sound level meter should be made to scan a path of at least 3 m at a constant speed. Alternatively, discrete averaging with a fixed number of microphones may be done to yield more accurate results at the expense of cost. Table 11.5 may be used to record the measurements.

Questions

(1) What is the check for diffusivity of the sound field in a chamber?
(2) What are the main sources of error in measurement?
(3) What materials can you use to line the walls, ceiling, and floor of a room to make it reflective?
(4) How does a warble oscillator help in the absence of a random source?
(5) Why is the sound source placed near the corner of the room?
(6) Why are reverberation times of rooms different in different frequency bands?

11.5   Sound Intensity Measurement Using Two-Microphone Method

> **INTERESTING FACTS:** Acoustic measurements and theory have not progressed in tandem. The quantity sound intensity was described in the book 'The Theory of Sound' by Lord Rayleigh in 1877. But it was only after a century that a practical method for measuring sound intensity was devised. Developments in electronics at the beginning of the twentieth century helped bring measurement in step with theory. In 1977, digital signal processing techniques were applied to acoustic theory independently by F. J. Fahy and J. Y. Chung. With this development and with advances in microphone design, measurement of sound intensity became possible with two closely spaced microphones. Today, sound intensity measurement is extensively used for the identification of the source of noise in a machine.

Introduction: Intensity measurement is a relatively new technique and helps locate and identify noise sources or sinks in a given complex environment. Generally, a good correlation exists between the the sound radiated by a source and its vibrational velocity. Therefore, the various mechanisms of sound generated by the source are accounted for in intensity measurements. Intensity measurements can be used for the determination of sound power of noise sources, sound absorption coefficient, and TL of structures in addition to the noise source location. Sound intensity is defined as the time-averaged rate of flow of energy through unit area of the medium as has been discussed in Chap. 4. Being a vector, it gives the direction of acoustical energy flow from the source to the sinks (or absorbers) and can be measured using a specially made probe, which incorporates a pair of highly phase-sensitive, phase-balanced condenser microphones, such that the gradient in sound pressure (the driving force of acoustical energy flow) and mean sound pressure level at any space point can be simultaneously taken into account to compute the intensity. The advantages of using intensity measurement are the following:

 (i) This technique excludes any influence from steady background noise, provided there is no absorption within the enclosing surface.
 (ii) No anechoic or reverberation test chamber is needed.
 (iii) Sound measurements can be carried in the near field or far field.
 (iv) There is no restriction on the shape of the surface on which measurements are made.
 (v) Sound power can be directly calculated from the measured intensity.

Aim of the Experiment: To measure the sound intensity of a source using the two-microphone method.

Theory: Sound intensity is the time-averaged product of the instantaneous pressure $p(t)$ and the corresponding particle velocity $u(t)$ in a given direction at a point in the sound field.

$$I = \frac{1}{T} \int_T p(t)u(t)\mathrm{d}t \tag{11.8}$$

The pressure may be computed as the average value of pressures p_1 and p_2 obtained from two identical closely placed microphones A and B.

$$p = \frac{p_1 + p_2}{2} \tag{11.9}$$

The particle velocity is related to pressure gradient according to the linearized Euler equation (Eq. 4.4) as

$$\rho_0 \partial u / \partial t = -\mathrm{grad}\ p \tag{11.10}$$

where ρ_0 is the mean mass density of air, p is the instantaneous pressure, and u is the corresponding instantaneous particle velocity in the direction of propagation.

The particle velocity can be computed by integrating the pressure gradient signal, which is measured by obtaining the pressures p_1 and p_2 at locations A and B and dividing the pressure difference by the microphone separation distance Δr. This computation is valid as long as the distance between the microphones is very small compared to the wavelength of the sound being measured. The particle velocity can be written as

$$\frac{\partial u}{\partial t} = -\frac{1}{\rho_0} \frac{p_2 - p_1}{\Delta r} \tag{11.11}$$

$$u \approx -\frac{1}{\rho_0 \Delta r} \int (p_2 - p_1)\mathrm{d}t \tag{11.12}$$

where u is the particle velocity in the r-direction. The sound intensity in the r-direction is then

$$I_r = \langle pu \rangle \approx \left\langle -\frac{(p_1 + p_2)}{2\rho_0 \Delta r} \int (p_2 - p_1)dt \right\rangle \tag{11.13}$$
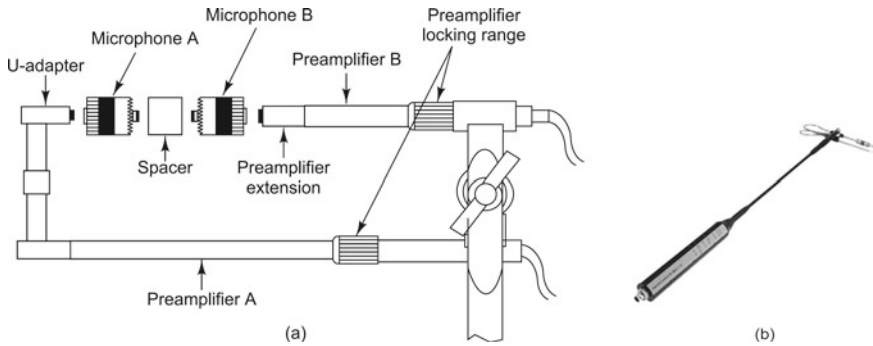
Fig. 11.7   Sound intensity probe: **a** schematic, **b** photograph. (Courtesy of Brüel & Kjær, Denmark)

Table 11.6   Microphone sizes, spacer sizes, and frequency range of operation

Sl. No.	Microphones (in. pair)	Spacer size (mm)	Frequency range (Hz)
1.	1/2	12	100–500
2.	1/2	50	50–1250
3.	1/4	6	200–10,000
4.	1/4	12	200–5000

where $\langle \; \rangle$ indicates expectation or time averaging.

Instrumentation and Procedure for Measurement of Sound Intensity

The sound intensity measurement system consists of an intensity probe, a real-time analyser, and a post-processing system. The probe kit itself has two closely spaced, face-to-face microphones with a spacer, as well as a control handle as shown in Fig. 11.7a, b. The 1/2-inch and 1/4-inch intensity microphone pairs available in the market are carefully manufactured and selected to have identical phase response and a flat amplitude response as a function of frequency. There are also phase-corrector units which allow for precision low-frequency phase matching, giving rise to increased measurement range and accuracy. Microphones of different sizes and different spacer lengths are used in these probes for different frequency ranges. Table 11.6 gives the commercially available sizes of microphones, spacers, and the frequency ranges of operation.

The principle of sound intensity measurement is as shown in Fig. 11.8. The sound intensity-measuring system uses a two-microphone probe with amplifiers, analogue-to-digital converters, and 1/1 and 1/3 octave filter sets, summing, multiplying, and averaging circuits which finally give the sound intensity in various frequency bands.
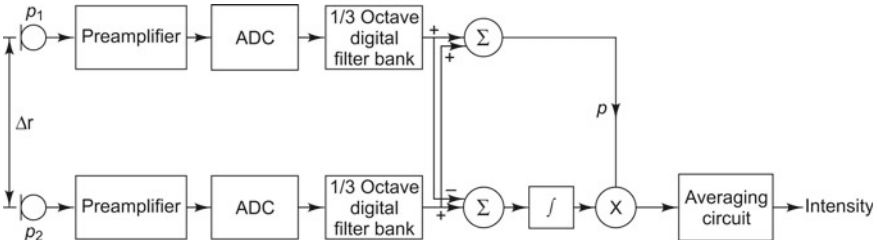
Fig. 11.8 Block diagram showing sound intensity measurement
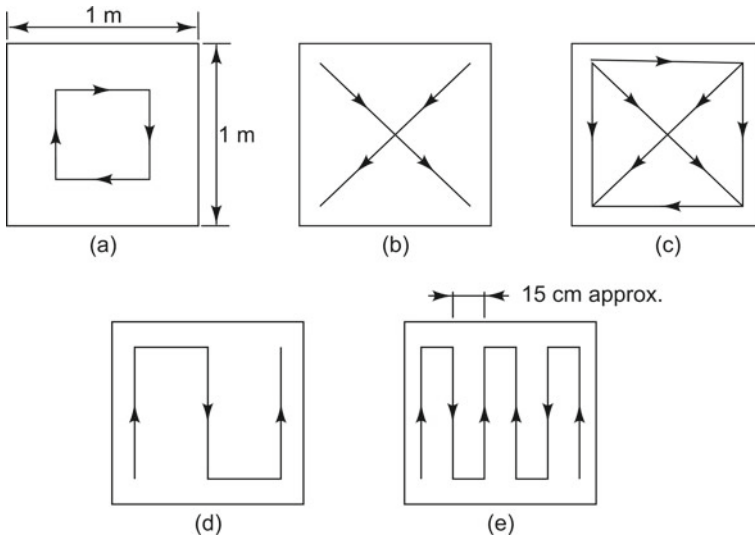
Fig. 11.9 Commonly used scanning patterns and errors: **a** 0.9 dB, **b** 0.2 dB, **c** 1.3 dB, **d** 0.9 dB, and **e** 0.7 dB

Methods of Sound Intensity Measurements The two methods of measurements commonly used are the sweep or scan method and the point method.

Sweep or Scan Method In this method, the intensity probe is slowly swept over the hypothetical surface enclosing the noise source, while the analyser is made to continuously average the measurements. Figure 11.9 shows the scanning patterns which can be employed.

These different patterns give different errors in the sound power obtained. This error can be quantified as ΔL_w and is the difference between the reference power source and the power measured by the intensity probe. This method is much faster and more accurate than the point method described subsequently. The errors obtained with the various scanning patterns employed and depicted in Fig. 11.9 are given by $\Delta L_w =$ (a) 0.9, (b) 0.2, (c) 1.3, (d) 0.9, and (e) 0.7 dB.

Point Method In this method, the intensity is computed based on measurements made at fixed points (usually forming a grid). These measurements can be used for obtaining intensity maps for noise source identification. For ΔL_w less than 1 dB, approximately $0.09\,\text{m}^2$ of area per point is required. For good results, $0.04\,\text{m}^2$ of area per point should be used.

Errors There are some errors associated with the sound intensity measurement technique and they are due to phase mismatch, finite difference approximation, near field, and statistical errors. These errors impose restrictions on the measurement frequency range, both the high-frequency and low-frequency limits. The low-frequency limit is decided by the amount of phase mismatch between the two microphones and the channels of the FFT analyser and can be minimized by the following precautions:

- Precise matching of the p and u channels, i.e., by choosing a pair of microphones with a low-phase mismatch (below 0.5° from 100 to 10,000 Hz).
- Adjusting Δr accurately.

The high-frequency limit in intensity measurements is due to the approximation of the pressure gradient by a finite pressure difference and this is called bias error. The orientation of the probe θ with respect to the source direction also gives rise to errors in measurement. Figure 11.10 shows the probe orientation θ.

The sum and difference of p_1 and p_2 are good approximations for p and $\text{grad}\,p$, respectively, if $\Delta r \approx \lambda$, where λ is the wavelength. If Δr remains constant (for a selected spacer), the error decreases with a decrease in θ. Therefore, by choosing a spacer and probe orientation of 0°, an error in intensity level for a given frequency can be determined.

Procedure Sound intensity measurements can be conducted in an open field or inside the laboratory depending on the sound source available in the laboratory. A hypothetical rectangular box may be chosen around the source to map the intensities. This box may be divided into many subareas which can be scanned and mapped for intensity. The sound analyser should be calibrated and configured for the required measurements. The probe may be kept in external gated mode, and surface areas of subsurfaces may be recorded for subsequent power calculations. A hypothetical rectangular box as shown in Fig. 11.11 with appropriate dimensions may be used. This box is divided into subareas for ease of scanning and analysis. The subareas may

Fig. 11.10 Probe orientation θ
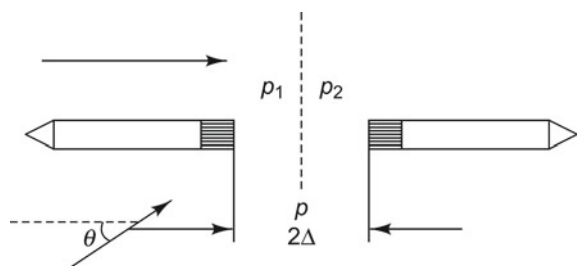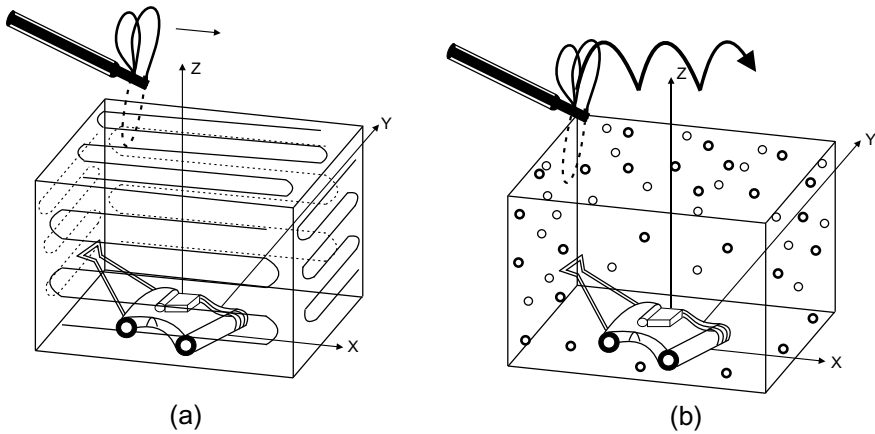
Fig. 11.11 Subarea and scanning pattern for sound intensity measurement: **a** sweep method, **b** point method

be made of uniform sizes, or sizes may be decided such as to highlight the probable sound sources to obtain maximum information. The rectangular box surrounding the source may be marked on the ground for reference. A mild steel or wooden bar grid may be fabricated for providing a reference grid for scanning the subareas, with the probe remaining perpendicular to the subareas during measurements. Each subarea may be scanned with an approximate speed of 0.25 m/s and with appropriate spacing between the sweeps. The measured data should be stored after each measurement and the grid shifted to the next subarea; this process should be repeated till the four sides around the source and the top have been scanned and intensity maps recorded for each subarea as a trace to complete the multispectrum in 1/1 or 1/3 octave bands in the required frequency range. Figure 11.11 shows a typical subarea and the scanning pattern. Figure 11.12 shows a typical intensity spectrum. The subareas with high intensities indicate the location of the sound source.

Questions

(1) Why are sound pressure and intensity levels the same?
(2) Why and how can sound intensity be used to determine sound power from near-field measurements?
(3) Why is sound power calculated from measured intensity values not affected by steady background noise?
(4) What is the effect of unsteady background noise?
(5) Why can you not measure intensity over the entire audio-frequency range using an intensity probe with a single spacer?

Fig. 11.12 Typical sound intensity spectrum
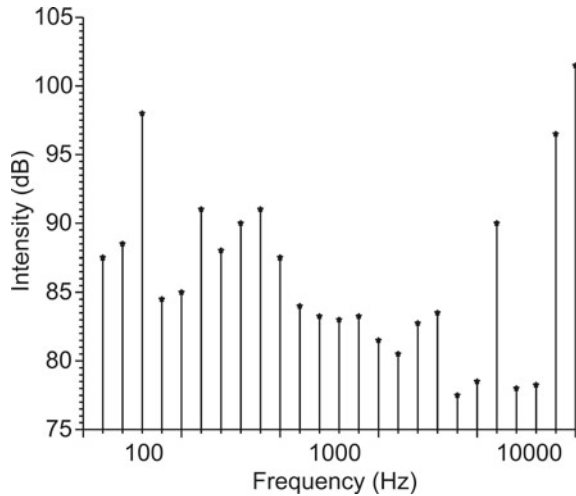
11.6   Sound Intensity Measurement Using a Fast Fourier Transform (FFT) Analyser

DID YOU KNOW that the measurement speed available in today's real-time analysers is the result of a long evolution? The earliest frequency analysers were of the swept spectrum type using sequential bandpass filters. The processing time with these analysers was enormous. To decrease analysis time, a set of parallel filters later became the norm, but these were expensive and heavy. The FFT algorithm proposed by Cooley and Tukey in 1965 revolutionized the way spectrum analysis was done. The computation time for the spectra using FFT was very small compared to the settling time of the narrowband filters, and this resulted in a great speed advantage over the traditional sweep method. The FFT crashed from hours to seconds the time required for a digital computer to calculate the frequency spectrum of a signal. Besides, the FFT algorithm, along with high-speed digital circuits and digital filters, changed spectrum analysis from scalar to vector. Over time, new refinements have made the FFT analyser ever faster, enabling computation of frequency spectra almost in real time with as many as ten spectra per second. This enables the user to watch the dynamic spectrum change even as changes occur in the system being analysed.

Introduction: Another approach to the measurement of sound intensity is through the use of an FFT analyser. In this method, the cross-spectrum between the two closely spaced microphones is obtained with the help of a dual-channel FFT analyser and the intensity calculated from it.

Aim of the Experiment: To analyse the sound intensity of a source using an FFT analyser.

Theory: The theory behind this is as follows. In Eq. (11.8), p and u are in general complex quantities and are not in phase. The Fourier transform of Eq. (11.9) is

$$P = \frac{P_1 + P_2}{2} \tag{11.14}$$

The transform of Eq. (11.12) is

$$U = \frac{1}{\omega \rho_0 \Delta r}(P_2 - P_1) \tag{11.15}$$

The expression for intensity in the frequency domain is

$$I(\omega) = \frac{1}{2}\text{Re}\{P(\omega)U^*(\omega)\} = \frac{1}{2}\text{Re}\left\{\left(\frac{P_1 + P_2}{2}\right)\left[\frac{i}{\omega \rho_0 \Delta r}(P_2 - P_1)\right]^*\right\} \tag{11.16}$$

This can be rearranged as

$$I(\omega) = \frac{1}{2}\text{Re}\left\{\frac{-i}{2\omega \rho_0 \Delta r}(P_2 P_2^* + P_1 P_2^* - P_2 P_1^* - P_1 P_1^*)\right\} \tag{11.17}$$

$P_1 P_1^*$ and $P_2 P_2^*$ are auto-spectra which are real quantities and which do not contribute to the real part of the total expression or the active sound energy. Therefore,

$$I(\omega) = \frac{1}{2}\text{Re}\left\{\frac{-i}{2\omega \rho_0 \Delta r}(P_1 P_2^* - P_2 P_1^*)\right\} \tag{11.18}$$

But

$$P_1 P_2^* = (P_2 P_1^*)^* \tag{11.19}$$

Hence,

$$P_1 P_2^* - P_2 P_1^* = 2i\,\text{Im}\{P_1 P_2^*\} \tag{11.20}$$

Plugging Eq. (11.20) into Eq. (11.18) gives

$$I(\omega) = \frac{1}{2}\text{Re}\left\{\frac{-i}{2\omega \rho_0 \Delta r} \cdot 2i\,\text{Im}\{P_1 P_2^*\}\right\} = \frac{\text{Im}\{P_1 P_2^*\}}{2\omega \rho_0 \Delta r} = \frac{\text{Im}\{P_{12}\}}{\omega \rho_0 \Delta r} \tag{11.21}$$

Here, $I(\omega)$ is the acoustic intensity component in the direction r. $\text{Im}\{P_{12}\}$ is the imaginary part of the cross-spectrum between the two microphone signals and can be directly measured using a dual-channel FFT analyser. Since cross-spectral density P_{12} is needed for the solution, a phase calibration becomes necessary. For this,

Table 11.7 Sound intensity measurement

Sl. No.	1/1 octave centre frequency (Hz)	Im $\{P_{12}\}$	Intensity I (W/m)^2	L_I (dB)
1.	31.5			
2.	63			
3.	125			
4.	250			
5.	500			
6.	1k			
7.	2k			
8.	4k			
9.	8k			
10.	16k			

the microphones are flush-mounted on a rigid circular plate attached to the end of the tube. Ideally, in this configuration, both microphones would measure the same pressure amplitude with zero phase shift.

Experimental Setup and Procedure The test procedure is the same as that outlined in Sect. 11.5. The outputs of the two microphones of the sound intensity probe are fed to a real-time analyser or spectrum analyser, and the imaginary part of the cross-spectrum is obtained. From this, the sound intensity in different bands is computed. The intensity at a particular location may be tabulated as shown in Table 11.7. As before the intensity may be mapped using a sweep method or by measuring at discrete points.

Questions

(1) Why is sound intensity used for noise source identification?
(2) How does sound intensity vary with distance from the source?
(3) If there is a phase mismatch between the microphones, how will the results get affected?
(4) What are auto- and cross-spectra?
(5) Why do auto-spectra disappear from the expression for sound intensity?

11.7   Sound Absorption Measurement Using a Standing Wave Tube

DID YOU KNOW that Kundt's tube was invented by German physicist August Kundt in 1866? This is a very simple apparatus comprising a transparent horizontal tube for the measurement of the speed of sound in air. Kundt caused a metal rod resonator to vibrate or 'ring' by rubbing it to produce longitudinal sound waves in the rod and in an air column. The tube contains a small amount of a fine powder such as cork dust or talc. The other end of the tube is fitted with a moveable piston which can be used to adjust the length of the tube. The standing waves created can be seen in the tube using cork powder. The distance between the nodes is measured to determine the wavelength λ. Present-day setups use a loudspeaker as the sound source, and this is fed with a sine wave from a function generator. The sound generator is turned on, and the piston is adjusted until the sound from the tube suddenly gets much louder indicating resonance. At this instant, the sound waves in the tube are in the form of standing waves. The powder moves with the moving air and remains motionless at nodal locations because the amplitude of vibrations of air is zero there. The distance between the nodal locations is one half wavelength or $\lambda/2$ of the sound waves. By measuring the distance between the nodes, the wavelength λ can be found. Knowing the frequency f of the sound, the speed of sound c_0 in the air can be found by multiplying f by the wavelength.

Introduction: Sound-absorbing materials play a very important role in architectural acoustics, the design of sound recording and broadcast studios, automobile interiors, cinema theatres, factories, workshops, firing ranges, etc. They are used to reduce reflected sound or reverberations and hence the steady-state noise levels in an enclosure, reduce or control reverberation, and improve listening environments for speech and music. Materials with high absorption coefficients are usually soft and porous, and sound absorption of a material is related to its physical properties, such as airflow resistance, porosity, elasticity, and density. Examples of common sound-absorbing materials are acoustic ceiling tiles and soft furnishings and curtains. With the growing focus on noise control issues, acoustic material testing is becoming increasingly important. The absorbing properties of the materials covering the surfaces of a room are required in the computation of its RT.

When sound waves strike absorbing surfaces, air flows in and out of the holes. Frictional forces convert the sound energy into heat, though the actual amount of this energy is small. When a sound wave impinges on a surface, part of the energy is absorbed and part of it is reflected. The sound-absorbing characteristic of a material is defined in terms of its absorption coefficient, which is the ratio of the energy absorbed by the surface to the energy incident on it. When all the incident sound energy is reflected, the value is zero, and when all the energy is absorbed, it is one. A material like foam insulation has an absorption coefficient close to 1, while marble which is

a very good reflector has an absorption coefficient very close to 0. The absorption coefficient also varies with the frequency and angle of incidence of the sound wave. Normal incidence sound absorption coefficients can be very useful in situations where the material is placed within a small acoustical cavity close to a sound source, such as a closely fitted machine enclosure. Generally, higher frequencies are more easily absorbed than the low frequencies found in human speech. Therefore, when specifying the absorbing qualities of a material, manufacturers specify it as a function of frequency in 1/1 or 1/3 octave bands. The selection of a sound-absorbing material has to be based on its absorbing characteristics for a particular noise source. Materials that are good absorbers permit sound to pass through them relatively easily. When sound waves are reflected from the surfaces of a room, they interact with the sound-absorbing materials and lose some energy. However, they require a large thickness or multiple paths for the sound energy to be significantly reduced. This is why sound absorbers are generally not good sound barriers; the latter are essential materials that prevent the passage of sound and are usually solid, fairly heavy, and non-porous. The absorption coefficients of different materials used for the construction of buildings are shown in Table 11.8. Absorption coefficients can be used to predict or design a space with a desired reverberation response.

There are two methods for measuring sound absorption: one involves the use of an impedance tube and is suitable for small samples, while the other involves the use of a reverberation chamber and is suitable for large objects, furniture, panels, etc. The first method is described in this experiment.

Aim of the Experiment: To measure the sound absorption coefficients and acoustic impedances of the given samples.

Theory: The simplest way to measure the absorption coefficients and acoustic impedances of samples is by using a Standing Wave Apparatus. This apparatus was developed essentially to measure the sound-absorbing properties of building materials and absorptive coatings for submarines (to protect them from sonar detection) during World War II. The standing wave tube method allows one to make quick and easy, yet perfectly reproducible, measurements of absorption coefficients and is based on measurement of the Standing Wave Ratio (SWR). It also allows for accurate measurement of the normally incident acoustic impedance and requires only small samples of the absorbing material. The small size and durable aluminium construction of commercially available tubes makes them easy to transport and they can be used for estimating the properties of walls, ceilings, installed building materials, road surfaces, and different ground surfaces or interiors of vehicles.

Figure 11.13 shows the pressure amplitude in a standing wave tube. Figure 11.13a shows the pressure amplitude in the tube with a rigid termination at $x = L$. All the sound energy incident on the termination is reflected with the same amplitude. The small reduction in amplitude shown is due to absorption along the walls as the waves travel back and forth along the tube. Figure 11.13b shows the standing wave when the tube is terminated at $x = L$ with some sound-absorbing material. Here, it is seen that some of the incident acoustic energy is absorbed by the material, resulting in a

Table 11.8 Absorption coefficients of construction materials

Material	1/1 octave band freq. (Hz)					
	125	250	500	1000	2000	4000
Bricks, unglazed	0.01	0.01	0.01	0.01	0.02	0.02
Marble	0.03	0.03	0.03	0.03	0.03	0.03
Heavy carpet on felt	0.08	0.27	0.39	0.34	0.48	0.63
Asphalt tile	0.02	0.03	0.03	0.03	0.03	0.02
Bricks, unglazed, painted	0.01	0.01	0.02	0.02	0.02	0.03
Concrete block	0.36	0.44	0.31	0.29	0.39	0.25
Concrete block, painted	0.10	0.05	0.06	0.07	0.09	0.08
Concrete	0.01	0.05	0.06	0.07	0.09	0.08
Wood	0.15	0.11	0.10	0.07	0.06	0.07
Glass, Ordinary Window	0.35	0.25	0.18	0.12	0.07	0.07
Plaster	0.013	0.015	0.02	0.03	0.04	0.05
Plywood	0.28	0.22	0.17	0.09	0.10	0.11
Fibre glass	0.30	0.34	0.50	0.80	0.93	0.96
Plate glass	0.18	0.06	0.04	0.03	0.02	0.02
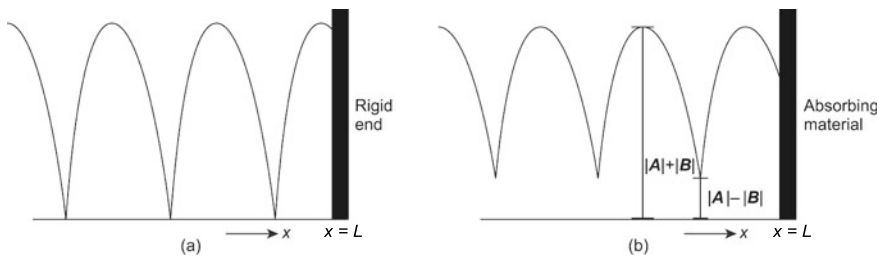
Fig. 11.13 Standing waves in Kundt's tube: **a** with rigid termination, **b** with sound-absorbing material

decrease in the amplitude of the reflected waves. Besides, the impedance mismatch at the absorbing material introduces a phase shift upon reflection.

From the plane wave assumption, the sound field inside the waveguide is described by

$$p = p_i + p_r = A e^{i(\omega t - kx)} + B e^{i(\omega t + kx)} \tag{11.22}$$

where the first term represents the incident wave and the second term the reflected wave, $k = \omega/c_0$ is the wavenumber, A is the complex pressure amplitude of the incident wave, B that of the reflected wave, and ω is the angular frequency.

The maximum pressure amplitude at an antinode is $|A| + |B|$, while the minimum amplitude at a node is $|A| - |B|$. It is not possible to measure A or B directly. However, we can measure $|A| + |B|$ and $|A| - |B|$ using the traversing microphone. The ratio of the pressure maximum (antinode) to the pressure minimum (node) is called the SWR S.

$$S = \frac{|A| + |B|}{|A| - |B|} \tag{11.23}$$

This ratio, which always has a value greater than or equal to unity, is used to determine α_r, the sample's reflection coefficient, α_a, its absorption coefficient, and z, its impedance. Sound power reflection coefficient is defined as

$$\alpha_r = \left| \frac{B}{A} \right|^2 = \frac{(S-1)^2}{(S+1)^2} \tag{11.24}$$

The sound power absorption coefficient for the test sample is

$$\alpha_a = 1 - \alpha_r = 1 - \frac{(S-1)^2}{(S+1)^2} \tag{11.25}$$

A pressure minimum occurs when

$$\cos\left(kx + \frac{\theta}{2}\right) = 0 \quad \text{and} \quad \sin\left(kx + \frac{\theta}{2}\right) = 1 \tag{11.26}$$

which requires that

$$kx + \frac{\theta}{2} = -\frac{\pi}{2} \tag{11.27a}$$

or

$$\theta = -2kx - \pi \tag{11.27b}$$

Here, the quantity x equals the distance from the test sample to the first pressure minimum and θ is the phase lag of B with respect to A.
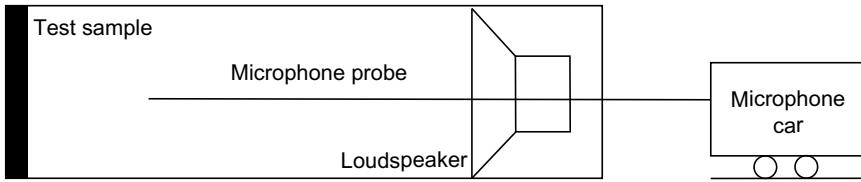
Fig. 11.14   Standing wave tube

The complex mechanical impedance of the test sample is then obtained as

$$z_n = z \frac{A + B}{A - B} \qquad (11.28)$$

z is the characteristic impedance of the medium and z_n is the specific acoustic impedance of the test specimen. Since the mechanical impedance of the test sample is generally a function of frequency, it may be required to repeat the above measurements over the frequency range of interest in various 1/1 and 1/3 octave bands. At low frequencies, the thickness of the material is important since absorption increases with thickness. Low-frequency absorption can be further increased by fixing the material at a distance of one-quarter wavelength from a wall, instead of directly on it.

Test Procedure for Measurement of Reflection and Absorption Coefficients and Impedance The standing wave tube or Kundt's tube is shown in Fig. 11.14. At one end of the tube is a loudspeaker which produces an acoustic wave that travels down the tube, impinges with normal incidence on the test sample placed at the other end, and gets reflected back. The amplitude of the reflected wave depends on the incident wave and the characteristic impedance of the test sample. Since this method is based on plane wave propagation, there is a maximum frequency that a tube can support and this is $f_{\max} = c_0/1.7D$, where D is in metres.

Both the forward and the backward travelling waves combine and there will be a resulting phase interference between the waves, leading to the formation of a standing wave pattern in the tube. At some particular points in space called points of destructive interference (nodes), the two waves will always be out of phase; the result is that they will cancel each other at these points. At points of constructive interference (antinodes), the two waves will always be in phase. If 100% of the incident wave is reflected, then the incident and reflected waves have the same amplitude, resulting in the nodes in the tube having zero pressure and the antinodes having double the pressure. If some of the incident sound energy is absorbed by the sample, then the incident and reflected waves have different pressure amplitudes and the nodes in the tube no longer have zero pressure. When there is a standing wave in an impedance tube, if a microphone is moved along a line between the sound source and the reflecting surface, a series of maxima and minima in sound levels is found. The pressure ampli-
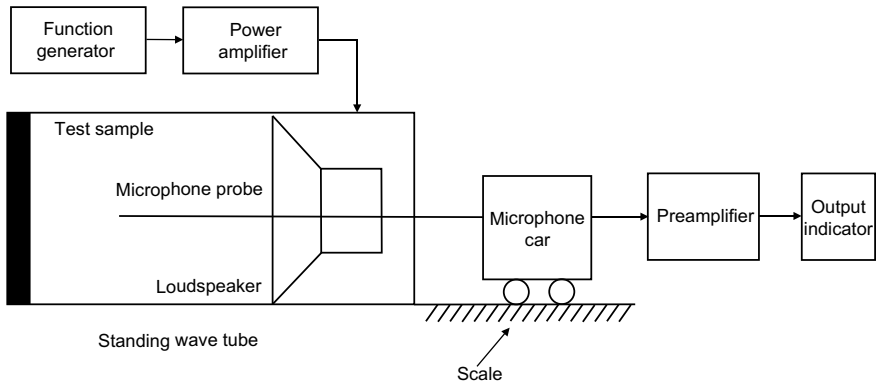
Fig. 11.15 Measurement setup for standing wave apparatus

Table 11.9 Measurements using standing wave apparatus

Sl. No.	1/1 octave band centre frequency (Hz)	$|A| + |B|$ (V)	$|A| - |B|$ (V)	S	α_r	α_a	x (m)	θ (rad)	z_n kgm^{-2} s^{-1} or MKS rayl
1.	31.5								
2.	63								
3.	125								
4.	250								
5.	500								
6.	1k								
7.	2k								
8.	4k								
9.	8k								
10.	16k								

tudes at nodes and antinodes are measured with a travelling microphone attached to a car which slides along a graduated ruler.

The test setup is as shown in Fig. 11.15. A typical measurement may be tabulated as shown in Table 11.9. The measurements may be repeated with the same material of different thicknesses and also with different materials.

Questions

(1) How does the sound absorption coefficient change with the thickness of the material?
(2) Which is the main source of error in the measurement?

(3) How does the sound absorption coefficient change with frequency?

(4) How can you measure the velocity of the sound wave?

(5) Repeat the experiment with a transparent tube with fine sawdust and observe the nodal locations.

11.8   Acoustic Impedance Measurement Using the Two-Microphone Method (The Transfer Function Method)

INTERESTING FACTS: Acoustic impedance (z) can be thought of as the resistance offered by a medium to longitudinal wave motion. This complex quantity is the ratio of the sound pressure (p) acting and the resulting particle velocity (u) and is $z = p/u$. This impedance is also called the specific acoustic impedance of the medium since it characterizes the medium itself. Acoustic impedance is analogous to electrical impedance where variations in electrical potential (or in voltage v) give rise to a moving charge (or electrical current i) and the electrical impedance $z = v/i$. The acoustic impedance is an important parameter that decides how much acoustic energy is transferred between two media. When a sound wave is incident on a boundary, part of the acoustic energy is reflected back and the remaining part gets transmitted into the other medium. There is no reflection (hence total transmission) if both media have nearly the same acoustic impedance. On the other hand, when there is an impedance mismatch, most of the energy is reflected and the power transmission is small.

Introduction: There are two standards for the measurement of acoustical properties in an impedance tube: ASTM E1050-98 and BS EN ISO 10534-2:2001. Both describe well-defined acoustical conditions and special instrumentation to ensure a high and known degree of repeatability and reproducibility. Both methods described are known as the 'two-microphone' or 'transfer function' method of measuring absorption and impedance of acoustical materials. This test method allows one to compare relative values of sound absorption when it is not possible to procure large samples for accurate random-incidence measurements in a reverberation room. This method is designed for the measurement of the following quantities: (i) absorption coefficient, (ii) reflection coefficient, (iii) complex acoustic impedance, (iv) complex acoustic admittance, and (v) transmission loss (TL). This technique can be used to evaluate acoustic properties very rapidly since no traversing microphone is necessary. It is also a much faster measurement technique compared to the measurement of sound absorption in a reverberation room according to the method specified in ISO 354:2003.
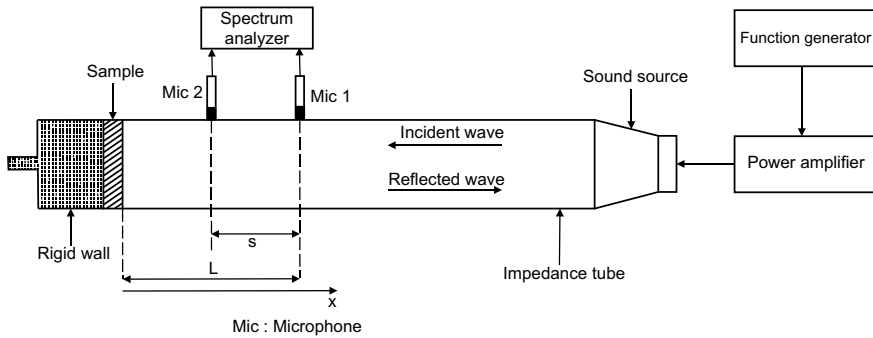
Fig. 11.16 Schematic of two-microphone method of sound absorption

Aim of the Experiment: To measure the acoustic impedance of the given samples using the two-microphone method or the transfer function method.

Theory: In both the ASTM E1050-98 and ISO 10534-2:2001 measurement techniques, the microphone closer to the source is taken as reference channel 1. Theory has been developed relating the auto- and cross-spectral densities of the incident and reflected waves to the auto- and cross-spectral densities of the two microphone signals. The origin of the coordinate system is at the termination of the tube (with the sample) and x_1 and x_2 are the distances to the two pressure measurement points. The two-microphone method is shown schematically in Fig. 11.16.

An expression for the reflection coefficient is obtained by measuring pressure at two different axial locations. From the transfer function H_{12}, the complex reflection coefficient α_r of the material is determined as

$$\alpha_r = \frac{H_{12} - e^{iks}}{e^{iks} - H_{12}} e^{i2k(L+s)} \tag{11.29a}$$

In the above expressions, H_{12} is the estimated frequency response function (FRF) between the two microphones and is given by

$$H_{12} = E\left[\frac{S_{12}}{S_{11}}\right] \tag{11.29b}$$

where $E[\ ]$ is the expectation operator, S_{12} is the estimated cross-spectral density, and S_{11} is the estimated auto-spectral density. L is the distance from the sample face to the first microphone in ASTM E1050-98 and s is the distance between the microphones.

$$k = 2\pi f/c_0 \tag{11.29c}$$

From the complex reflection coefficient, the absorption coefficient α_a is determined using Eq. (11.30a) and normal specific acoustic impedance z_n of the sample may be

determined from Eq. (11.30b).

$$\alpha_a = 1 - |\boldsymbol{\alpha_r}|^2 \tag{11.30a}$$

$$\boldsymbol{z_n} = \rho_0 c_0 (1 + \boldsymbol{\alpha_r})/(1 - \boldsymbol{\alpha_r}) \tag{11.30b}$$

A quick absolute test for calibrating this setup is done by measuring the imaginary part of impedance of a closed tube of length L_0. The theoretical normalized impedance $z/\rho_0 c_0$ of a closed tube of length L_0 for calibration is given by

$$\mathrm{Im}(z/\rho_0 c) = \frac{1}{\tan k L_0} \tag{11.31}$$

Test Setup and Procedure: This method uses an impedance tube, two microphones, and a digital frequency analyser for the determination of the sound absorption coefficient of sound absorbers for normal sound incidence, that is, at $0°$, typically in the frequency range 150–6000 Hz. The test specimens are small circularly cut samples of the sound-absorbing materials. A loudspeaker or a high-output acoustic driver (sound source) is mounted at one end of an impedance tube and a small sample of the material to be tested is placed in a sample holder at the other end. A rigid plunger with an adjustable depth is placed behind the sample to provide a reflecting surface. The sound source is typically made to generate broadband, stationary random sound waves which propagate as plane waves in the tube, impinge on the sample normally, and are reflected back, resulting in a standing wave interference pattern inside the tube.

The sound pressure is measured by a pair of identical microphones mounted flush with the inner wall of the tube near the sample end, and the complex transfer function between the two measurements is calculated using a multichannel spectrum analyser. This method separates the incident and reflected energy from the measured transfer function and then estimates the required acoustic properties of the test sample. The useful frequency range depends on the diameter of the tube and the spacing between the microphone positions. A 100 mm diameter tube is used to cover measurements over the frequency range 150–1600 Hz, and a 30 mm diameter tube to cover measurements in the frequency range 1200–6000 Hz. Tubes of two different diameters may be used to cover the full frequency range. Figure 11.17 shows a photograph of the impedance tube. The measured values and results are tabulated as shown in Table 11.10.

Precautions to be Taken During Measurement

(i) *Construction of the tube:* The tube is made preferably of a material with high damping such as brass, which has damping several times that of materials such as aluminium and density also three times as much. It is required that the tube be sufficiently massive and rigid to avoid transmission of noise into the tube

Fig. 11.17 Photograph of impedance tube. (Courtesy of https://www.bswa-tech.com)

Table 11.10 Measurements and calculation of impedance

Sl. No.	f (Hz)	H_{12}	R	α_a	z_n
1.					
2.					
3.					
4.					

from outside and to prevent vibration excitation of the tube by the sound source or by background sources. The tube must be long enough to present a stable plane wave sound field to the sample under test. The standards recommend a tube length of at least three diameters, but a length of at least 10–15 diameters is preferred. Typically, the upper working frequency f_u is limited by $d < 0.58\lambda_u$ (for a circular tube) or $d < 0.50\lambda_u$ (for a rectangular tube) where d is the inside diameter of the circular tube (m) or side of the rectangular tube. A wall thickness that is 10% of the tube diameter is preferred. The lower limit is determined by the speaker's lower frequency limit.

(ii) *Microphone mounting and frequency range:* The microphones must be mounted flush with the inside wall of the tube and isolated from the tube (so that they are not affected by vibration). Both the microphones should be of laboratory grade and of the same type and size. Their diameter should be small in comparison with their spacing to reduce the influence of their acoustic centres, i.e., $d_{mic} < 0.2s$ where d_{mic} is the diameter of the microphones and s the spacing (m) between them. Their diaphragm diameter should be small to minimize high-frequency spatial averaging across the face of the diaphragm: $d_{mic} \ll \lambda_{max}$ where $\lambda_{max} = c_0/f_{max}$ corresponds to the wavelength of the highest frequency sound. The upper frequency limit is chosen to ensure the occurrence of plane wave mode propagation and accurate phase detection; typically, $\frac{1}{2}$ or $\frac{1}{4}$ inch microphones are used. The spacing s between microphones is chosen such that $s < (0.40 \text{ to } 0.45)\lambda_{max}$ where λ_{max} is the maximum side length of the rectangular tube (m) and f_{max} is the highest frequency (Hz). The spacing between the sample and the closest microphone

must be large enough to avoid proximity distortions to the acoustic field; for a sample with a strongly asymmetrical layer, $\Delta r > 2d$ is preferred.

(iii) *Signal-to-noise ratio (SNR):* The standards recommend that the background noise inside the tube be lower than the sound field, by at least 10 dB, but preferably by 20–30 dB. For this, the TL of the tube should be greater than 45 dB. This can be achieved by ensuring that the tube is constructed of heavy materials and is properly sealed at the microphone ports, sample holder, sound source, etc. The minimum SNR occurs at the minima of the standing waves which can be as much as 25 dB below the maxima. Besides, it is often required to measure acoustic absorption at relatively low levels to match field conditions (e.g., approximately 65 dB to simulate conditions in offices or vehicles). This poses a requirement for a low background noise level in the tube.

(iv) *Sample holder:* The sample holder should also be massive and rigid and should have a massive back plate of length at least 20 mm behind the sample. It should also have a smooth, machined surface for mating with the tube to ensure that no sound leaks into or out of the tube. The back plate or plunger should be adjustable so that materials of different thicknesses may be tested and to allow for an air space behind the sample, if necessary.

(v) *Sound source:* The sound source should have the required frequency range and a high power rating (e.g., 50–100 W) so that high-intensity sounds may be generated inside the tube for certain types of testing. For certain materials, the absorption depends on the intensity of the sound source and such materials should be tested at several levels above and below field conditions. The sound source should also have a frequency response that is fairly flat (within ± 10 dB).

(vi) *Mounting the sample:* The impedance tube method requires samples of the test object which are of the same size as the cross section of the tube. According to standards, the sample must fit snugly in the specimen holder so that there is no space between its edge and the holder. The test specimen is generally sealed around the edges using Vaseline or layers of tape to minimize leakage of sound. The sample should also be in contact with the rigid plunger. In applications where the acoustic material is used with an air gap, the test should duplicate field conditions by maintaining the same air gap between the sample and the plunger.

(vii) *Temperature correction:* The temperature should be held constant during the test to within 1 °C.

Questions

(1) What are the main sources of experimental error?
(2) How will you decide the frequency resolution for measurements?
(3) Will averaged spectra lead to better results than instantaneous spectra?
(4) What will be the effect of errors in measurements in length on the final results?
(5) How does the distance between microphones affect the results?

11.9 Absorption Coefficient Measurement Using a Reverberation Chamber

DID YOU KNOW that reverberation chambers cause multiple reflections of sound waves to produce a non-directional or diffuse sound field within the chamber? They are typically used in a variety of standardized measurements, such as in measuring the absorption of materials, the sound power of noise sources, for microphone calibration, and for measuring the transmission loss of partitions. They are also used for testing satellite structures at high sound pressure levels. When designing a space, such as a concert hall, the reverberation time of the space should be kept optimal so that music sounds good and speech is intelligible. Hence, it is required to know the absorption characteristics of architectural materials to be used in a room for good acoustics. Reverberation chambers are used to measure the sound absorption characteristics of absorbing materials or specimens such as soft panels, screens, or pieces of furniture (e.g., theatre seats, chairs, and sofas).

Introduction: An alternate method (to standing wave measurements) for determining the absorbing properties of a material, also in wide use today, involves the use of a special reverberation chamber. The reverberation room method ideally determines the sound absorption coefficient for diffuse sound incidence. This method requires a large (not less than $125\,\text{m}^3$, preferably $> 200\,\text{m}^3$) and carefully qualified reverberation chamber and test specimens which are rather large (more than $5.4\,\text{m}^2$, and with an aspect ratio not too high). Hence, this technique is not convenient for research and development work, where only a few samples of the absorber are available. This method is generally more expensive, since it requires a reverberation chamber and precisely calibrated sensors and is more cumbersome than the standing wave method. Besides, it does not yield normally incident acoustic impedance data. However, it is superior for determination of absorbing properties that depend on the size of the material and for measuring absorption characteristics due to randomly incident sound waves since the standing wave tube measures the absorption coefficient only for normal incidence and the measured coefficients are generally a little more than those measured using the reverberation room method. The sound absorption measurement in a reverberation chamber may be carried out according to ASTM C-423 or BS EN ISO 354:2003 for testing plane absorbers, i.e., flat areas of sound absorptive material such as carpeting or acoustic tiling, or for testing discrete sound absorbers such as chairs, free-standing screens, persons, pads, or baffles.

Aim of the Experiment: To measure the absorption coefficient of the given samples in a reverberation chamber.

Theory: The sound source is initially kept on and then suddenly turned off, causing the sound pressure level to decrease. The RT is then measured as the time taken for

the SPL in the room to decay by 60 dB after the sound source has been switched off. The equivalent sound absorption area of the test specimen, a, is computed using Sabine's equation (Eq. 4.161); this is calculated as the difference in Sabine with and without the object under test present in the reverberation chamber. Because of the sound absorption, the RT is shorter with the test specimens present. The expression for RT is

$$T = 0.161 \frac{V}{S\bar{\alpha}} \tag{11.32a}$$

where $\bar{\alpha}$ is the average absorption coefficient of all absorbing surfaces S_1, S_2, S_3, ..., S_n with sound absorption coefficients $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n$.

$$\bar{\alpha} = \frac{S_1\alpha_1 + S_2\alpha_2 + \cdots + S_n\alpha_n}{S_1 + S_2 + \cdots + S_n} = \frac{S_1\alpha_1 + S_2\alpha_2 + \cdots + S_n\alpha_n}{S} \tag{11.32b}$$

The Sabine absorption equation is obtained as

$$a_{ij} = \frac{0.9210 V d_j}{c_0} \tag{11.33}$$

where a_{ij} is the Sabine absorption of the reverberation chamber under the ith measurement condition in the jth frequency band (in metric Sabines), V is the chamber volume in m^3, d_j is the decay rate in dB per second for the jth frequency band, and c_0 is the speed of sound.

Test Setup and Procedure: In this method, the reverberant sound field within the chamber should be kept as diffuse as possible during the steady-state and decay conditions since computation is based on this assumption. Hence, effort should be made to qualify a chamber for sound absorption measurements according to international acoustic standards. The source is kept on for a sufficiently long time to ensure that the sound pressure level in the room has reached steady-state conditions. The source is then suddenly turned off, causing the sound pressure level to decrease. The RT is initially measured in the empty room, later with the test item/items placed in the room. This measurement may be made using a sound level meter having the facility. The item is typically in the form of a large piece of the sample (say 10 m^2) and is placed in the chamber. The difference in the RT with and without the material yields the absorbing properties of the material. A band of random noise generated by a sound source and amplified by a power amplifier is generally used as the test signal since this enables measurements over the entire frequency range of interest in one pass; a harmonic signal may also be used. Figure 11.18 shows the typical test setup used for measurement.

At least one loudspeaker position and three (preferably five) microphone positions with two readings in each case should be used, making the measured RT the average value of six (or 10) decay measurements in each frequency band to get space- and time-averaged decay rates. This process is done for each octave or third octave band from 400 to 10,000 Hz. For the computation, the absorption coefficient of the
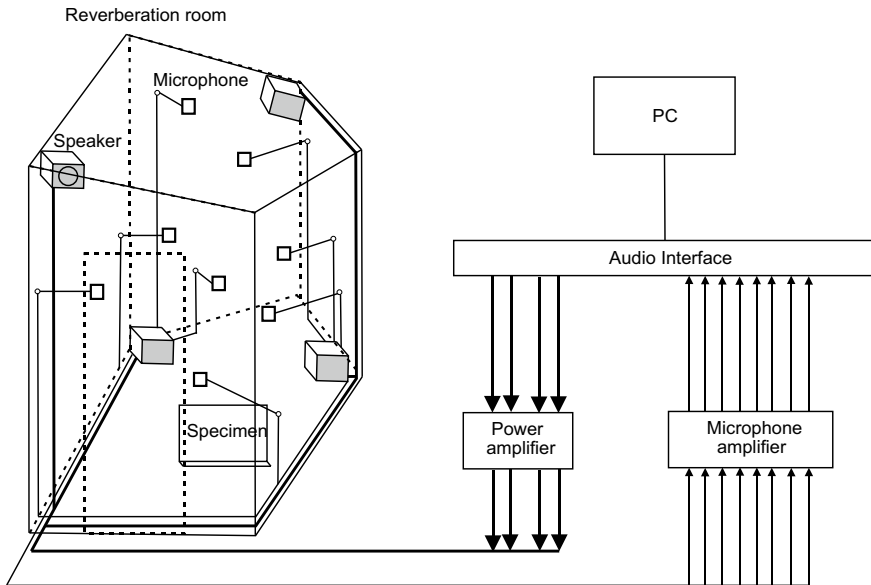
Fig. 11.18 Test setup for measurement of absorption coefficient

materials of the room surface should be considered if they are significant; however, values for smooth, hard, and rigid surfaces such as reverberation room floor may be neglected. Readings may be tabulated as shown in Table 11.11.

Typical Difficulties Involved in this Method There may be spatial variations in the decay rates in the reverberation chamber due to the formation of standing waves. Such variations are significant at low frequencies and for test specimens with high sound absorption coefficients. Variations may be reduced by adding stationary and/or rotating diffusers suspended from the ceiling. Multiple microphone positions may be used, as also multiple sample positions. Standardized sample locations are defined in ASTM E795 and ISO 354, Amendment 1. Changes in temperature and humidity during the test, when the sample is taken into or out of the chamber, give rise to errors, especially at high frequencies. Insufficient amplifier and loudspeaker power or intrusive ambient noise create problems. 'Non-linear' decays, that is, the decay rate not being constant, but varying with time, cause errors, especially at low frequencies, in small rooms, and when the sound absorption coefficient of the test specimen is high. Specimens with too small area or too high aspect ratio, as well as reverberation chambers with improper dimensions, are other pitfalls.

For some test specimens, sound absorption coefficients greater than 1.00 are obtained. This happens in cases where the absorption footprint is larger than the area of the specimen and wave diffraction happens at the edges of the specimen. The smaller the frequency and specimen size and the larger the aspect ratio and the sound absorption, the larger is the diffraction effect.

Table 11.11 Measurements of absorption of test specimen

i th measurement	jth frequency band (Hz)	RT without specimen (s)	d_j without specimen (dB/s)	a_{ij} without specimen (metric Sabines)	RT with specimen (s)	d_j with specimen (dB/s)	a_{ij} with specimen (metric Sabines)	a of test specimen (metric Sabines)
1.	500							
1.	500							
2.	500							
2.	500							
3.	500							
3.	500							
1.	630							
1.	630							
2.	630							
2.	630							
3.	630							
3.	630							
1.	708							
1.	708							
. . .	. . .							

Absorption Coefficient of Plane Absorbers These are generally placed directly against a surface of the reverberation chamber, usually on the floor. Due to the non-parallel walls of the chamber, a sample placed on the floor will affect the decay time of all of the modes. The edges of the test specimen are sealed with an acoustically reflective frame made of steel, gypsum board, or wood to prevent the edges from absorbing sound. For a plane absorber or a specified array of test objects, the acoustic absorption coefficient is obtained by dividing absorption a by the treated surface area S. For plane absorbers, the sound absorption coefficient is reported as

$$\alpha_j = (a_j/S) + \alpha_{1j} \text{ for ASTM C423}$$

$$\alpha_j = a_j/S \text{ for ISO 354} \tag{11.34}$$

where the quantity α_{1j} is the estimated sound absorption coefficient of the surface covered by the plane absorber during the test.

Discrete Sound Absorbers Discrete objects such as chairs, screens, or persons are installed in the same manner as they are installed in practice. At least three individual test specimens should be used to provide a measurable change in the equivalent sound

absorption area of the room. These individual objects are arranged randomly with a spacing of at least 2 m from each other and at least 1 m from any room edge. The sound absorption of the specimen is

$$a_j = a_{2j} - a_{1j} \tag{11.35}$$

where a_{2j} is the Sabine absorption of the reverberation chamber with the specimen present in the jth frequency band (in metric Sabines) and a_{1j} is the Sabine absorption of the chamber with the chamber 'empty' in the jth frequency band (in metric Sabines).

For arrays of discrete sound absorbers, the absorption a_j of the array or the averaged sound absorption of each element may be reported as a_j/N, where N is the number of elements in the array.

Noise Reduction Coefficient (NRC) A single number rating called the NRC is used to specify the absorbing properties of materials. It is computed as the average of the sound absorption coefficients measured in the 1/1 octave bands centred at 250, 500, 1000, and 2000 Hz and is rounded to the nearest multiple of 0.05 Sabine/m^2. The frequency range of interest may be extended down to 63 Hz octave band in large chambers of concrete construction and sometimes to 5000 Hz on the upper side. Outside this frequency range, measurements are typically performed in an impedance tube as described in Sects. 11.7 or 11.8.

Questions

(1) What is the effect of background noise on the measurement of reverberation time?
(2) What are the main sources of error in measurement?
(3) When is NRC preferred over absorption coefficient as a function of 1/1 or 1/3 octave bands?
(4) What are the frequency constraints associated with reverberation chambers?
(5) How can you do spatial and temporal averaging?
(6) How do you prevent the formation of standing waves in the chamber?

11.10   Measurement of Reverberation Time of a Chamber

DID YOU KNOW that it was in October 1898 that Wallace Sabine published his equation for calculating the reverberation time of a room? Fogg Hall which was within the new art museum of Harvard University had a lot of echoes and the room was unusable. Sabine was asked to investigate the behaviour of sound in that room and thus was born the field of architectural acoustics. After thousands

of painstaking measurements using just a portable pipe organ and a stopwatch, Sabine with the help of some keen-eared assistants discovered that the reverberation time (a term coined by him), i.e., the time taken for the sound to decay by 60 dB was related only to the volume of the room and the absorption of the materials on the room surfaces. The simplifying assumptions that Sabine made were that the sound field in the room is perfectly diffuse, absorption is uniform around the room, and that the reverberation time obtained for 512 Hz (more recently, 500 Hz) tone would be representative of all frequencies in the room. Owing to his pioneering work, he was asked to help with the design of the New Boston Music Hall. This was opened to the public in October 1900 and is still considered to be one of the three finest concert halls in the world. His work has been the starting point for acoustic calculations for over a century.

Introduction: RT is a quantity of great significance in architectural acoustics and is obtained from spatial acoustic analysis for characterizing the acoustic properties of a room or enclosure. RT is used to describe room acoustics by analysing the decay of sound intensity and is defined as the time it takes for the sound pressure level in an enclosure to decay by 60 dB or to one millionth of the original energy after a sound source has been switched off. This time depends on the absorbing properties of the room surfaces; the larger the amount of absorbing material, the smaller is the RT. In practice, all surfaces absorb some amount of sound energy, leading to a finite RT. Values of RT may vary from fractions of a second to a few seconds, depending upon the size of the room and the nature of the materials used in its construction. Highly reflective materials, such as a concrete or tiled floor, brick walls, and windows, will increase the RT as they are very rigid. Absorptive materials, such as curtains, heavy carpet, and people, on the other hand, reduce the RT. In fact, people tend to absorb quite a bit of energy, reducing the RT. Bigger rooms have longer RTs since the sound waves have to travel a longer distance between reflections. The air in the room itself will also attenuate the sound waves, reducing the RT. Table 11.12 gives an indication of the RTs of some commonly used spaces.

While designing an enclosure, it is imperative to identify appropriate values of RT depending on its intended use and then to decide on the required building materials to be used in the construction. In the case of rooms with existing acoustic problems, the measurement of RT is usually the first step in identifying a suitable remedy, and depending on the application, RT may need to be corrected. A small RT is desirable to minimize the masking effects of sound, while a large RT is desirable if a weak source of sound is to be audible everywhere; hence, a compromise between the two extremes is desirable.

Aim of the Experiment: To measure the reverberation time of a known chamber.

Theory and Procedure: The basic theory has been described in Sect. 4.6.3. According to ASTM C423, the decay rate is to be determined as follows:

Table 11.12   Reverberation times of some common enclosed spaces

Enclosure	Reverberation time (s)
Recording studio	0.4–0.5
Living room	0.5
Lecture hall	0.5–1
Opera house	1.1–1.3
Small music hall	1.5
Concert hall	1.5–2.1

(a) Once the noise source is turned off, sound pressure levels in all one-third octave bands of interest are to be sampled every 50 ms or less. An arithmetically averaged decay curve is to be obtained over N or more decay curves, where N should be at least 50.

$$\langle L_m \rangle = \frac{1}{N} \sum_{n=0}^{N} L_{mn} \tag{11.36}$$

where $\langle L_m \rangle$ is the average of the sound pressure levels measured at the mth data point across the N decays and L_{mn} is the sound pressure level measured at the mth data point during the nth decay.

(b) In each band, the following are to be found: the first data point that occurs 100 ms or more after the noise source is turned off and the last data point that is within 25 dB of the earlier value.

(c) In each band, the unadjusted decay rate d is to be calculated; this is the negative of the slope of the linear, first-order regression on the averaged decay curve computed as

$$d' = \frac{6}{M(M^2 - 1)\Delta t} \left[(M + 1) \sum_{m=1}^{M} \langle L_m \rangle - 2 \sum_{m=1}^{M} \langle L_m \rangle \right] \tag{11.36a}$$

where M is the number of data points in the jth band and Δt is the time interval with which the sound pressure level spectra were sampled. The unadjusted decay rate may be optionally adjusted for air absorption as

$$d = d' - d_{air} \tag{11.36b}$$

where the correction is based on coefficients taken from ANSI S1.26 for the appropriate temperature and relative humidity.

Instrumentation and Measurement: To measure the RT, a sound source and instrumentation for capturing the sound decay are required (Fig. 11.19). The source could be a loudspeaker or power oscillator. A diffuse field environment is to be created for the measurement. The experimental procedure involves generating a sound using a

source and capturing the decay after the source is switched off. One primary difficulty encountered with a pure tone sound source is the formation of a pattern of standing waves, especially pronounced at low frequencies. Hence, a method commonly employed to average out the anomalies is to use a warble oscillator, in which by slowly warbling (adopting frequency modulation of) the frequency of the sound emitted by a loudspeaker, continuous changes in the standing wave patterns are produced. There are two other options for the sound source: impulse excitation or noise excitation (ISO 140 series of standards). It is important for the noise to have sufficient bandwidth to cover the entire frequency range of interest. The source is usually positioned in a corner of the room to ensure that it undergoes as many reflections as possible. It is suddenly switched off after steady-state conditions have been reached. RT is measured by collecting response data throughout the space using calibrated microphones. These signals are then digitized using an analogue-to-digital converter and then sent to a personal computer with software for mathematical data analysis. Sampling at a high frequency is necessary to ensure that the signal reconstructed from the samples is an accurate representation of the analogue input signal. Besides, the record length should be long enough to ensure that the decay event is fully captured. The earlier practice was to amplify, rectify, and record the level of the signal by a stylus on a moving strip of paper. The rate of decay in dB/s could be obtained from the slope of the curve, and from this, RT could be calculated.

Although the RT is defined as the time it takes for the sound to decay 60 dB, this is often difficult to measure in practice, the reason being the high existing background noise in many rooms. To be able to measure a 60 dB decay with a background noise level of 55 dB, the acoustical output level of the noise source should typically be more than 120 dB ($55 + 60+$ another 5 dB), which is a requirement very hard to meet with many practical sources. Therefore, it has become a practice to measure 15, 20, 30, or 40 dB decays instead of the complete decay and subsequently to extrapolate them to 60 dB, assuming that the measured decay is representative of the entire one. Hence, the decay used, i.e., T20 or T30 or T40, is specified and a check for consistency is made by comparing T20 and T40 records to ascertain if the decay is linear (when plotted as a graph with a logarithmic level scale). Most instruments currently available for building acoustics have at least two ways of calculating the decay simultaneously. The frequency range of interest is 100–3150 Hz, but many measurements are made in 50–5000 Hz range, this being the requirement for laboratory measurements. RT measurements are made in 1/3 octave bands in this frequency range since measurements made at a single frequency are insufficient to completely characterize the reverberation characteristics of an enclosure due to the frequency dependence of the absorption coefficients of materials.

An ideal sound decay will form a straight line when drawn with logarithmic axes. However, the actual sound decays always contain fluctuations, the influence of which may be reduced by a triangular weighting function applied to the middle part of the decay curve and ignoring the fluctuations at the beginning and end. In order
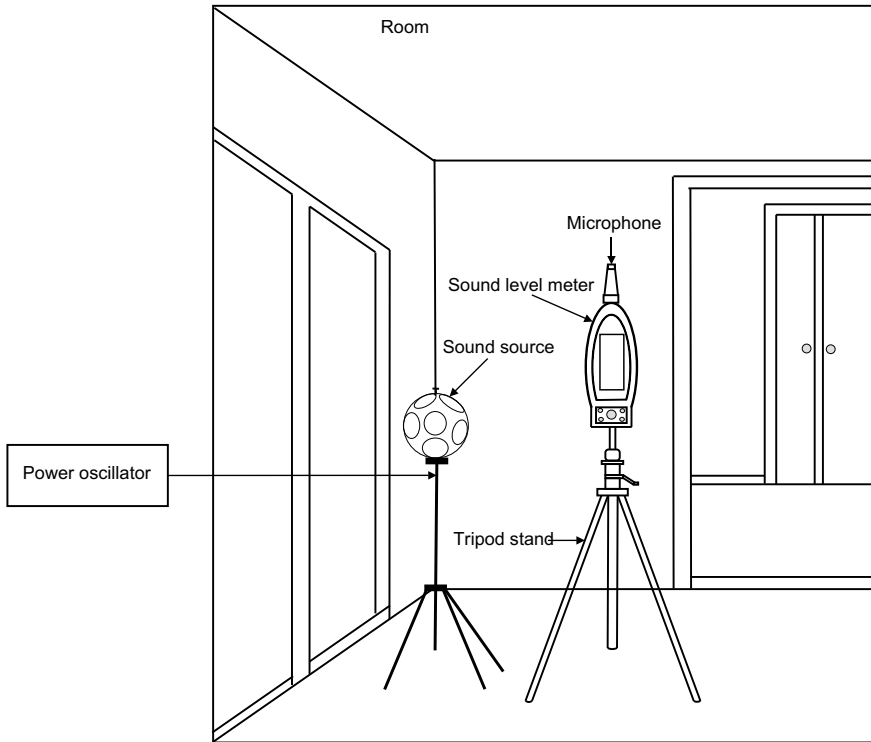
Fig. 11.19 Test setup for measurement of reverberation time

to account for any possible influence from the background noise level, a minimum distance to the noise floor may be specified (ISO 354 requires a minimum of 15 dB distance). Figure 11.20 shows a typical decay curve.

Questions

(1) Do you expect larger errors in measured reverberation time at low frequencies or at high frequencies?
(2) How does the position of the source affect the measured reverberation time?
(3) How does the position of the microphone affect the measured reverberation time?
(4) Why is it beneficial to have the sound source in the corner of a room?
(5) How does the modal density of the room affect measurements?
(6) Why is it that concert halls are expected to have a reverberation time of 1.5 to 2.1 s?
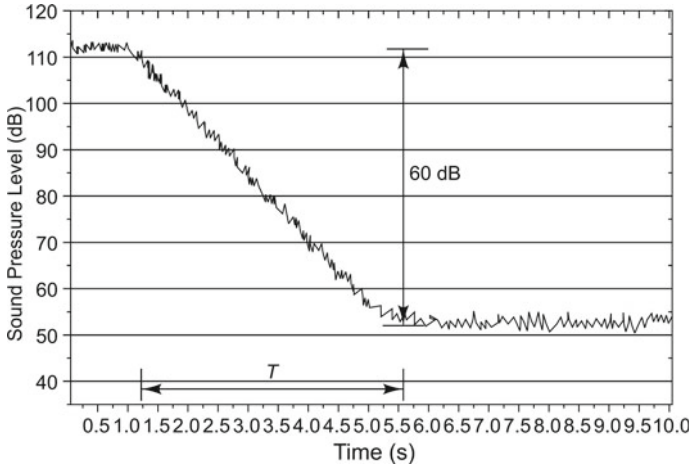
Fig. 11.20 Typical decay curve

11.11 Measurement of Sound Transmission Loss (TL) Using Sound Pressure Level and Reverberation Time: ASTM E-90

> **INTERESTING FACTS:** Sound transmission loss is a measure of the sound attenuation brought about by a partition dividing two spaces. It is not applicable to incomplete partitions such as barriers where sound can diffract around the edges. The transmission loss of cylindrical shells has long been a subject of research. This has a lot of applications in the aerospace and automotive industry. Such studies are useful in the context of, say, the cabin of a helicopter, where passengers are in close proximity to the excitation sources that contribute to interior noise such as main and tail rotors, engines, main gearbox, and aerodynamic turbulence. Transmission loss studies are also required in building acoustics.

Introduction: Acoustic barriers are used as acoustic insulators to block or prevent the passage of sound. Common barrier materials are gypsum board, plywood, concrete, etc. They are generally non-porous, heavy, and air-tight materials. Table 11.13 shows the TL of some common building materials.

Soft porous materials and lightweight materials like rock or glass fibre on the other hand do not prevent the passage of sound; they absorb sound. When acoustic energy is incident on a barrier surface, some of it will be reflected and some will be transmitted through the barrier. Sound TL measures how effectively a material

Table 11.13 Transmission loss of building materials

Barrier	TL (dB)
60 mm thick heavy wood door	25–30
100 mm brick	45
100 mm brick, plastered	47
200 mm brick, plastered	50
50 mm thick hardwood panel	30

performs as a sound barrier. Facilities for sound TL measurements are used for evaluating windows, doors, and other elements that are built into walls.

The fraction of incident energy transmitted is called the sound power transmission coefficient. The sound TL is 10 times the logarithm of the reciprocal of the ratio of the energy transmitted through the specimen to the energy incident on it and is expressed in dB. It represents the loss in sound power due to transmission through the specimen.

Aim of the Experiment: To measure the sound transmission loss using sound pressure level and reverberation time as per ASTM E-90.

Theory: Theory behind sound transmission is described in Sect. 4.5. The expression for TL in dB is

$$TL = 10 \log_{10} \left(\frac{1}{\tau} \right) \tag{11.37}$$

where τ = the sound power transmission coefficient. The higher the TL, the lower is the sound passing through the wall.

There are two approaches to the measurement of TL. Both involve the measurement of SPL. The difference between the two is that one additionally uses RT, while the other uses sound intensity measurements. The mathematical expression for TL is

$$TL = L_{p(source)} - L_{p(receive)} + 10 \log_{10}(S/a)$$

$$TL = NR + 10 \log_{10}(S/a) \tag{11.38}$$

where

$L_{p(source)}$ = space–time-averaged sound pressure level in the source chamber in dB,

$L_{p(receive)}$ = space–time-averaged sound pressure level in the receive chamber in dB,
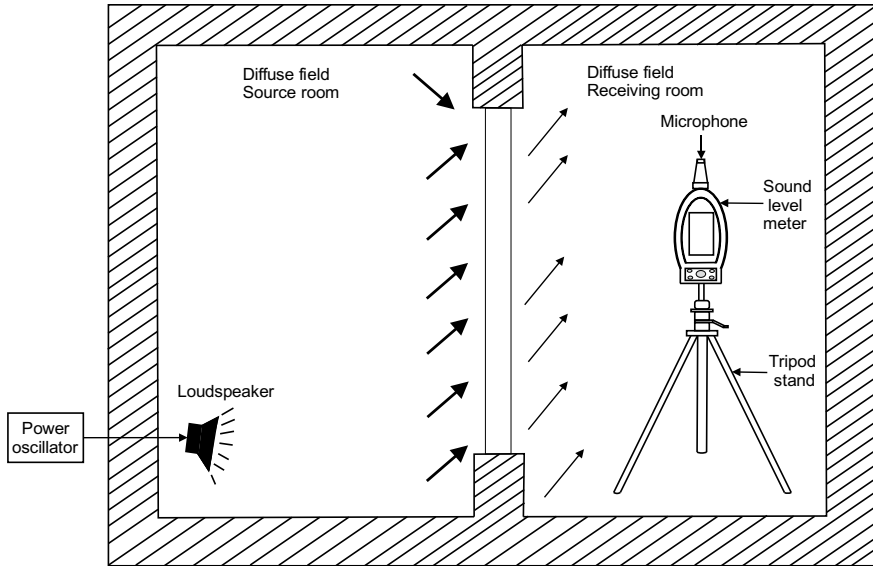
Fig. 11.21 Facility for measuring sound transmission loss

NR = noise reduction,

S = surface area of the test specimen in m^2, and

a = total absorption in Sabines (m^2) in the receiver chamber with the test specimen in place.

Test Procedure and Instrumentation: This method involves the measurement of the sound TL of a specimen based on the difference in the space- and time-averaged SPLs between a source chamber and a receiving chamber. Each chamber should have a minimum volume of 50 m^3 coupled through an opening equal to the size of the test specimen. The measurement is based on the assumption that the sound fields in the two chambers are diffuse. Noise is generated in the source room and a part of the sound energy is transmitted through the test specimen into the receiving room. For the generation of acoustic energy, a pink/white noise generator is typically used in conjunction with a power amplifier, and speakers of sufficiently high rating are chosen so as to generate SPLs in the receiving chamber that are at least 10 dB higher than the ambient SPLs. They are measured in both the source and receiver rooms in 1/3 octave bands in the frequency range 100 and 5000 Hz using an integrating SLM with 1/3 octave band analysis and RT measurement features. The frequency range of interest may be extended down to 63 Hz octave band in large chambers and up to the 8,000 Hz octave band. The microphones chosen should have a flat frequency response over the frequency range of interest. Figure 11.21 shows a facility for measuring sound TL.

It is advisable to have both the rooms supported on isolation springs to ensure efficient vibration and sound isolation between the two rooms. The support frame for the specimen is generally fixed to a concrete platform isolated from both chambers

Table 11.14 Transmission loss measurements

Sl. No.	Frequency (Hz)	$L_{p(source)}$	$L_{p(receive)}$	TL (dB)
1.	63.5			
2.	125			
3.	250			
4.	500			
5.	1000			
6.	2000			
7.	4000			
8.	8000			

to reduce the transmission of structure-borne vibration from the structure to the two rooms and to ensure that the only path between the rooms is through the specimen being tested. The TL limit of the test facility and any filler walls used to adapt the test specimen to the test opening should be known and the difference in levels should be corrected to account for their acoustical properties. The absorption coefficient of the receiving chamber is to be determined from measurements of the RT. A facility for testing floors may also be created which has one room above the other. There may be multiple paths other than through the test specimen in this case.

The major sources of measurement errors in the TL measurement which are highest at low frequencies are spatial/temporal variations in (i) the SPLs in both chambers and (ii) the decay rates in the receive chamber. These uncertainties may be reduced by spatial/temporal sampling and averaging of these parameters with a sufficiently large number of samples so as to ensure 95% confidence interval in the TL to the following values: 3 dB in the 125 and 160 Hz bands, 2 dB in the 200 and 250 Hz bands and 1 dB in the 315–4000 Hz bands. Table 11.14 shows how the readings may be tabulated.

Questions

(1) Why are the spatial/temporal variations in the SPLs in both chambers highest at low frequencies?
(2) Why are there spatial/temporal variations in the decay rates in the receive chamber?
(3) What properties of a material decide whether it will perform effectively as a sound barrier?
(4) When is a pink noise generator preferred to a white noise generator?
(5) What does the PSD of pink noise look like?

11.12   Measurement of Transmission Loss Using Sound Pressure Level and Sound Intensity

DID YOU KNOW that TL is used in many fields of engineering? Transmission loss in building acoustics has already been defined as the decrease in sound intensity caused by a wall or other structure at a given frequency. TL along with insertion loss (IL) is used to describe the acoustic performances of mufflers. The higher the transmission loss of a system, the better will be its noise reduction capabilities. In underwater acoustics, TL describes the reduction in sound intensity caused by a damping structure in the wave path at a given frequency. Transmission loss in telephony is defined as the ratio of power into the cable to power out of the cable and is defined in dB.

Introduction: The measurement of the sound TL of a specimen by this method is similar to the earlier method, except that it is based on the assumption that the sound field in the source chamber is diffuse, and that in the receiving chamber it is essentially a free field. The quantities to be determined are sound pressure level in the source reverberation chamber and the average sound intensity radiated by the specimen into the receiving chamber.

Aim of the Experiment: To measure the sound transmission loss of the given specimen using the sound pressure level measured in a source reverberation chamber and sound intensity in a receiver chamber.

Theory: The mathematical expression for transmission loss is

$$\text{TL} = L_{p(source)} - 6 - L_{I(receive)} - 10\log_{10}(S_m/S) \tag{11.39}$$

where

$L_{p(source)}$ = space–time-averaged sound pressure level in the source chamber in dB measured with a microphone on a rotating boom,

$L_{I(receive)}$ = averaged sound intensity level radiated normally to and averaged over the measuring surface in the receive chamber,

S_m = surface area of the measuring surface in the receiving room in m^2, and

S = surface area of test specimen (area radiating sound to the receiving room) in m^2.

The assumption behind this computation is that the angle of incidence of the energy in the source chamber is random in direction and that the SPL in the source chamber is used to get the total energy incident on the specimen. The constant 6 dB term arises from the fact that the source radiates into $\frac{1}{4}$ space, resulting in a 6 dB difference between the sound pressure level in the source room and the incident sound intensity.
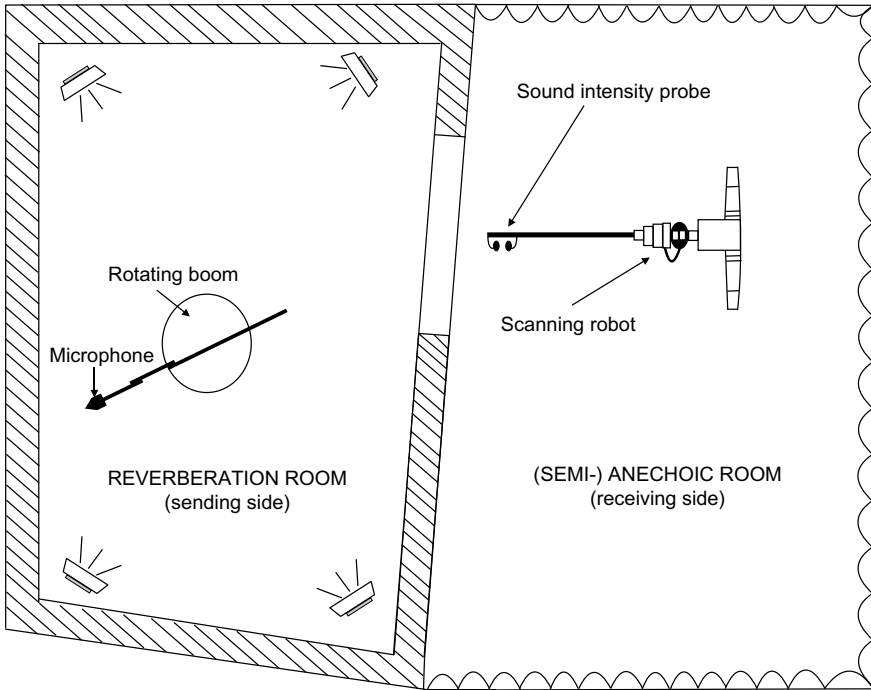
Fig. 11.22   Setup for transmission loss measurements

Instrumentation and Test Procedure: The test panel is fixed in an aperture between the reverberant source chamber and the semi-anechoic receive chamber as shown in Fig. 11.22. Most of the instrumentation is similar to what is required for the earlier method described in Sect. 11.11. Additionally, a sound intensity probe with two phase-matched microphones with uniform frequency response over the frequency range of interest, a dual-channel analyser with cross-spectrum analysis, and 1/1 or 1/3 octave band filters is required. Broadband noise is generated to produce a diffuse field in the reverberant room. Sound pressure level which is averaged over space and time is obtained in the source chamber. Sound intensity level normal to the measuring surface of the specimen and averaged over it is measured in the receiving room. This technique for TL has not been standardized, and the uncertainties in measurement are not well documented. Results may be recorded as shown in Table 11.15.

Questions

(1)   How can measurement errors at low frequencies be reduced?
(2)   How can you reduce flanking noise?
(3)   How will specimen size affect transmission loss?
(4)   What do you mean by mass law in the context of panels?

Table 11.15 Measurements for sound TL

1/1 band frequency (Hz)	$L_{p(source)}$ (dB)	$L_{i(receive)}$ (dB)	TL (dB)
63.5			
125			
250			
500			
1 k			
2 k			
4 k			
8 k			

11.13 Measurement of Transmission Loss and Single Number Ratings (STC) Using Impedance Tube

DID YOU KNOW that in 1961 Sound Transmission Class (STC) was introduced as an integer rating of soundproofing effectiveness? It is a measure of how well a partition attenuates sound and is used in the United States to rate transmission loss of interior partitions, ceilings and floors, doors, windows, and exterior wall configurations of buildings. However, it should not be totally relied upon for real-world soundproofing expectations for the reason that STC is computed by taking the experimentally measured TL values at only 16 standard frequencies over the range 125 to 4000 Hz.

Introduction: TL can be measured by the four-microphone method in a plane wave tube. This method uses two sets of microphones, similar to that used in the measurement of absorption coefficient using the two-microphone method. The sound source (loudspeaker) is fixed at one end of the tube and the test sample is placed almost midway in a holder. A pair of microphones (Mic. 1 and Mic. 2) is mounted in the upstream tube and another pair of microphones (Mic. 3 and Mic. 4) is mounted in the downstream tube so as to measure both incident and reflected waves.

Theory: The four microphone positions are denoted by x_1, x_2, x_3, and x_4 in Fig. 11.23. The reference position ($x = 0$) is at the front surface of the specimen. The microphones are placed equidistant such that distance

$$s = |x_1 - x_2| = |x_3 - x_4| \tag{11.40}$$

If a 1-D plane wave in the tube represented as $pe^{i(\omega t - kx)}$ hits the sample, part of the energy gets reflected back, part of it gets absorbed by the sample, and part of it passes
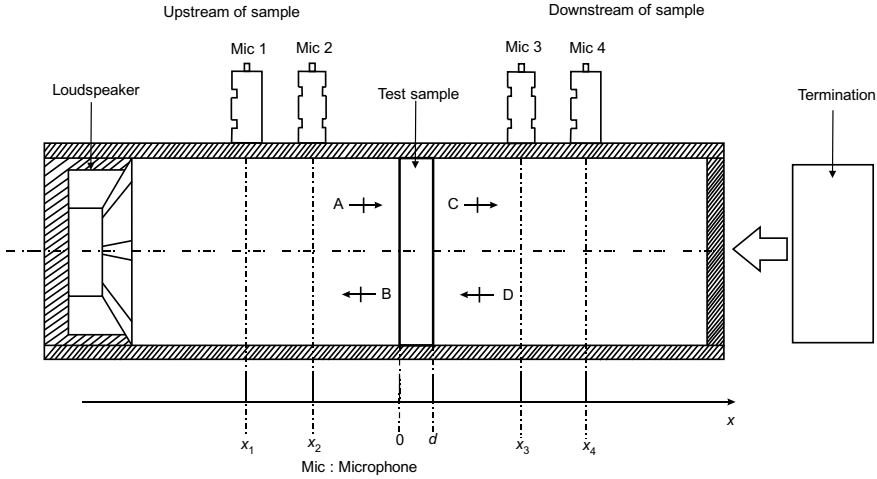
Fig. 11.23 Schematic of transmission loss measurement using impedance tube

into the receiving tube. The plane wave that exits the test specimen impinges on the termination at the end of the receiving tube and is reflected back. A 400 mm thick urethane foam may be used to obtain an anechoic termination in the downstream tube.

The sound pressures at locations 1, 2, 3, and 4 (after eliminating the time-dependent parts) may then be expressed as

$$p_1 = Ae^{-ikx_1} + Be^{ikx_1}$$
$$p_2 = Ae^{-ikx_2} + Be^{ikx_2}$$
$$p_3 = Ce^{-ikx_3} + De^{ikx_3}$$
$$p_4 = Ce^{-ikx_4} + De^{ikx_4} \tag{11.41}$$

The transmission coefficient $\tau = C/A$ and transmission loss TL is given by

$$\text{TL} = -20\log_{10}|\tau| = 20\log_{10}\left|\frac{e^{iks} - H_{12}}{e^{iks} - H_{34}}\right| - 20\log_{10}|H_t| \tag{11.42}$$

where

$$H_t = \sqrt{|S_d/S_u|}$$

Here, k is the wavenumber, A and B are the amplitudes of the incident and reflected waves in the upstream tube, and C and D are the amplitudes of the transmitted and reflected waves in the downstream tube; $H_{12} = p_2/p_1$ is the transfer function between pressures p_2 and p_1, and $H_{34} = p_4/p_3$ is the transfer function between pressures p_4 and p_3; and S_u is the auto-spectrum in the upstream tube and S_d that in

Table 11.16 Measurements for sound TL using four-microphone method

1/1 band frequency (Hz)	H_{12}	H_{34}	H_t	τ	TL (dB)
63.5					
125					
250					
500					
1 k					
2 k					
4 k					
8 k					

the downstream tube.

Test Procedure: A broadband stationary random noise from a function generator and amplified by a power amplifier is used to excite the loudspeaker in the source tube and the resulting plane waves propagate in the tube. By measuring the sound pressures at the four fixed microphones, two in the source tube and two in the receiving tube, a complex transfer function is calculated using a four-channel digital analyser. The contributions of the incident, reflected, and transmitted waves are separated and TL is calculated as shown in Eq. 11.42. Results may be tabulated as shown in Table 11.16.

Single Number Ratings Plots of sound TL data are complex and hence the sound TL of a specimen is often expressed in terms of a single number rating. Sound Transmission Class (STC) is the most common rating of the specimen and is determined in accordance with the calculation procedures of ASTM E413. The STC curve is a sliding contour that is fitted to the measured data plotted such that

- the maximum deviation in any 1/3 octave band (i.e., for a TL value which falls below the contour) does not exceed 8 dB.
- the sum of the deviations falling below the reference contour in all bands does not exceed 32 dB.

Once the appropriate contour satisfying these conditions has been selected, the STC rating of a specimen is determined by comparing the measured TL values in the 16 contiguous 1/3 octave bands with centre frequencies between 125 and 4000 Hz, with the values of the STC reference contour shown in Fig. 11.24. The STC rating of a specimen corresponds to the TL value at 500 Hz for the highest reference contour which meets the above two conditions simultaneously and is expressed as a single number. As in the case of TL, the higher the STC, the better the sound insulation.

Outside the USA, the Weighted Sound Reduction Index (SRI) or R_w defined as per ISO standard is used. This is defined in terms of the transmission coefficient, but uses a slightly different frequency range (100–3150 Hz). Measurement of SRI of a

Fig. 11.24 Reference
contour for sound
transmission class
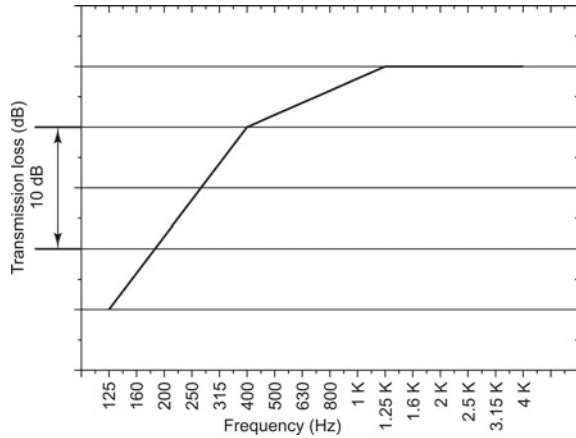
partition is also conducted in a pair of reverberation chambers as described for STC. In general, the two ratings give either the same number or only 1–2 dB difference, so they can generally be interchanged.

Questions

(1) What is the problem relying on an STC number alone for sound isolation by a panel?
(2) What are the main sources of error in this method?
(3) How does the termination at the downstream end of the tube affect the measured TL?
(4) What happens if the output of the loudspeaker is non-linear?
(5) If the spacing between microphones is not uniform as assumed in this method, how do the results change?

Bibliography

1. American National Standards Institute, ANSI S1.40-Specifications for Acoustic Calibrators, 1984, R1997.
2. ASTM E1332, Standard Classification for Determination of Outdoor-Indoor Transmission Class.
3. ASTM E413-04, Classification for Rating Sound Insulation.
4. ASTM E90, Standard Test Method for Laboratory Measurement of Airborne Sound Transmission Loss of Building Partitions.
5. Barnard, A. R., & Rao, M. D. (2004). Measurement of sound transmission loss using a modified four microphone impedance tube. In *Proceedings of the NOISE-CON 2004*, Baltimore, Maryland, 12–14 July 2004.
6. Bies, D. A., & Hansen, C. H. (2003). *Engineering noise control: Theory and practice*. New York, USA: Taylor & Francis.

7. Brüel & Kjær. *Acoustic Noise Measurements, BT 0010-12*. Denmark.
8. Brüel & Kjær. *Frequency Analysis of Sound, BA 7660-06*. Denmark.
9. Brüel & Kjær. (1986). *Noise control-principles and practice*. Denmark.
10. BS EN 1793-1:1998, Road Traffic Noise Reducing Devices-Test Method for Determining the Acoustic Performance, Part 1: Intrinsic Characteristics of Sound Absorption.
11. BS EN 1793-3:1998, Road Traffic Noise Reducing Devices-Test Method for Determining the Acoustic Performance, Part 3: Normalized Traffic Noise Spectrum.
12. BS EN ISO 140-1:1997, Acoustics-Measurement of Sound Insulation in Buildings and of Building Elements, Part 1: Requirements for Laboratory Test Facilities with Suppressed Flanking Transmission.
13. BS EN ISO 140-2:1991, Acoustics-Measurement of Sound Insulation in Buildings and of Building Elements, Part 2: Determination, Verification and Application of Precision Data.
14. BS EN ISO 140-3:1995, Acoustics-Measurement of Sound Insulation in Buildings and of Building Elements, Part 3: Laboratory Measurement of Airborne Sound Insulation of Building Elements.
15. BS EN ISO 354:2003, Acoustics-Measurement of Sound Absorption in a Reverberation Room.
16. BS EN ISO 717-1:1997, Acoustics-Rating of Sound Insulation in Buildings and of Building Elements, Part 1: Airborne Sound Insulation.
17. BSI BS EN 1793-2, Road Traffic Noise Reducing Devices-Test Method for Determining the Acoustic Performance, Part 2: Intrinsic Characteristics of Airborne Sound Insulation.
18. Burns, W., & Robinson, D. W. (1970). *Fundamentals of acoustics hearing and noise in industry*. London: Her Majesty's Stationery Office.
19. Chung, J. Y. (1977). Cross-Spectral Method of Measuring Acoustic Intensity, Research Publication of General Motors Research Laboratory (GMR-2617), Warren (USA).
20. Chung, J. Y., & Pope, J. (1978). Practical measurement of acoustic intensity-the two-microphone cross-spectral method. In *Inter-Noise 78: Designing for Noise Control, Proceedings of the International Conference*, San Francisco, CA, May 8–10, 1978 (A79-15551 04-71), Noise Control Foundation, Poughkeepsie, N.Y. (pp. 893–900)
21. Diehl, G. M. (1973). *Machinery acoustics*. New York: Wiley.
22. Doebelin, E. O. (2004). *Measurement systems: Application and design*. New York: McGraw-Hill Professional.
23. Fahy F. J. (1977). Measurement of acoustic intensity using the cross-spectral density of two microphones signals. *Journal of the Acoustical Society of America, 62* (L), 1057–1059.
24. Fahy F. J. (1989). *Sound intensity*. London, (UK): Elsevier Applied Science.
25. Foreman, J. (1990). *Sound analysis and noise control*. New York: Van. Nostrand Reinhold.
26. Fry, A. (1988). *Noise control in building services*. Oxford, New York: Pergamon Press.
27. Goelzer, B., Hansen, C. H., & Sehrndt, G. A. (2001). Occupational Exposure to Noise: Evaluation, Prevention, and Control, Special Report S 64. Dortmund and Berlin: Federal Institute for Occupational Safety and Health.
28. Hassal, J. R., & Zaveri, K. (1979). *Acoustic noise measurements*. Brüel & Kjær.
29. ISO 3740:2000, Acoustics-Determination of Sound Power Levels of Noise Sources-Guidelines for the Use of Basic Standards and for the Preparation of Noise Test Codes.
30. ISO 3741:1999, Acoustics-Determination of Sound Power Levels of Noise Sources Using Sound Pressure-Precision Methods for Reverberation Rooms.
31. ISO 3743-1:1994, Acoustics-Determination of Sound Power Levels of Noise Sources-Engineering Methods for Small, Moveable Sources in Reverberant Fields, Part 1: Comparison Method for Hard-Walled Test Rooms.
32. ISO 3743-2:1994, Acoustics-Determination of Sound Power Levels of Noise Sources Using Sound Pressure-Engineering Methods for Small, Moveable Sources in Reverberant Fields, Part 2: Methods for Special Reverberation Test Rooms.
33. ISO 3744:1994, Acoustics-Determination of Sound Power Levels of Noise Sources Using Sound Pressure-Engineering Method in an Essentially Free Field over a Reflecting Plane.
34. ISO 3745:2003, Acoustics-Determination of Sound Power Levels of Noise Sources Using Sound Pressure-Precision Methods for Anechoic and Hemi-Anechoic Rooms.

35. ISO 3746:1995, Acoustics-Determination of Sound Power Levels of Noise Sources Using Sound Pressure-Survey Method Using an Enveloping Measurement Surface over a Reflecting Plane.
36. Kinsler, L. E., Frey, A. R., Coppens, A. B., & Sanders, J. V. (2000). *Fundamentals of acoustics*. New York: Wiley.
37. Moore, B. C. J. (2003). *An introduction to the psychology of hearing*. Orlando, Florida: Academic Press.
38. Pieper, R. M., Alexander, J., Bolton, J., & Yoo, T.-W. (2007). Assessment of Absorbers in Normal Incidence Four-Microphone Transmission-Loss Systems to Measure Effectiveness of Materials, in Lateral-Flow Configurations of Filled or Partially Filled Cavities, *SAE 2007-01-2190*, May 2007.
39. Reynolds, D. D. (1981). *Engineering principles of acoustics: Noise and vibration*. USA, Boston: Allyn & Bacon.
40. Robinson, D. W. (1971). Towards a unified system of noise assessment. *Journal of Sound and Vibration, 14*, 279–98.
41. Sujatha, C. (2010). *Vibration and acoustics: Measurement and signal analysis*. Tata McGraw-Hill Education.
42. Suter, A. H. (1998). *Noise, ILO Encyclopaedia of Occupational Health and Safety* (4th ed.). Geneva: International Labour Organization.
43. Wilson, C. E. (1994). *Noise control: Measurement, analysis and control of sound and vibration*. Malabar, Florida: Krieger Publishing Company.

Chapter 12
Common Experiments in Stress Analysis

DO YOU KNOW the history of the resistance strain gauge which is a valuable transducer for experimental stress analysis? In 1856, William Thomson (later Lord Kelvin) submitted a paper describing his investigation of the electrical properties of metals to the Royal Society of London. He reported that the electrical resistance of copper and iron wires, when subjected to a tensile load, varied with the tensile strain. Besides, he found that the iron wire showed a greater increase in resistance than did the copper wire for the same strain. Finally, he established three important facts: (i) the dependence of the changes in resistance of the wire to changes in strain, (ii) different sensitivities of different materials, and (iii) the use of the Wheatstone bridge to measure these resistance changes accurately, all of which have helped the development of the electrical-resistance strain gauge. However, it was not until the late 1930s that strain gauges based on Lord Kelvin's experiment became commercially available.

12.1 Strain Gauge Installation

Strain gauge installation by itself is an art. If gauges are not bonded properly, the strain readings obtained will be erroneous and the whole exercise of experimental stress analysis becomes a waste.

Handling of Strain Gauges

   i. Strain gauges should never be touched by hand. They are meant to be handled with rounded tweezers only.
  ii. Gauges are to be held at the backing support, never at the active grid region.

iii. They do not require to be cleaned before bonding, unless the user has accidentally contaminated them.

Surface Preparation

i. Special care has to be taken to ensure that the strain gauge is not contaminated by cutting oil containing silicon, since even a very small amount of silicon will make a bond useless.

ii. The specimen on which the gauge has to be bonded has to be degreased using a solvent for at least 100 to 150 mm around the gauge. This can be done with solvents such as Acetone, Isopropyl Alcohol (IPA), or Methyl-Ethyl-Ketone (MEK) to remove greases, oils, organic contaminants, and soluble chemical residues. Porous materials like cast iron, cast aluminium, and titanium may be degreased using a hot vapour degreaser or an aerosol-type spray degreaser or may be placed in an ultrasonically agitated liquid bath to drive off the absorbed contaminants

Surface Abrasion and Conditioning

i. Any scale, paint, rust, or galvanized coating on the surface of the specimen (extending beyond the space required for the strain gauge) must be removed using sandpaper with grit size 320–400 or by sandblasting with aluminium oxide grit of size 120 to 400, and any remaining dust must be removed using a cloth and solvent in order to obtain a suitable surface texture for bonding.

ii. An appropriate conditioner should be applied and the surface of the specimen should be scrubbed with a soft applicator until there is no dirt remaining on the specimen.

iii. After conditioning, the surface of the specimen must be cleaned with a fresh sponge.

iv. The surface must be set at the pre-defined optimum alkalinity of strain gauge adhesives (around 7.0 to 7.5 pH) by using a neutralizer and then it must be cleaned once again with a fresh sponge.

Marking the Gauge Location Layout Lines

The location of the strain gauge can be marked using a brass needle or a 4H pencil. For accurate positioning, the axis along which strain measurement is to be made and a perpendicular line must be drawn.

Bonding the Gauges

i. Bonding is to be done on a dust-free workbench since utmost cleanliness is required.

ii. The application area should be cleaned with a cotton-tipped applicator dipped into a solvent; the cotton tips should be replaced frequently until the application area is completely clean.

iii. It is desirable to have the surface area cleaned to extend beyond the space required for the strain gauge.

iv. The strain gauge should be taken out of the packing and put on a chemically cleaned glass plate with its bonding side on top.

v. Both the area of the strain gauge and the workpiece should be coated with a thin layer of an appropriate adhesive.

vi. The strain gauge should be positioned properly and smoothened out by hand using a piece of teflon film to press out any excess adhesive.

Curing of the Adhesive and Inspection of Bond Quality

i. During the setting of the adhesive, the gauge has to be clamped with constant pressure so as to obtain a constant thickness of the adhesive.

ii. Once the adhesive has set, the pressure may be released and the teflon film removed.

iii. Inspection can be done using a magnifying glass or a microscope with a magnification of 10.

iv. Voids or unbonded areas or round bubbles indicate improper clamping or adhesive application or drying time, in which case it may be desirable to paste a fresh strain gauge.

Soldering of Strain Gauges

i. For soldering the gauges, it is desirable to use a temperature-controlled soldering station set approximately to 380–400 °C.

ii. The soldering iron tip has to be cleaned well and wet with solder.

iii. Some fluxes such as liquid rosin may be applied to the strain gauge lead while tinning.

iv. The external wire to be soldered is to be cut to the required length, insulation is to be removed, and the wire is to be tinned at the ends.

v. Holding the wire flat on the gauge lead in the same direction for approximately 1 s, soldering is to be done till the solder becomes solid.

vi. After soldering, it is to be ensured that the resulting solder has a good connection to the gauge lead.

vii. The solder joint is to be checked to see if it is homogeneous, smooth, and shiny and if the wire is soldered on the gauge.

viii. Any excess flux residues may be removed by brushing the part with a medium-sized brush soaked in solvent.

Protecting the Strain Gauge

The strain gauge and surrounding area should be protected with a coating against moisture and other chemical influences. Any coating suiting the environmental conditions and expected temperatures may be used.

Check for Insulation Resistance

The insulation resistance between the gauge grid and the workpiece should be greater than 1000 MΩ, a value which is not difficult to achieve. If the insulation resistance is too low, it may be required to paste a fresh strain gauge or do rewiring.

12.2   Axial Load: Quarter Bridge with Single Active Gauge

INTERESTING FACTS: The Wheatstone bridge for measuring resistance in a circuit was 'discovered' in 1843, but had been invented a decade earlier. The inventor was not Wheatstone; on the other hand, it was a British physicist and mathematician, Samuel Hunter Christie, who developed the circuit to measure unknown electrical resistances and described it in 1833. By using a galvanometer to balance the current flowing through the two branches of the circuit, Christie could, with a little mathematics, determine the value of the unknown resistor. The bridge worked because of the special diamond-shaped arrangement of the four resistors. It was Wheatstone, however, who was responsible for popularizing the diamond-shaped bridge with its battery and galvanometer. He called the circuit a 'Differential Resistance Measurer' and gave Christie full credit in his 1843 Bakerian lecture. The circuit soon came to be called the Wheatstone bridge and was used in telegraphy. This bridge can be used to determine unknown capacitances and inductances by changing the type of elements contained in its arms and by using an alternating current (AC) excitation source. This sensitive and accurate method for measuring resistance is still widely used today.

Introduction: The uniaxial tensile test is the simplest and most useful test that can be conducted on a strip of the material to be characterized. Such a test is more generally conducted to measure important material properties like tensile strength, yield strength, modulus of elasticity, resilience, and elongation. The quarter bridge with a single active gauge is the simplest configuration that can be used for the measurement of axial strain, though it has a considerable amount of non-linearity.

Aim of the Experiment: To experimentally measure the axial strain due to tensile loading of a bar.

Theory: Figure 12.1a shows an axial load P applied to a bar of width b and thickness t. Figure 12.1b shows a single active gauge R_1 pasted on the top surface to sense the strain. The bridge is given a direct current (DC) excitation voltage E_i. Initially, the bridge is balanced without applying any load. On application of a load, the strain ε_1 and output voltage E_o change.

   Load P causes the axial strain given by

$$\varepsilon = \frac{P}{Ebt} \tag{12.1}$$

where E is Young's modulus of elasticity.
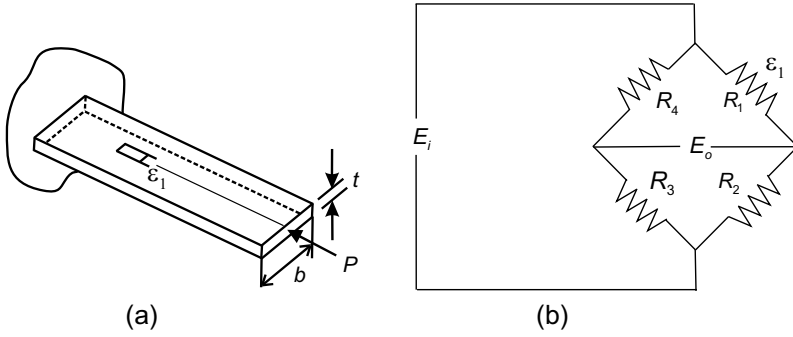
Fig. 12.1 Axial loading: **a** bar with load and strain gauge, **b** quarter bridge with gauge

The output of the bridge with this single active gauge $\varepsilon = \varepsilon_1$ after loading and bridge unbalance is given by

$$\frac{E_o}{E_i} = \frac{R_1 + \Delta R_1}{R_1 + \Delta R_1 + R_2} - \frac{R_4}{R_3 + R_4} = \frac{F\varepsilon_1}{4 + 2\,F\varepsilon_1} \qquad (12.2a)$$

Here, the following relation has been used.

$$\frac{\Delta R/R}{\Delta L/L} = F \quad \text{Hence} \quad \frac{\Delta R}{R} = F\varepsilon \qquad (12.2b)$$

where F is the gauge factor. When $F = 2$ and with $R_1 = R_2 = R_3 = R_4$ before loading, the above equation becomes

$$\frac{E_o}{E_i} = \frac{2\epsilon_1}{4 + 4\varepsilon_1} \qquad (12.3)$$

This equation with the output in terms of mV/V and with strain expressed as microstrain is

$$\frac{E_o}{E_i} = \frac{2\epsilon_1 \times 10^{-3}}{4 + 4\varepsilon_1 \times 10^{-6}} \simeq \frac{\varepsilon_1 \times 10^{-3}}{2} \qquad (12.4)$$

Strain being in terms of μm/m, the term $F\varepsilon$ in the denominator is usually much less than 4; this gives rise to some non-linearity with a single active gauge. The output in terms of mV/V may also be written as

$$\frac{E_o}{E_i} = K\varepsilon_1 \times 10^{-3}(1 - \eta) \qquad (12.5a)$$

where $K = F/4$ and η is the non-linearity given by

$$\eta = \frac{F\varepsilon_1 \times 10^{-6}}{2 + F\varepsilon_1 \times 10^{-6}} \qquad (12.5b)$$

The magnitude of this non-linearity is roughly 0.1% per 1000 $\mu\varepsilon$ of strain. This error can be neglected in most applications, especially at low strain levels. The output voltage can be increased by increasing the excitation voltage E_i. The typical input voltages in commercially available systems are 0.5, 1, 2, 2.5, and 5 V. This bridge configuration is sensitive to temperature variations; this drawback can be overcome by using either a self-temperature-compensated strain gauge or by using a half bridge. The bridge output voltage being proportional to excitation voltage, the higher the excitation voltage, the higher is the sensitivity of the circuit. But with a high excitation voltage, the power dissipated in the gauge increases, leading to heating.

Test Setup and Procedure: A common setup for the measurement of axial strain using a universal testing machine (UTM) is shown in Fig. 12.2. It consists of a fixed and a moving crosshead, between which the specimen on which axial load is to be applied is fixed. The specimen is in the form of a dog bone with wide ends and a narrow middle; this is a standard configuration for tensile testing of strip materials as per ASTM E8. Since it takes time to machine a dog bone specimen, straight-sided samples are also permitted by ASTM. The grips of the UTM firmly hold the specimen at the wide ends. The midsection of the specimen has a narrower width which ensures that the stress is concentrated in this test area and that the fracture also occurs here. The strain is measured by pasting a strain gauge R_1 here axially and its output is connected to a Wheatstone bridge. The gauge may be a 3 mm or
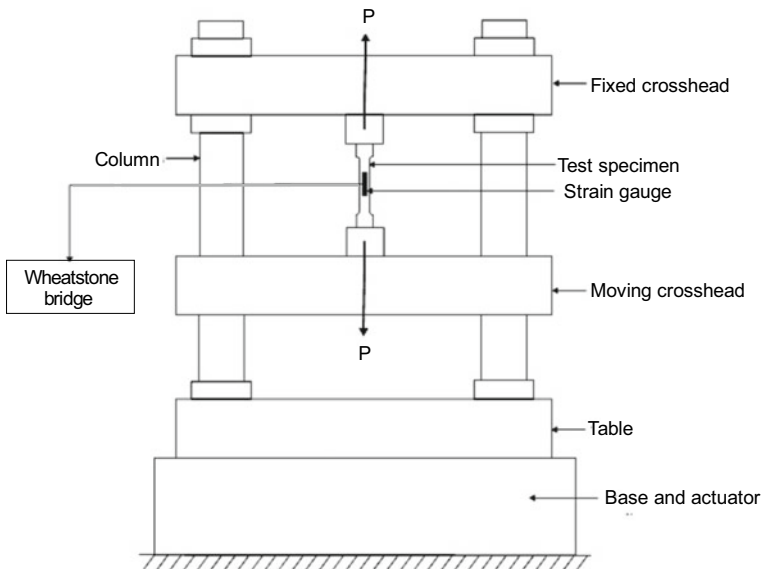
Fig. 12.2 Experimental setup for axial loading
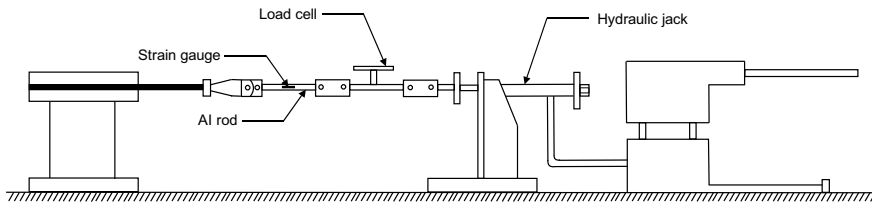
Fig. 12.3 Low-cost setup for axial loading

Table 12.1 Table for measurement of axial strain

S.No.	Load, N	O/p voltage, V	Measured Strain μɛ	Non-linearity η	Computed Strain μɛ	Discrepancy, %
1.						
2.						
3.						

6 mm foil-type gauge, and stress is calculated knowing the axial load on the grips. Any strain or deformation occurring outside the test area can lead to inaccuracies in the test results.

Alternately, a typical low-cost setup for axial loading can be fabricated in the workshop as shown in Fig. 12.3. This consists of a hydraulic jack for loading, a load cell for measuring the load, and the test sample in the form of a bar. The active gauge is pasted on the bar and connected to one arm of the Wheatstone bridge, which along with the excitation source and output voltmeter is called a carrier frequency amplifier.

Initially, the bridge is balanced with the active gauge and three dummy gauges without applying any load. It is preferable to choose resistance-matched gauges from the same lot so that the bridge is balanced easily. The dummy gauges are typically three identical gauges pasted on a block of the same material as the bar and at the same temperature. It is to be ensured that the block is not subjected to any load. Alternately, the internal resistors in the carrier frequency amplifier may be used as the dummy resistors. The axial load has to be increased in steps after null balancing the bridge, and the output voltage, which is a function of the strain, is to be recorded. It is to be noted that this arrangement of the strain gauge is sensitive to both axial and bending loads. Hence, care has to be taken in the application of a true axial force. Besides, any variation in the resistance of the strain gauge due to temperature effects is not automatically compensated for.

Results and Analytical Validation: A typical tabulation may be made as shown in Table 12.1. Plots of output voltage versus load, load versus strain, and non-linearity versus strain can be made. Knowing the dimensions of the bar, the applied load, and Young's modulus of the material, the strains for the applied loads may be calculated and the discrepancy between the measured and calculated strains may be found out. The experiment may also be conducted while gradually reducing the load to zero.

Exercises

(1) Calculate the non-linearity as a function of strain and make a plot as a function of strain.
(2) Why is a gauge resistance of 120 Ω so common?
(3) Is there any precaution to be taken while choosing the active and dummy gauges?
(4) Why is a gauge factor of 2 so common?
(5) How do you ensure that the active gauge is not affected by bending loads?
(6) Make a plot of strain while loading and unloading. Comment on the results obtained.

12.3   Axial Load: Half Bridge with Two Active Gauges in Opposite Arms

INTERESTING FACTS: The Wheatstone bridge is electrically equivalent to two parallel voltage divider circuits with resistors R_1 and R_2 forming one arm and resistors R_3 and R_4 forming the second arm (Fig. 12.1b). This bridge can be used (i) for finding the absolute value of a resistance by comparing it with a known resistance or (ii) for finding relative changes in resistance. It is the second method which is used in strain gauge techniques. Relative changes of resistance in the strain gauge are typically of the order of 10^{-4} to 10^{-2} Ω/Ω and can be measured with great accuracy using this bridge. A half bridge circuit has two active gauges; both could be pasted in the axial direction or one in the axial direction and the other in the transverse or Poisson direction. The outputs would be different in the two cases.

Introduction: The half bridge with two active gauges in opposite arms can be used for the measurement of axial strain with better sensitivity than what is obtained with a quarter bridge. The same experimental setup as shown in Fig. 12.2 or 12.3 may be used here also, the only difference being in the bridge configuration.

Aim of the Experiment: To experimentally measure the axial strain due to tensile loading of a bar as before, but using a half bridge circuit.

Theory: In Fig. 12.4a, an axial load P is applied to a bar of width b and thickness t as before. It may be noted that a second strain gauge is pasted on the underside of the bar as a mirror image of the gauge on the upper side. Hence, the two active gauges R_1 and R_3 are now electrically connected in opposite arms of the Wheatstone bridge as shown in Fig. 12.4b.
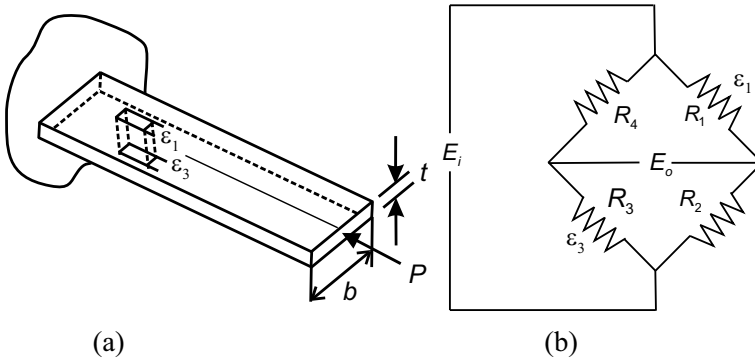
Fig. 12.4   Axial loading: **a** bar with load and strain gauges, **b** half bridge with gauges

The expression for the output voltage for this configuration may be obtained as described below.

With R_1 and R_3 as active gauges

$$\varepsilon_1 = \frac{P}{Ebt}; \quad \varepsilon_3 = \varepsilon_1 \tag{12.6}$$

$$\frac{E_o}{E_i} = \frac{F\varepsilon_1}{2 + F\varepsilon_1} \tag{12.7a}$$

When $F = 2$

$$\frac{E_o}{E_i} = \frac{2\varepsilon_1}{2 + 2\varepsilon_1} \tag{12.7b}$$

With a gauge factor F of 2, the above equation with output in mV and strain in microstrains is

$$\frac{E_o}{E_i} = \frac{2\varepsilon_1 \times 10^{-3}}{2 + 2\varepsilon_1 \times 10^{-6}} \simeq \varepsilon_1 \times 10^{-3} \tag{12.8}$$

The above equation in terms of mV/V may also be written as in Eq. 12.5(a), where $K = F/2$ and η is the non-linearity given by

$$\eta = \frac{F\varepsilon_1 \times 10^{-6}}{2 + F\varepsilon_1 \times 10^{-6}} \tag{12.9}$$

Since changes in opposite arms (R_1 and R_3) are of the same sign, they have a numerically additive effect and double the output of the bridge, giving rise to a larger bridge output than that of the quarter bridge. Besides, this configuration cancels bending strains with equal magnitudes and opposite signs. This arrangement, like the quarter bridge, has some non-linearity (approximately 0.1% per 1000 $\mu\varepsilon$ produced by axial loads). Besides, temperature compensation is the poorest of all configurations because any thermal output from the active gauges is additive for this configuration.

Later, it will be seen that the same physical configuration of a beam with gauges on the top and bottom surfaces may be used for the measurement of bending strains using a half bridge; however, the gauges will be connected in adjacent arms.

Test Setup and Procedure: The setup and procedure are the same as outlined in Sect. 12.2. The test rig corresponding to either Fig. 12.2 or Fig. 12.3 may be used. The bridge is null-balanced without any load. The axial load is gradually increased in steps and the output voltage noted. Table 12.1 may be used to tabulate these results. Similar graphs can also be made.

Points to Ponder:

(1) What will the output of a half bridge with two active gauges in adjacent arms be when it is subjected to axial load?
(2) How can you further increase the output with this configuration?
(3) For the same physical configuration of a beam used for the measurement of bending strains using a half bridge with gauges on the top and bottom surfaces, why should the gauges be connected in adjacent arms?
(4) Fill in the blanks: Changes of resistance in adjacent arms (R_1 and R_4 for instance) have a numerically additive effect on the bridge output when the changes are of _____ sign.
(5) Fill in the blanks: When the changes of resistance in opposite arms (R_1 and R_3 for instance) are of the same sign, they have a numerically _____ effect on the bridge output.

12.4 Axial Load: Half Poisson Bridge

INTERESTING FACTS: Conventionally, tensile deformation is considered positive and compressive deformation negative. By definition, Poisson's ratio has a minus sign so that conventional materials have a positive ratio. Almost all common materials, such as rubber, increase in length and become narrower in cross section when they are stretched. This is because, from a continuum point of view, most materials easily undergo changes in shape as determined by the shear modulus G, but are resistant to changes in volume as determined by the bulk modulus K. Materials that have a negative Poisson's ratio are referred to as anti-rubber or dilatational materials because these materials become wider in cross section when stretched. Such materials easily undergo volume changes easily, foam being a typical example.
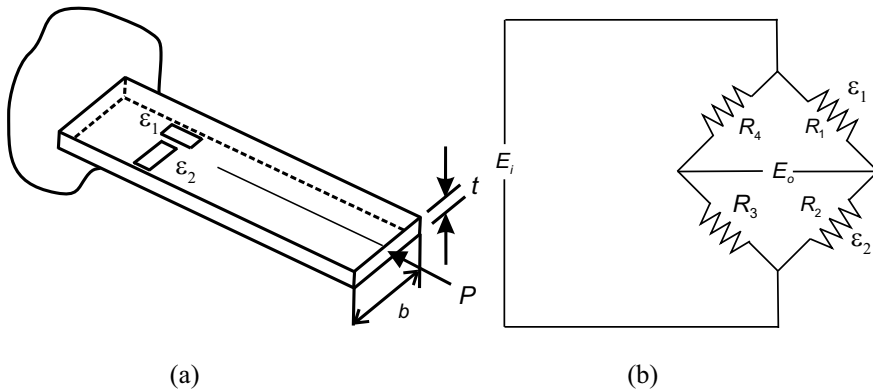
Fig. 12.5 Axial loading: **a** bar with load and strain gauges, **b** half Poisson bridge with gauges

Introduction: This half bridge configuration for axial loading with a longitudinal and Poisson gauge gives higher output than a quarter bridge, but less than that of a half bridge with two active gauges in opposite arms. So, what advantage does this bridge have over the latter? Since the longitudinal and transverse gauges are in adjacent arms, resistance changes of thermal origins will be cancelled when both the active gauges and the specimen undergo the same temperature variation as the derivation that follows will show.

Aim of the Experiment: To experimentally measure the axial strain due to tensile loading of a bar using a Poisson half bridge.

Theory: Figure 12.5a shows an axially loaded bar with a half Poisson bridge and Fig. 12.5b shows the position of the active longitudinal gauge R_1 and lateral Poisson gauge R_2 in the Wheatstone bridge.

With axial load P acting on the bar, the longitudinal strain is

$$\varepsilon_1 = \frac{P}{Ebt}; \quad \varepsilon_2 = -\nu\varepsilon_1 \tag{12.10}$$

where ν is Poisson's ratio.

The ratio of the output to the input voltage is

$$\frac{E_o}{E_i} = \frac{F\varepsilon_1(1+\nu)}{4 + 2F\varepsilon_1(1-\nu)} \tag{12.11}$$

With E_o in mV, the above equation may be reduced to a form shown in Eq. 12.5(a) with

$$K = \frac{F(1+\nu)}{4} \quad \text{and} \quad \eta = \frac{F\varepsilon_1(1-\nu) \times 10^{-6}}{2 + F\varepsilon_1(1-\nu) \times 10^{-6}} \tag{12.12}$$

Test Setup and Procedure: The test setup and procedure are the same as described in Sect. 12.2. Table 12.1 may be used for tabulation.

Questions

(1) Why does this half bridge have better temperature compensation than that described in Sect. 12.3?
(2) Comment on the non-linearity of this bridge.
(3) What material is used for manufacturing foil gauges?
(4) What is ν for the material used?

12.5   Axial Load: Full Poisson Bridge

INTERESTING FACTS: Since time immemorial, devices for the accurate measurement of loads have existed. Some of the earliest load-measuring devices were the mechanical lever scales. These devices can weigh everything from a fraction of a Newton to thousands of Newtons accurately and reliably. Later came the actual pneumatically and hydraulically operated load cells. The present-day industrial load cells did not develop until the nineteenth century. Though the Wheatstone bridge was popularized in 1843 and the bonded wire gauges were developed in the 1940s, it was not until developments in modern electronics happened that strain gauge-based load cells began to be made. They are used everywhere today in the industrial world due to their high levels of accuracy and relatively low cost. The earliest column (direct stress)-type load cells typically have the gauges bonded onto a structural member that deforms when a load is applied. In most applications, four strain gauges are used to obtain maximum sensitivity and temperature compensation, with two of them in the longitudinal direction and two in the Poisson direction, the so-called full Poisson bridge.

Introduction: This 'full bridge' configuration with a longitudinal gauge and a transverse 'Poisson' gauge on both the top and bottom surfaces is the most popular bridge arrangement for the measurement of axial loads and is used in many load cells. This bridge minimizes the effects of temperature change in the gauges.

Aim of the Experiment: To experimentally measure the axial strain due to tensile loading of a bar using a full Poisson bridge.

Theory: Figure 12.6a shows the bar with axial loading and all the four gauges on the bridge with their proper orientation. Figure 12.6b shows the Wheatstone bridge with the locations of all four active gauges.
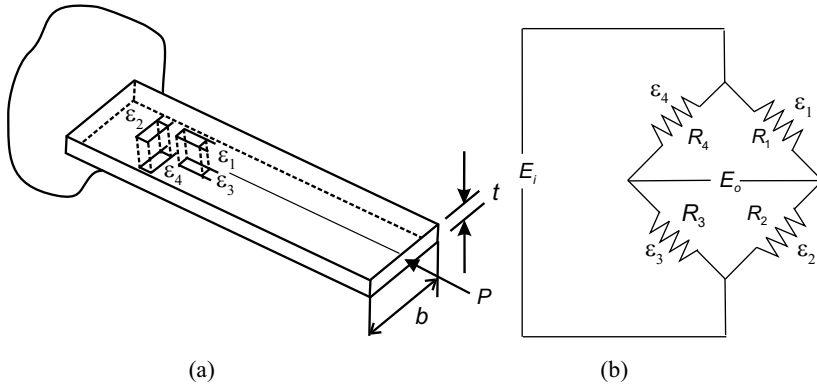
Fig. 12.6 Axial loading: **a** bar with load and strain gauges, **b** full Poisson bridge with gauges

With all four as active gauges

$$\varepsilon_1 = \varepsilon_3 = \frac{P}{Ebt}; \quad \varepsilon_2 = \varepsilon_4 = -\nu\varepsilon_1 \tag{12.13}$$

$$\frac{E_o}{E_i} = \frac{F\varepsilon_1(1+\nu)}{2 + F\varepsilon_1(1-\nu)} \tag{12.14a}$$

When $F = 2$

$$\frac{E_o}{E_i} = \frac{2\varepsilon_1(1+\nu)}{2 + 2\varepsilon_1(1-\nu)} \tag{12.14b}$$

For this configuration, constant K and non-linearity η in Eq. 12.5(a) are given by

$$K = \frac{F(1+\nu)}{2}; \quad \eta = \frac{F\varepsilon_1(1-\nu) \times 10^{-6}}{2 + F\varepsilon_1(1-\nu) \times 10^{-6}} \tag{12.15}$$

The output is not only higher by approximately a factor of $(1 + \nu)$ than for the quarter gauge version shown in Sect. 12.2, but is also less non-linear (approximately $[(1-\nu)/10]\%$ per $1000\,\mu\varepsilon$ produced by axial loads). Since all the four gauges experience the same temperature change, temperature compensation is inherently built into the bridge circuit. It is worth mentioning that both the gauges on a given surface are in adjacent arms of the bridge. If the applied force is not concentric to the bar, or if it is at an angle, there is a bending moment on the bar. To avoid errors due to bending strains, the strain gauges are fixed at the centres of the faces.

Test Setup and Procedure: The test setup and procedure are the same as described in Sect. 12.2. Table 12.1 may be used for consolidating the results.

Questions

(1) Why does this circuit reject the effect of bending strain?
(2) How is temperature compensation inherently built into this configuration?
(3) Why can you not have a full bridge with all four gauges in the longitudinal direction for measuring axial strain?
(4) What are the common resistance values in commercially available strain gauges?

12.6   Bending Strain Analysis of Cantilever Beam with Quarter Bridge

INTERESTING FACTS: Anyone on an aircraft experiencing light to moderate turbulence would have watched the wings flex up and down, bringing about a smooth ride for passengers. One of the most dramatic tests that manufacturers of aircraft subject them to is a wing load or flex test in order to assess how much a wing can bend before it breaks. An aeroplane in flight experiences a bending load on its wing caused by the aerodynamic lift trying to raise the wing. This lift force actually causes compression on the skin on the upper surface of the wing and tension on that on the bottom surface. On the other hand, in the case of the wings of an aeroplane sitting on its landing gear, the gravitational force causes the wing to bend downward, subjecting the skin at the bottom of the wing to compression and that at the top to tension. Hence, during flight certification, the wings of an aeroplane are intentionally bent up and down to ensure that they can take the stress without failing. For example, Boeing flexed the wings of the new 787 Dreamliner approximately 7.5 m which amounts to about 150% of the maximum flex that it would experience during normal flight conditions. It is even reported that passenger aircraft can endure a wing flex of almost 90°!

Introduction: When an external force is applied transverse to a structural element, it results in bending. Bending moment is a measure of the amount of bending that can occur and can be used to calculate where and how much bending may occur when a force is applied. The beam configuration with a single longitudinal gauge responds well to bending loads and is the simplest of all configurations for the measurement of bending strain.

Aim of the Experiment: To experimentally measure the bending strain due to transverse loading of a beam using a quarter bridge.

Theory: The specimen considered here is the same as before with the same dimensions, except that it has a transverse load P at a longitudinal distance L from the
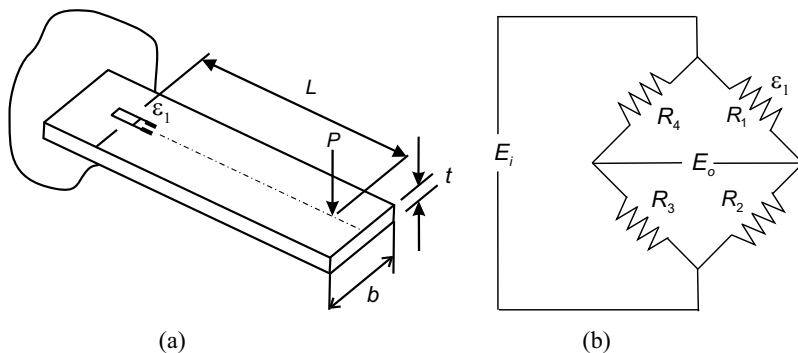
Fig. 12.7 Beam with transverse loading: **a** beam with load and strain gauge, **b** quarter bridge with strain gauge

longitudinal strain gauge. Figure 12.7a shows the transversely loaded beam with a single strain gauge, and Fig. 12.7b shows the gauge in a Wheatstone bridge.

From the point of view of mechanics, the strain at a point on the extreme fibre of the rectangular cross section of a cantilever beam is

$$\varepsilon_L = PL/EZ \tag{12.16}$$

where ε_L = longitudinal strain, P = transverse load, L = distance between the load and the gauge, E = modulus of elasticity, and Z = section modulus = $bt^2/6$ where b = beam width and t = thickness.

With R_1 as an active gauge,

$$\varepsilon_1 = \frac{6PL}{Ebt^2} \tag{12.17a}$$

$$\frac{E_o}{E_i} = \frac{F\varepsilon_1}{4 + 2F\varepsilon_1} \tag{12.17b}$$

With $F = 2$,

$$\frac{E_o}{E_i} \simeq \frac{\varepsilon_1}{2} \tag{12.17c}$$

The output in terms of mV/V may be written as shown in Eq. 12.5(a) where the constant $K = F/4$. It is seen that there is some non-linearity with a single active gauge, the non-linearity (which is roughly 0.1% per 1000 $\mu\varepsilon$ of strain) being the same as shown in Eq. 12.5(b). This non-linearity can be neglected at low strain levels. Care must be taken while applying the load, because the gauge will also respond to any axial loads present. If the gauge is mounted on the centre line, it is not affected by torsional loads. This configuration is sensitive to changes in temperature since bending moment PL, beam thickness and width (and hence section modulus), and modulus of elasticity are all temperature dependent; however, this change is very
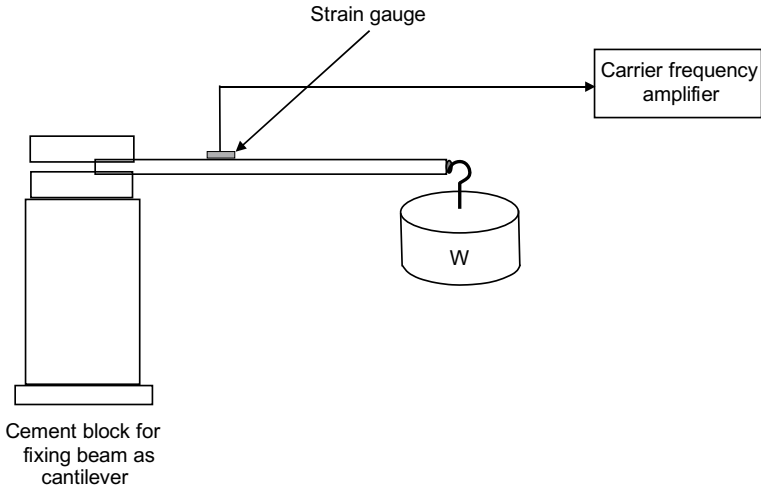
Fig. 12.8 Experimental setup for bending strain measurement on cantilever

small. Of all the variables affecting strain at a point, the only one unaffected by temperature change is the load. Since this quarter bridge configuration produces a small amount of non-linearity and is sensitive to changes in temperature, the half bridge configuration is generally preferred. When a half bridge cannot be used, a single active self-temperature-compensated strain gauge may be used; it also helps to zero-balance the bridge before the load is applied.

Test Setup and Procedure: The test setup is as shown in Fig. 12.8. It consists of a cantilever beam with a longitudinal gauge pasted on the top surface. There is a hook with provisions for hanging weights near the free end. The bridge is null-balanced without load. The beam is then loaded in steps by increasing the weight W hanging on the hook, and the strain at each stage is noted using the carrier frequency amplifier. Tabulations may be made as in Table 12.1. The analytically computed strains are compared with the measured values. The experiment is repeated changing the locations of the strain gauges longitudinally.

Questions

(1) How can you get large strains with a shorter beam?
(2) Why do some configurations use a beam with tapered sides?
(3) Which is the best location for pasting the strain gauge?
(4) How do self-temperature-compensated strain gauges work?

12.7 Measurement of Bending Strain: Half Bridge with Gauges in Adjacent Arms

> **INTERESTING FACTS:** Designers and engineers have to meticulously plan all aspects of bridge design and construction before it is built. More bridges fail during construction than while in service. This is because the loading during construction is higher than the operational loads. In 1907, a bridge failed during the construction phase in Quebec City in Canada. At a certain point in the construction process, designers were told that the bridge weighed around 4000 tonnes more than estimated. However, they did not believe this to be an issue significant enough to redesign the bridge. Soon, an engineer observed that the frame of the bridge had started bending. However, when he brought this to the notice of other engineers, they did not take it seriously, claiming that the beams were bent even when they were delivered. Soon, the bridge collapsed, killing 75 workers. A post-mortem revealed that the beams were not designed to handle the additional weight and had bent.

Introduction: The beam is the most common element that is subjected to bending and may bend at any point along its length. When the tensile stress due to the applied force is equal to or greater than the yield stress of the element, failure due to bending can occur.

Aim of the Experiment: To experimentally measure the bending strain due to transverse loading of a beam using a half bridge.

Theory: Figure 12.9a shows the transversely loaded beam with two strain gauges pasted as mirror images of each other in the longitudinal direction. Figure 12.9b shows the Wheatstone bridge configuration with gauges R_1 and R_2 connected in adjacent arms of the bridge.

With R_1 and R_2 as active gauges,

$$\varepsilon_1 = \frac{6PL}{Ebt^2}; \quad \varepsilon_2 = -\varepsilon_1 \tag{12.18}$$

$$\frac{E_o}{E_i} = \frac{F\varepsilon_1}{2} \tag{12.19a}$$

With $F = 2$

$$\frac{E_o}{E_i} = \varepsilon_1 \tag{12.19b}$$
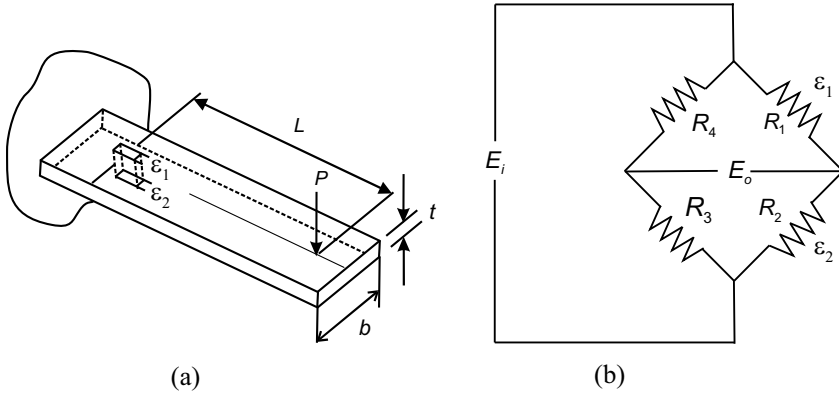
Fig. 12.9 Beam with transverse loading: **a** beam with load and strain gauges, **b** half bridge with gauges in adjacent arms

The constant $K = F/4$ and the non-linearity factor η is as shown in Eq. 12.5(b). The gauge on the lower surface is bonded exactly below that on the top surface; therefore, the gauges measure bending strains of equal magnitudes, but of opposite signs, and the bridge output is not just linear, but is double that produced by a single active gauge for the same loading conditions. Any changes in resistance in the active gauges produced by axial loads are due to strains of the same sign and will be cancelled since the two active gauges are in the adjacent arms of the Wheatstone bridge. Similarly, resistance changes due to temperature variations will get cancelled out since both the strain gauges and the specimen undergo the same thermal variations.

Test Setup and Procedure: The test setup and procedure are the same as indicated in Sect. 12.6.

Questions

(1) What will the error be if the two gauges are not pasted as mirror images of each other?
(2) Which is better for a non-homogeneous material like concrete, a short gauge or a long gauge?
(3) What length of strain gauge is preferred for localized strain measurement?
(4) How does the resistance of the lead wires affect the measured strain?
(5) When is this arrangement of gauges affected by torsion?

12.8   Measurement of Bending Strain: 1/2 Poisson Bridge

> **INTERESTING FACTS:** We all know that Poisson's ratio in elasticity is the ratio of transverse strain (in compression) to longitudinal strain (in tension). But did you know that the name of Siméon-Denis Poisson, the person after whom this ratio is named, is attached to many other ideas, as well, such as Poisson's equation in potential theory, Poisson's integral, Poisson's brackets in differential equations, Poisson's distribution in probability, and Poisson's constant in electricity? Poisson's father had wanted him to enter the medical profession, but as luck would have it for us, he was apparently extremely clumsy with his hands, but extremely good at theoretical work and mathematics. Some of Poisson's important contributions were the application of mathematics to electricity, magnetism, electrostatics, mechanics, and other areas of physics. Poisson also contributed to planetary theory by extending the work of Lagrange and Laplace on the stability of planetary orbits. He also studied applications of a number of physical problems using ordinary and partial differential equations such as the dynamics of a pendulum and the theory of sound. For people working in mechanics, his 'Traité de Mécanique' (1811 and 1833) or 'Treatise on Mechanics' was the standard book for many years.

Introduction: The half Poisson bridge is an extension of the quarter bridge for the measurement of bending strain and as can be expected has a higher sensitivity than the latter.

Aim of the Experiment: To experimentally measure the bending strain due to transverse loading of a beam using a half Poisson bridge.

Theory: This configuration for the measurement of strain due to a bending load P has one gauge R_1 aligned in the axial direction and the second gauge R_2 in the transverse direction, i.e., a Poisson gauge as shown in Fig. 12.10a. This may be implemented by pasting a 90° rosette on the top surface of the beam or by pasting one gauge in the longitudinal direction and one in the transverse direction. These gauges are connected in the adjacent arms of the bridge as shown in Fig. 12.10b.

With R_1 and R_2 as active gauges,

$$\varepsilon_1 = \frac{6PL}{Ebt^2}; \quad \varepsilon_2 = -\nu\varepsilon_1 \tag{12.20a}$$

$$\frac{E_o}{E_i} = \frac{R_1 + \Delta R_1}{R_1 + \Delta R_1 + R_2 + \Delta R_2} - \frac{R_4}{R_3 + R_4} = \frac{F\varepsilon_1(1 + \nu)}{4 + 2F\varepsilon_1(1 - \nu)} \tag{12.20b}$$
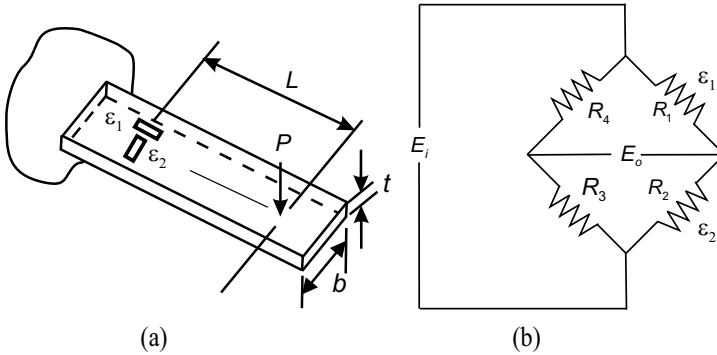
Fig. 12.10 Beam with transverse loading: **a** beam with load and strain gauges, **b** half Poisson bridge with gauges in adjacent arms

With $F = 2$

$$\frac{E_o}{E_i} = \frac{2\varepsilon_1(1 + v)}{4 + 4\varepsilon_1(1 - v)} \tag{12.20c}$$

The output in terms of mV/V may also be written as

$$\frac{E_o}{E_i} = \frac{2\varepsilon_1(1 + v) \times 10^{-3}}{4 + 2\varepsilon_1(1 - v) \times 10^{-6}} = K\varepsilon_1 \times 10^{-3}(1 - \eta) \tag{12.20d}$$

where

$$\eta = \frac{2\varepsilon_1(1 - v) \times 10^{-3}}{2 + 2\varepsilon_1(1 - v) \times 10^{-6}}$$

and

$$K = \frac{2(1 + v)}{4} \tag{12.20e}$$

Because the longitudinal and the transverse Poisson gauges are connected in the adjacent arms of the bridge, the bridge output is increased by a factor of approximately $(1 + v)$. Strains of opposite signs in adjacent arms have a numerically additive effect. This configuration does not give as high an output as the half bridge described in Sect. 12.7, but may be beneficial when there is no access to the bottom side and when improved output as compared to the quarter bridge is required. Resistance changes of thermal origins will get cancelled to some extent, as the beam as well as both the gauges experience like changes in temperature. The non-linearity is reduced to approximately $[(1 - v)/10]\%$ per $1000\,\mu\varepsilon$ of longitudinal strain.

Test Setup and Procedure: This is the same as described in Sect. 12.6.

Questions

(1) How do you ensure that axial strains are not sensed?

(2) Is the setup sensitive to torsional strains?

(3) How will a rosette be beneficial?

(4) If the gauges are not oriented as required, how much will the error be?

12.9 Bending Strain: Full Bridge

DID YOU KNOW that sensor spring elements that employ bending beam structural design are the most common in present-day transducers? This is because the bending beam is typically a high-strain, low-force structural member, as compared to a column of the same cross section. The bending beam design is used in lower capacity load cells. It is ideal for floor scales, hopper weighing, platform scales, weigh bridges, and truck weigh stations. There are also displacement transducers with strain gauges mounted on a cantilever. The simple structure and high sensitivity to displacement enable this transducer to make accurate and stable measurements in a small space.

Introduction: This four-gauge version is the most popular arrangement for a bending beam. This configuration yields the highest output for bending strain measurements.

Aim of the Experiment: To experimentally measure the bending strain due to transverse loading of a beam using a full bridge.

Theory: A beam with a cross section which is symmetrical about the bending axis offers two equal and opposite surfaces subjected to equal strains of opposite signs. Strain gauges bonded to these surfaces may be used in a full bridge circuit to offer the simplest beam configuration. This is not the best though and has led to designs which induce multiple bending in beam elements.

Figure 12.11a shows a beam with four identical gauges, two pasted on top and two at the bottom as mirror images in the axial direction. Figure 12.11b shows the Wheatstone bridge with all four active gauges.

With all four gauges as active gauges,

$$\varepsilon_1 = \varepsilon_3 = \frac{6PL}{Ebt^2}; \quad \varepsilon_2 = \varepsilon_4 = -\varepsilon_1 \tag{12.21}$$

$$\frac{E_o}{E_i} = F\varepsilon_1 \tag{12.22a}$$

(a)   (b)
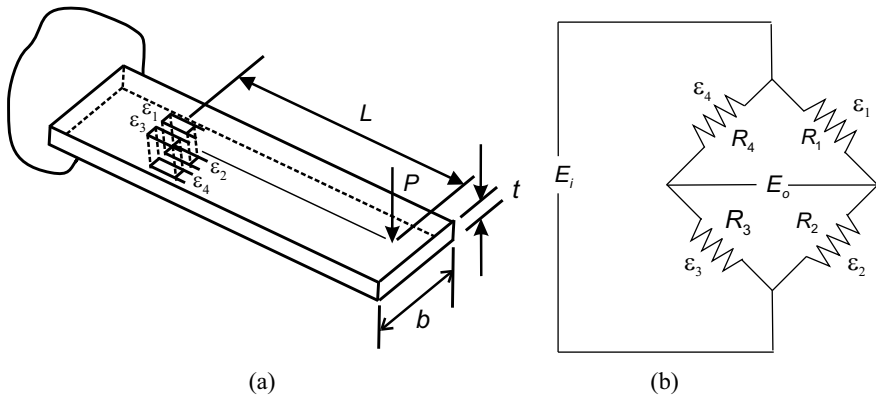
Fig. 12.11 Beam with transverse loading: **a** beam with load and strain gauges, **b** full bridge configuration

With $F = 2$

$$\frac{E_o}{E_i} = 2\varepsilon_1 \qquad\qquad (12.22b)$$

The output in terms of mV/V may also be written as

$$\frac{E_o}{E_i} = 2\varepsilon_1 \times 10^{-3} \qquad\qquad (12.22c)$$

For this configuration, $K = F$ and $\eta = 0$. The linear bridge output is twice that of the half bridge version described in Sect. 12.7. It may be noted that the two gauges on the top surface are in opposite arms of the Wheatstone bridge, and so also the two gauges on the bottom surface. This configuration gives the highest output among all configurations for the measurement of bending strain.

Test Setup and Procedure: This is the same as described in Sect. 12.6.

Questions

(1) Does this configuration respond to torsional strain?
(2) Can you somehow double the output of this configuration?
(3) How does this configuration reject axial strain?
(4) What can you say about the temperature compensation of this circuit?

12.10   Bending Strain: Full Poisson Bridge

> **INTERESTING FACTS:** All of us working in engineering mechanics have heard of the Timoshenko beam formulation which accounts for transverse shear deformation and rotational bending effects, making it capable of modelling thin or thick beams. But did you know that Prof. Stepan Prokofyevich Timofeyevich (1878–1972) or Stephen P. Timoshenko is considered as 'the father of applied mechanics'? So great was the widespread international influence of this renowned expert, teacher, and writer that his active years in the engineering field have come to be known as 'the Timoshenko era' both in the United States and in Russia. Timoshenko authored thirteen excellent textbooks and many papers in the areas of statics and dynamics, mechanics of materials, structural theory, elasticity, vibrations, stability, torsion and buckling, and plates and shells, which are widely studied even today. Much of the works he pioneered are now classical subjects that are taught as standard topics in engineering courses, and his textbooks have been translated into as many as thirty-five languages!

Introduction: This Poisson full bridge configuration measures only bending strain. It has four active strain gauges. Two of them are mounted longitudinally in the direction of bending strain, with one gauge on the top surface of the beam and the other on the bottom surface as a mirror image of the first one. Two other gauges are pasted in the transverse direction as Poisson gauges, one on each surface of the specimen and as a mirror image of each other. This configuration has the advantage that it rejects axial strain and compensates for temperature variations.

Aim of the Experiment: To experimentally measure the bending strain due to transverse loading of a beam using a full Poisson bridge.

Theory: Figure 12.12a shows the test arrangement with strain gauges and Fig. 12.12b the arrangement of gauges in the Wheatstone bridge.

With all four gauges as active gauges,

$$\varepsilon_1 = \frac{6PL}{Ebt^2}; \quad \varepsilon_2 = -\varepsilon_1; \quad \varepsilon_4 = -\nu\varepsilon_1; \quad \varepsilon_3 = \nu\varepsilon_1 \tag{12.23}$$

$$\frac{E_o}{E_i} = \frac{F(1+\nu)\varepsilon_1}{2} \tag{12.24a}$$
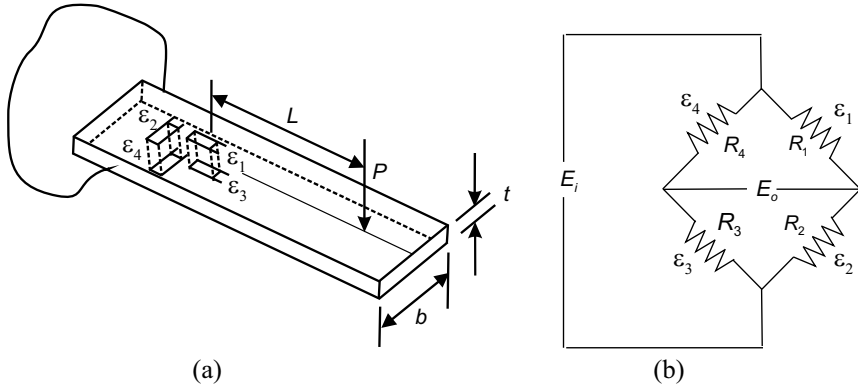
Fig. 12.12 Beam with transverse loading: **a** beam with load and strain gauges, **b** full Poisson bridge configuration

The output in terms of mV/V is

$$\frac{E_o}{E_i} = \frac{F(1 + v)\varepsilon_1 \times 10^{-3}}{2} \tag{12.24b}$$

With a gauge factor of 2, this becomes

$$\frac{E_o}{E_i} = (1 + v)\varepsilon_1 \tag{12.24c}$$

This version may not be as popular as the earlier one in Sect. 12.9, but is better than the half bridge version in Sect. 12.8.

Test Setup and Procedure: The setup is the same as shown in Fig. 12.8 in Sect. 12.6. The procedure and tabulation are also identical to those given earlier.

Questions

(1) Does this arrangement have any advantage as compared to the full bridge in Sect. 12.9?
(2) Is this arrangement affected by axial strains?
(3) Comment on the effects of temperature on this arrangement of gauges.
(4) Is this arrangement affected by torsion?
(5) Is resistance matching of the active gauges important while selecting them?

12.11   Torsional Strain Measurement Using Half Bridge

INTERESTING FACTS: Simple tools like a screwdriver develop a twist and torsional stress when a screw is driven into the wood; likewise, a torque wrench used to tighten a nut on a bolt applies a torque to the bolt causing it to twist and subjecting it to internal torsional stresses. When the torsional stresses developed exceed the shear strength of the material, failure occurs. Another example of torsion is that of a spark plug which must be loosened using a swivelled socket wrench. Yet another example is that of the steering wheel of a car which helps in manoeuvering or turning it. If the radius of the steering wheel from the pivot point is increased, the amount of force required for applying the torque reduces and facilitates the driver to turn the steering wheel.

Introduction: The half bridge arrangement used for the measurement of bending strains as shown in Fig. 12.9b is used here also. The major difference is in the orientation of the strain gauges on the rod.

Aim of the Experiment: To find the torsional strain in a rod subjected to torsion using a half bridge.

Theory: Two active strain gauges R_1 and R_2 are pasted at 45° to the longitudinal axis on the top surface of a rod of radius R subjected to a torque T as shown in Fig. 12.13a. Figure 12.13b shows the location of the gauges on the periphery of the rod, and Fig. 12.13c shows the gauges connected to a Wheatstone half bridge circuit to sense the torsional strains induced due to the application of the torque.
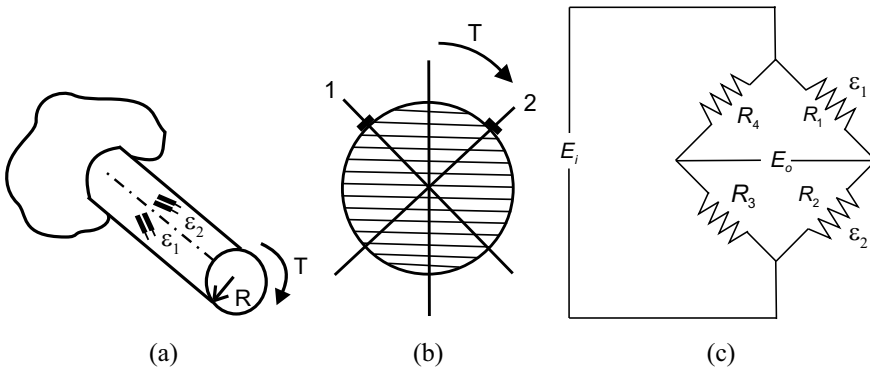
Fig. 12.13 Strain with torque loading: **a** rod with torque and strain gauges, **b** location of gauges on the circumference, and **c** half bridge configuration

For this configuration,

$$\text{Shear strain } \gamma_{\text{max}} = |\varepsilon_1 - \varepsilon_2| = 2\varepsilon_1 \qquad (12.25\text{a})$$

Here, gauges ε_1 and ε_2 are oriented at 45° to the longitudinal axis.

$$\text{Shear strain } \gamma_{\text{max}} = \frac{\tau}{G} = \frac{\text{shear stress}}{\text{modulus of rigidity}} \qquad (12.25\text{b})$$

where

$$\tau = \frac{TR}{I_p} \qquad (12.25\text{c})$$

I_p is the polar moment of inertia

$$I_p = \frac{\pi D^4}{32} = \frac{\pi R^4}{2} \qquad (12.25\text{d})$$

Here, τ is torsional stress, T is applied torque, and R is distance from centre to outer fibre. Substituting for I_p

$$\tau = \frac{2T}{\pi R^3} \qquad (12.25\text{e})$$

$$G = \frac{E}{2(1 + \nu)} \qquad (12.25\text{f})$$

Substituting for τ from Eq. (12.25c) and G from Eq. (12.25f) into Eq. (12.25b)

$$\gamma_{\text{max}} = \frac{\tau}{G} = \frac{2\tau(1 + \nu)}{E} = \frac{4T}{\pi E R^3}(1 + \nu) \qquad (12.26)$$

$$\frac{E_o}{E_i} = \frac{R_1 + \Delta R_1}{R_1 + \Delta R_1 + R_2 + \Delta R_2} - \frac{R_4}{R_3 + R4} \qquad (12.27\text{a})$$

Simplifying,

$$\frac{E_o}{E_i} = \frac{F\gamma_{\text{max}}}{4} = \frac{F\varepsilon_1}{2} \qquad (12.27\text{b})$$

Test Setup and Measurements: Figure 12.14 shows a circular brass rod of radius R subjected to a torque T. The slender brass rod is rigidly fixed to the clamp shown at the right end and is attached to the mandrel of a pulley at the left end. A wire passing over the pulley of radius r carries a hook with a weight W at the end, applying a torque $T = Wr$ on the rod. Two torque gauges oriented as shown in Fig. 12.13a, b are bonded onto the rod as shown in Fig. 12.14. Before the start of the test, the maximum applicable load that can safely be applied without breaking the rod can be computed assuming the safe shear stress in brass to be 75 MPa. The load may
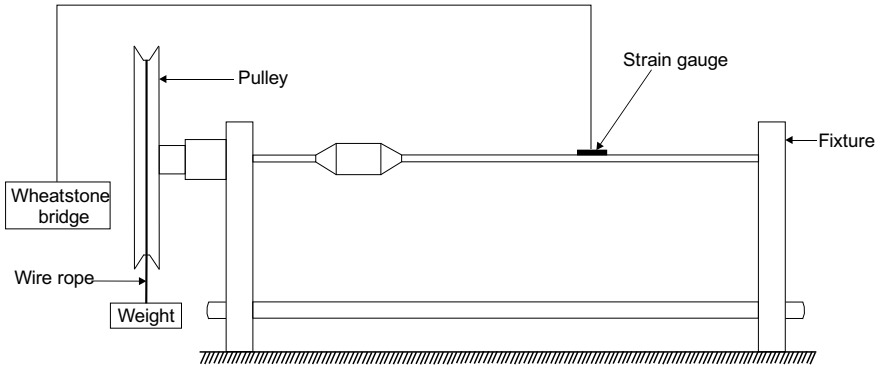
Fig. 12.14   Test setup for torsional strain measurement

Table 12.2   Table for torsional strain measurement

Sl. No.	Weight W, N	Torque, Nm	O/p voltage, V	Strain $\mu\varepsilon$	
1.					
2.					
3.					

be increased in steps starting from zero to increase the torque and the bridge output may be noted for each torque. Readings may be tabulated as shown in Table 12.2. A graph of torque versus torsional strain may be plotted.

Questions

(1) Comment on the temperature compensation of this configuration.
(2) What are the major sources of errors in the measured stress?
(3) Are the above equations applicable for composite materials? Why?
(4) What are the precautions to be taken in the above experiment?

12.12   Torsional Strain Measurement with Full Bridge

DID YOU KNOW that it is not uncommon for large motors to suddenly stop running due to excessive torsional stress and a cracked shaft? This is because torsional loads induce severe twisting of shafts. Most rotating machineries experience torsional vibration during coast up or coast down or during continuous operation. The only indication of severe torsional vibration is often gear

noise or coupling wear. Torsional vibration of large amplitudes may cause gear wear, gear tooth failures, shrink fit slippage, key failures, and broken shafts in very severe cases. There is torsional vibration in rotating machines such as synchronous and induction motor-driven machines, reciprocating machines, long propeller drive rotors in ships, machines operating under variable torques, and in those transmitting high power. Torsional vibrations result not only from direct torsional oscillations, but also from common lateral mode excitations due to torsional/lateral mode coupling. Typical examples are turbines, shafts in engines, transmissions, torque converters, electric motors, and similar devices.

Introduction: A full bridge circuit with four active gauges may be used to obtain twice the sensitivity of the half bridge configuration for torsional strain measurement.

Aim of the Experiment: To find the torsional strain in a rod subjected to torsion using a full bridge.

Theory: A full bridge circuit with four strain gauges as shown in Fig. 12.15a, b is used. Gauges 3 and 4 are located on the bottom side and are pasted as mirror images of gauges 1 and 2 pasted on the top. Figure 12.15c shows the four active gauges in the bridge.

$$\text{Shear strain } \gamma_{\max} = |\varepsilon_1 - \varepsilon_2| = |\varepsilon_3 - \varepsilon_4| = 2\varepsilon_1 \tag{12.28}$$

$$\frac{E_o}{E_i} = \frac{R_1 + \Delta R_1}{R_1 + \Delta R_1 + R_2 + \Delta R_2} - \frac{R_4 + \Delta R_4}{R_3 + \Delta R_3 + R_4 + \Delta R_4} \tag{12.29a}$$
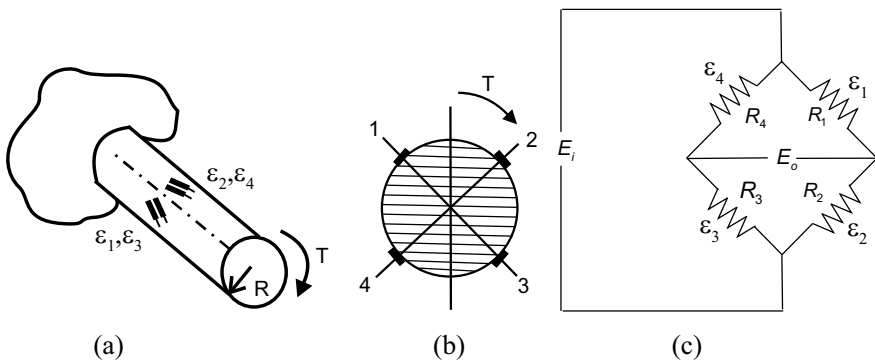
Fig. 12.15 Rod with torque loading: **a** rod with torque and strain gauges, **b** location of gauges on the circumference, and **c** full bridge configuration

$$\frac{E_o}{E_i} = \frac{F\gamma_{\max}}{2} = F\varepsilon_1 \qquad (12.29b)$$

As in the case of the full bridge configuration for bending loads, this torsional full bridge arrangement has a linear output and good temperature compensation. The effects of both axial and bending loads are cancelled in this configuration for torque measurement. However, very accurate gauge orientation and placement of all four gauges are very important for correct results.

Measurement Setup: The test setup as shown in Fig. 12.14 may be used here also for the determination of torsional strains, except that four gauges now have to be bonded to sense torsional strain. The load may be increased in steps as before to increase the torque and the bridge output may be noted for each case. Table 12.2 may be used for tabulating the readings.

Questions

(1) How do you ensure the accurate placing of the gauges?
(2) How do you bring out leads from strain gauges pasted on a rotating shaft to the external carrier frequency amplifier?
(3) How do you obtain shear modulus G?

12.13 Determination of Longitudinal and Hoop Stresses and Fluid Pressure in a Thin-Walled Pressure Vessel with Gauges Pasted Along the Principal Axes

INTERESTING FACTS: Pressure vessels are containers meant for holding liquids, vapours, or gases at high pressures, examples being vessels in the petroleum refining and chemical processing industries, boilers and heat exchangers, storage tanks, and cylinders for diving and in medicine. Working pressures of cylinders have gone up from 150 bars in 1950 to around 200 bars in 1975 and 300 bars in 1995 to even 4000 bars today. Each pressure vessel has its own design pressure and design temperature, operation beyond which could damage the equipment, leading to loss of fluid or catastrophic failure. These vessels are most commonly in the form of spheres, cylinders, or cones. In any pressure vessel subjected to internal pressure, triaxial stresses are set up in the shell wall. The three principal stresses are longitudinal stress, circumferential or hoop stress, and radial stress. Besides, there may be bending and shear stresses, all of which need to be measured.

Introduction: For the present experiment, a cylindrical pressure vessel in the form of an aluminium beverage can is considered, though in reality thin-walled pressure vessels may be either spherical or cylindrical.

Aim: To measure the longitudinal and hoop stresses in a thin-walled pressure vessel when pressure is released suddenly and to compute the original fluid pressure which existed.

Theory: A pressure vessel may be assumed to be thin-walled if the ratio of its inside radius R to its wall thickness t is greater than 10.

$$\frac{R}{t} > 10 \tag{12.30}$$

Such a condition greatly simplifies the analysis of pressure vessels. The two principal stresses in a cylindrical pressure vessel are σ_l in the longitudinal or axial direction and σ_h in the circumferential direction, also known as hoop stress. From basic engineering mechanics, the values of these stresses in terms of the hydrostatic pressure, p, within the cylinder producing the stresses are given by

$$\sigma_h = \frac{pR}{t} \tag{12.31a}$$

$$\sigma_l = \frac{pR}{2t} \tag{12.31b}$$

It is evident that

$$\sigma_h = 2\sigma_l \tag{12.31c}$$

It is assumed that the material out of which the cylinder is made is compositionally homogeneous and mechanically isotropic. It is also assumed that the test is conducted within the purely elastic range of the material. The air pressure on the outer surface of the cylinder is ignored. With the biaxial state of stress assumed for the cylinder, Hooke's law gives the following equations as derived in Sect. 6.4.

$$\sigma_h = \frac{E(\varepsilon_h + \nu\varepsilon_l)}{(1 - \nu^2)} \tag{12.32a}$$

$$\sigma_l = \frac{E(\varepsilon_l + \nu\varepsilon_h)}{(1 - \nu^2)} \tag{12.32b}$$

where E is the modulus of elasticity of the material of the can, ν is Poisson's ratio, and ε_h and ε_l are the circumferential and longitudinal strains (obtained from the raw strain readings after correcting for the effect of transverse sensitivity of the gauges) corresponding to the circumferential and longitudinal stresses σ_h and σ_l. Both the stresses can be computed as shown below.
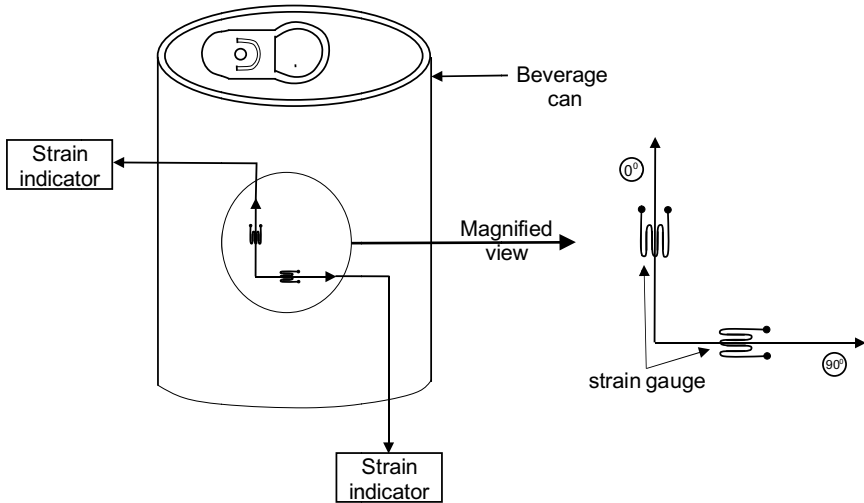
Fig. 12.16 Beverage can with strain gauges

$$\sigma_h = \frac{2E\varepsilon_h}{2 - \nu} \tag{12.33a}$$

$$\sigma_l = \frac{E\varepsilon_l}{(1 - 2\nu)} \tag{12.33b}$$

Finally, the internal pressure in the can may be computed as

$$p = \frac{2tE\varepsilon_l}{R(1 - 2\nu)} = \frac{2tE\varepsilon_h}{R(2 - \nu)} \tag{12.34}$$

Test Setup and Procedure: An aluminium beverage as shown in Fig. 12.16 can be used as an inexpensive thin-walled pressure vessel for this experiment. The length and diameter of the can may be measured using calipers and the thickness using a micrometer. Two linear gauges may be pasted on the can at its longitudinal centre, one in the longitudinal direction and the other in the circumferential direction. Alternately, a two-element 90° strain gauge rosette with gauges oriented at 0° and 90° may be pasted such that the gauge at 0° is aligned in the longitudinal direction to measure axial strain and the one at 90° is aligned to measure the circumferential strain. Each of the two individual strain gauges is to be connected in a quarter bridge configuration to a separate strain indicator. Initially, the bridges are balanced. The can has to be shaken well and the lid has to be popped open suddenly to depressurize the can. The output of each quarter strain gauge bridge has to be recorded. The results may be tabulated as shown in Table 12.3. It may be noted that the longitudinal strain is lower than the hoop strain.

Table 12.3 Table for torsional strain measurement

Case	Gauge		
	Longitudinal Strain ε_l (μm/m)	Circumferential ε_h (μm/m)	Pressure p (N/m^2)
Pressurized			
Depressurized			

Questions

(1) Why do you get two different values for the hydrostatic pressure when you compute from the longitudinal and circumferential strains?
(2) The cylinder is not infinitely long. What difference does it make to the measured pressure?
(3) Do the different geometries at the top and bottom make a difference to the measured pressure?
(4) Does the measured thickness of the can justify the assumption of a thin-walled cylinder?
(5) Are the assumptions of the can being homogeneous and isotropic justified?

12.14 Determination of Longitudinal and Hoop Stresses and Fluid Pressure in a Thin-Walled Pressure Vessel with Gauges Pasted with Arbitrary Orientation: Biaxial Stress

DID YOU KNOW that it is possible to do controlled planar biaxial tests by applying a controlled tension–tension biaxial state of stress and strain to flat specimens? Such tests help characterize the anisotropic behaviour of the material of the specimens and provide inputs for its constitutive modelling. Experimental planar biaxial systems can be configured for static, dynamic, and fatigue testing applications. Both the implementation of the experimental loading and the interpretation of the results are challenging in the case of biaxial tension, as compared to uniaxial tension. Typical specimens on which biaxial tension testing is conducted are tissues, including skin, cardiac, aortic, lung, tendon, and ligaments. The biaxial stresses in the warp and weft directions of airbag fabrics are found out by the application of tensile forces in both these directions. Inflation tests are also carried out on tubular or cylindrical specimens such as pressure vessels clamped in rigid, circular jaws. In these tests, hydrostatic pressure in the tube/cylinder can create longitudinal and circumferential stresses.

Introduction: The experiment is conducted as in Sect. 12.13, but this time a three-element rosette is pasted with arbitrary orientation. It is assumed that the stress fields and the directions of principal stresses are not known; it is known that three strain gauge elements are needed to determine the state of strain if the gauges are not aligned with the principal axes.

Aim: To measure the principal stresses in a thin-walled pressure vessel when pressure is released.

Theory: Given the three strain gauge readings from a rectangular rosette with three strain gauges placed at $0°, 45°$, and $90°$, it is possible to calculate the principal strains. If gauges 'a' and 'c' are tilted to the circumferential (x) and longitudinal (y) directions, respectively, as in Fig. 12.17, then the equations for the strains are

$$\varepsilon_a = \varepsilon_x; \quad \varepsilon_b = \frac{1}{2}(\varepsilon_x + \varepsilon_y + \gamma_{xy}); \quad \varepsilon_c = \varepsilon_y \quad \text{and} \quad \gamma_{xy} = (2\varepsilon_b - \varepsilon_a - \varepsilon_c)$$

$$(12.35a)$$

If the gauges are oriented as in Fig. 12.17, then the principal strains are

$$\varepsilon_1 = \frac{1}{2}(\varepsilon_a + \varepsilon_c) + \frac{1}{2}\sqrt{(\varepsilon_a - \varepsilon_c)^2 + (2\varepsilon_b - \varepsilon_a - \varepsilon_c)^2} \qquad (12.35b)$$

$$\varepsilon_2 = \frac{1}{2}(\varepsilon_a + \varepsilon_c) - \frac{1}{2}\sqrt{(\varepsilon_a - \varepsilon_c)^2 + (2\varepsilon_b - \varepsilon_a - \varepsilon_c)^2} \qquad (12.35c)$$

where ε_1 and ε_2 are the two in-plane principal strains. The principal angle ϕ between the 'a'-axis and the axis corresponding to maximum principal strain ε_1 may be obtained as

$$\tan(2\phi) = \frac{2\varepsilon_b - \varepsilon_a - \varepsilon_c}{\varepsilon_a - \varepsilon_c} \qquad (12.36)$$

With the assumption that the material of the test specimen is linear, isotropic, and homogeneous, the principal stresses can be obtained from the strains as shown below:

$$\sigma_1 = E\left[\frac{(\varepsilon_a + \varepsilon_c)}{2(1-\nu)} + \frac{\sqrt{(\varepsilon_a - \varepsilon_c)^2 + (2\varepsilon_b - \varepsilon_a - \varepsilon_c)^2}}{2(1+\nu)}\right] \qquad (12.37a)$$

$$\sigma_2 = E\left[\frac{(\varepsilon_a + \varepsilon_c)}{2(1-\nu)} - \frac{\sqrt{(\varepsilon_a - \varepsilon_c)^2 + (2\varepsilon_b - \varepsilon_a - \varepsilon_c)^2}}{2(1+\nu)}\right] \qquad (12.37b)$$

where E is the modulus of elasticity of the material and ν is its Poisson's ratio.

Test Setup and Procedure: The three-element strain gauge rosette with gauges oriented at $0°, 45°$, and $90°$ should be pasted on the can at its longitudinal centre with
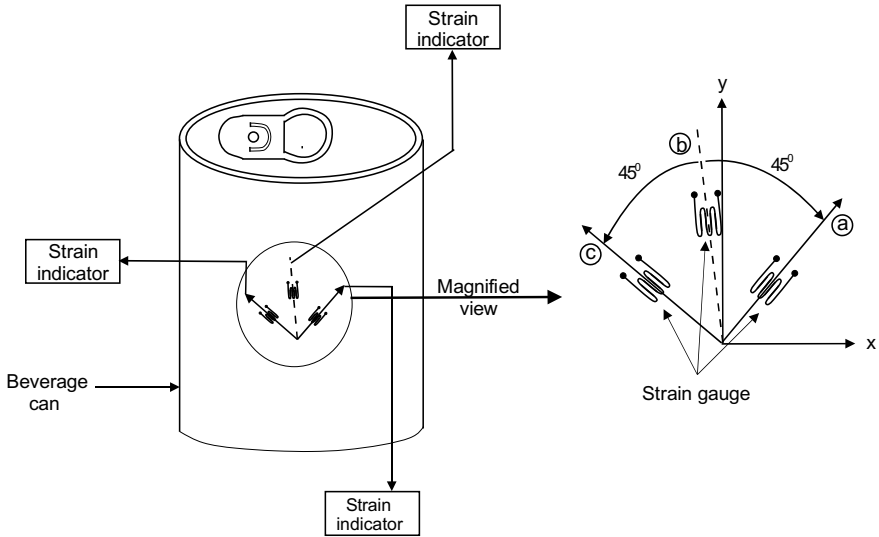
Fig. 12.17 Beverage can with a three-element strain gauge oriented arbitrarily

Table 12.4 Strain readings

Case	Quantity							
	ε_a	ε_b	ε_c	ε_1	ε_2	ϕ°	σ_1	σ_2
	(μm/m)						N/m^2	
Pressurized								
Depressurized								

arbitrary orientations not aligned with the principal axes as shown in Fig. 12.17. Each of the three individual strain gauges is to be connected to a separate strain indicator connected in a quarter bridge configuration as before. Initially, the bridge is balanced in the pressurized condition, then the lid of the can has to be popped open suddenly after shaking it well. The output of each quarter bridge has to be recorded and the results tabulated as shown in Table 12.4.

Questions

(1) Is choice of the material of the strain gauge important?
(2) How will humidity affect the measured strains?
(3) What are stacked and planar rosettes?
(4) Does a stacked rosette have any advantages over a planar rosette?
(5) Calculate the initial pressure in the can.

12.15   Determination of Young's Modulus of Elasticity and Poisson's Ratio Using Strain Gauges

DID YOU KNOW that Thomas Young, who is credited with Young's modulus of elasticity in mechanics, the Young–Laplace and Young–Dupré equations in fluid mechanics, the Young temperament, a method of tuning musical instruments, and Young's rule to calculate the infant dose of a drug, was actually a medical professional? He was a child prodigy, and by the time he left school at the age of 13, he was knowledgeable in many languages and had also acquired a good foundation in Newtonian physics, had studied optics, and had made several instruments in school. He was a polymath and scientific investigator whose articles for the Encyclopaedia Britannica spanned 20 fields of knowledge. He is described in the literature as a man who knew too much. For someone as talented as Young, he received relatively few honours.

Introduction: We all know that elastic modulus is an intrinsic material property that is extensively used in engineering design. Accurate knowledge of this property is very important for design, finite element modelling, and experimental validation. It is also required for obtaining reliable fits to the constitutive equations for the stress–strain curve. There are many test methods available for measuring modulus, but due to doubts within the engineering community of the accuracy of these methods, many engineers use standard handbook values in their calculations and designs. Accurate values can be got through good experimental methods and careful measurement. Most of the standards discuss the use of extensometers; however, strain gauges offer better results than extensometers in many respects in modulus testing. For the highest possible accuracy, a Class 0.2 averaging high-resolution extensometer is recommended. Unfortunately, such an extensometer is not widely available, nor do many laboratories have the equipment to calibrate these devices and this is where the strain gauge-based method comes in handy. ASTM E132-17 gives the standard method for the measurement of Poisson's ratio at room temperature.

Aim: The aim of this experiment is to find Young's modulus of elasticity and Poisson's ratio using an aluminium bar. In this experimental method, the value of Poisson's ratio is obtained from strains resulting from uniaxial stress only.

12.15.1   Determination of Poisson's Ratio

Theory, Instrumentation, and Test Procedure: It is known that when a bar of isotropic material is subjected to uniaxial stress, it deforms in the direction of the stress, as well as in the perpendicular direction, but with opposite sign. The axial
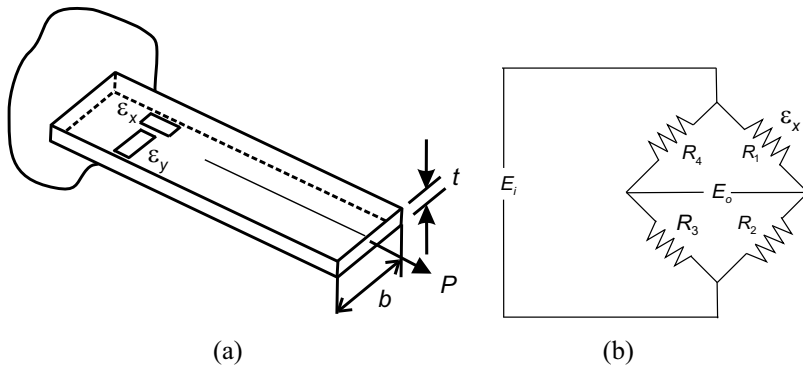
Fig. 12.18 Axial loading: **a** bar with load and strain gauges, **b** quarter bridge for longitudinal configuation

loading setup as shown in Fig. 12.2 or 12.3 may be used for the measurement of Poisson's ratio.

When an axial load P is applied to the bar, it expands longitudinally, while contracting laterally. Poisson's ratio is defined as the absolute value of the ratio of transverse strain ε_y to the axial strain ε_x. This ratio is, however, valid only in the linear range of the stress–strain curve.

$$\nu = \left| \frac{\varepsilon_y}{\varepsilon_x} \right| \tag{12.38}$$

Two linear gauges may be pasted on the bar at its longitudinal centre, one in the longitudinal direction and the other in the transverse direction. Figure 12.18a shows the arrangement for the measurement of longitudinal strain, as also the bridge circuit for the same. A similar arrangement may be used for the measurement of lateral strain, except that the gauge is now oriented as a Poisson gauge as shown in Fig. 12.18a. Each of the two individual strain gauges is to be connected in a quarter bridge configuration as shown in Fig. 12.18b to a separate strain indicator. Alternately, a two-element 90° strain gauge rosette with gauges oriented in the longitudinal and transverse directions may also be used, if available, as shown in Fig. 12.18a; here again, each gauge has to be connected to a separate strain indicator. The measured strains in the longitudinal and lateral directions are given by Eq. (12.39).

$$\varepsilon_x = \frac{P}{Ebt}; \quad \varepsilon_y = -\nu\varepsilon_x \tag{12.39}$$

Initially, the bridges are to be balanced. Then the bar is to be loaded in steps, and the longitudinal and transverse strain readings are to be noted as shown in Table 12.5. For most metals, Poisson's ratio is very close to 0.3.

Questions

Table 12.5   Strain readings

Sl. No.	Load P, N	ε_x, μm/m	ε_y, μm/m	ν
1.				
2.				
3.				

(1) How sensitive do you think this method is to the values of the measured forces?
(2) How can you increase the sensitivity of the method?
(3) What is the effect of the transverse sensitivity of the gauges used?
(4) Can this method be used for materials that are not homogeneous and isotropic?
(5) What if the material is not linearly elastic?

12.15.2   Determination of Young's Modulus of Elasticity

In order to obtain Young's modulus of elasticity, two methods may be used:

Method 1: The bar is loaded axially and the axial strain is measured for each load.

Theory, Instrumentation, and Test Procedure: The test setup as shown in Fig. 12.2 or 12.3 may be used for the experiment. Figure 12.4a, b shows the axially loaded bar and strain gauges and the corresponding half bridge configuration, respectively. Using the theory highlighted in Sect. 12.3, the following formula can be used to calculate E. We know that

$$E = \frac{\sigma}{\varepsilon_1} = \frac{P}{\varepsilon_1 bt} \tag{12.40a}$$

Also, the output voltage divided by the input voltage in terms of mV/V with strain being expressed as microstrain is

$$\frac{E_o}{E_i} \simeq \varepsilon_1 \times 10^{-3} \tag{12.40b}$$

The dimensions of the aluminium bar should be noted; the applied load and strain are to be measured in steps. Table 12.6 shows how the readings may be arranged. A linear fit may be made between the obtained load and strain values.

Method 2:

Theory, Instrumentation, and Test Procedure: The beam is subjected to a transverse load, and the bending strain is measured for each load. For this method, the

Table 12.6   Typical measurements

Sl. No.	Load P, N	ε_1, μm/m	E, N/m^2
1.			
2.			
3.			

test setup as shown in Fig. 12.11 with the beam subjected to transverse loading may be used. Figure 12.11a, b shows the beam with load and strain gauges and the corresponding full bridge configuration, respectively. Using the theory highlighted in Sect. 12.9, the following formula can be used to calculate E.

$$E = \frac{\sigma}{\varepsilon_1} = \frac{6PL}{\varepsilon_1 bt^2} \tag{12.41a}$$

The output in terms of mV/V is

$$\frac{E_o}{E_i} = 2\varepsilon_1 \times 10^{-3} \tag{12.41b}$$

A table similar to Table 12.6 may be used, except that the load in now transverse.

Questions

(1) How can you increase the sensitivity in this experiment?
(2) Compare the results obtained by the two methods for finding the modulus of elasticity.
(3) Which method gives a larger output for the same load?
(4) What are the main sources of error in these methods?

12.16   Three-Point Bending Experiment for Determination of Flexural Modulus of Elasticity

INTERESTING FACTS: Modulus of elasticity is a very important property of a material. It can be reported as (i) tensile modulus from tensile testing, (ii) compressive modulus of elasticity from compression tests, (iii) flexural modulus of elasticity from flexural tests, or (iv) torsional modulus of elasticity from torsion tests. The product designer should be clear about which modulus is to be used and this depends on the application. The flexural strength of a material is defined as its ability to resist deformation under load. Many machines and structures have parts which are primarily intended to carry loads that may cause bending. Examples of such parts are beams, plates, and columns. Flexural modulus is also used as a measure of the strength of adhesives, laminates, and polymers. Many plastics often have their flexural modulus (obtained from a bending test) listed, but not their elastic modulus (obtained from a tensile test).

Fig. 12.19 Schematic
drawing of three-point
bending
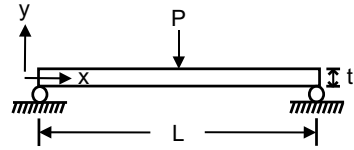
Introduction: Flexural modulus is a measure of the stiffness during the first or initial step of the bending process and is represented by the slope of the initial straight-line portion of the flexural stress–flexural strain curve. The bending test helps determine the limits of structural beams of various shapes and sizes. For materials that deform to a large extent, but do not break, the load at yield, typically measured at 5% deformation/strain of the outer surface, is reported as the flexural strength or flexural yield strength. For ASTM D790 standard for determination of structural strength, a specimen of size $0.3175 \, \text{cm} \times 1.27 \, \text{cm} \times 12.7 \, \text{cm}$ is placed on two supports and a load is applied at the centre. The analogous ISO equivalent is ISO 178. Bending tests on beams are usually made to determine strength and stiffness in bending. A uniform slender beam will generally fail when the tensile strength of the material is exceeded. Any beam may also fail prematurely due to defects in the material as in wooden beams with knots or cracks. Defects near the neutral axis of a beam will, in general, have very little effect on its strength.

Objectives: The objectives of this experiment are to obtain the flexural modulii of elasticity for small beams of different materials, all with the same length and rectangular cross section.

Theory: Consider a beam of rectangular cross section with supported span L, width b, and thickness t (Fig. 12.19). In a three-point bending test, the fibres close to the top side of the beam are in compression and the fibres close to the bottom side of the beam are in tension, with the outer fibres being subjected to maximum stress and strain. Failure will occur when the strain or elongation exceeds the material's limits.

The moment of inertia of the cross section of the beam about its longitudinal axis is given by

$$I = \frac{bt^3}{12} \tag{12.42}$$

Flexural stress in the outer fibres at the midpoint of the beam is given by

$$\sigma_f = \frac{My}{I} \tag{12.43a}$$

where

$$M = \frac{PL}{4} \tag{12.43b}$$

Here, P is the load at a given point on the load–deflection curve. Substituting for I from Eq. (12.42) and for M from Eq. (12.43b), the expression for flexural stress in the outer fibres at the midpoint reduces to

$$\sigma_f = \frac{3PL}{2bt^2} \qquad\qquad (12.44)$$

Flexural strain on the outer surface is given by

$$\varepsilon_f = \frac{6\delta_{\text{max}}t}{L^2} \qquad\qquad (12.45a)$$

where δ_{max} is the maximum deflection of the centre of the beam and is given by

$$\delta_{\text{max}} = \frac{PL^3}{48EI} \qquad\qquad (12.45b)$$

Flexural modulus of elasticity is given by

$$E_f = \frac{L^3 m}{4bt^3} \qquad\qquad (12.46a)$$

The measured value of E_f is given by

$$E_f = \frac{P}{\delta_{\text{max}}} \qquad\qquad (12.46b)$$

Here, m is the gradient or slope of the initial straight-line portion of the load–deflection curve.

Test Setup and Procedure: The 'three-point bending' experiment is usually conducted by using a specifically designed test fixture on a universal testing machine. The test sample is placed on two supporting pins spaced a known distance apart (Fig. 12.20). A third loading pin is attached to the lower end of the mobile traverse and is lowered onto the test sample from above at a constant rate. The position of the loading pin is measured, as this gives the deflection of the beam. The lowering is done in discrete steps till the sample fails. At each step, the load and deflection are noted. The load at yield is the sample material's flexural strength. Results may be tabulated as shown in Table 12.7. Graphs of force versus measured deflection and force versus calculated deflection are to be drawn and compared. This is done by connecting the outputs of the force sensor and the displacement output of the mobile traverse to a personal computer. From these graphs, E_f may be found out. It is to be borne in mind that the results of the testing method are sensitive to the specimen and loading geometry, as well as strain rate.
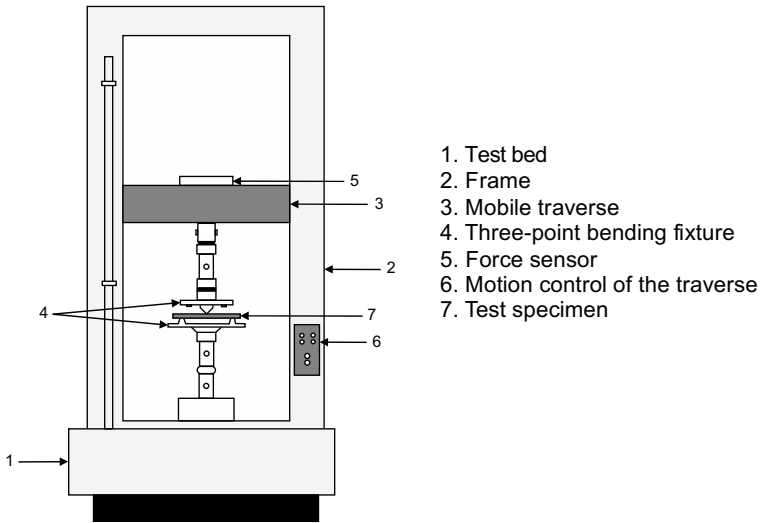
Fig. 12.20 Test setup for three-point bending test

Table 12.7 Load and deflection

Sl. No.	Load P, N	Deflection δ, mm
1.		
2.		
3.		

Questions

(1) How does a rectangular cross section help resist bending?

(2) Where, along the length of the beam, does the maximum deflection occur?

(3) What is the maximum value of the bending moment in terms of P and L?

(4) Draw the shear force and bending moment diagrams for the given loading conditions.

(5) What is the reason for the discrepancy between graphs of force versus measured deflection and force versus calculated deflection?

(6) Calculate Young's modulus of elasticity from Eq. 12.46(b) and compare it with the nominal value of E_f.

(7) How do ductile and brittle materials behave under bending?

12.17   Measurement of Dynamic Strains

INTERESTING FACTS: Machine members subjected to large static loads develop large deflections and the parts are usually replaced before fractures occur, because the stresses have exceeded the yield strength. Static failures, thus, give visible warnings in advance. Machine members subjected to dynamic/repetitive/ fluctuating loads/stresses, however, fail at stresses well below the ultimate strength of the material and at stresses which are much lower than those required to cause a fracture in a single application of a load. This is because the fluctuating stresses, alternating between tension and compression, have occurred for a very large number of cycles. Such failures are sudden and dangerous and are called fatigue failures; these phenomena are complicated and are not fully understood. Due to the search for lighter weight materials and structures with cost benefits, the analysis of structural vibration and fatigue life has become important. Fatigue failure can impact the safety of personnel and also reduce the availability of equipment. It is estimated that fatigue accounts for approximately 90% of all service failures of mechanical origin; they manifest as cracks at certain locations in the structure. Cracks can appear in many types of structures such as aeroplanes, boats, bridges, frames, machine parts, cranes, piping systems, reactors vessels, turbines, canal lock doors, offshore platforms, masts and chimneys, transmission towers, gas turbine exhaust systems, and composite structures. Fatigue testing is difficult since it must employ millions of fatigue cycles to provide meaningful design data.

Introduction: Measurement of dynamic strains involves measurement of strains in a vibrating member. The test setup differs from one with static loads because a dynamic shaker is used. The strain gauges and bridge configurations are identical to those used for the measurement of static strains.

Aim of the Experiment: To measure the strains in a cantilever beam subjected to dynamic bending strains.

Theory, Test Setup, and Procedure: An aluminium or mild steel beam is fixed between two solid cement blocks as a cantilever. An electrodynamic shaker is used to excite the beam near the free end. A function generator is used to obtain a sinusoidal signal which is amplified using a power amplifier and fed to the shaker coil. The complete setup is as shown in Fig. 12.21. Two identical strain gauges are pasted longitudinally along the beam near the root as discussed in Sect. 12.7 and shown in Fig. 12.9 for a half bridge bending configuration. The outputs of the two gauges

1. Rigid block   5. Power amplifier
2. Cantilever beam   6. Display device
3. Strain gauge   7. Funciton generator
4. Electrodynamic exciter   8. Carrier frequency amplifier
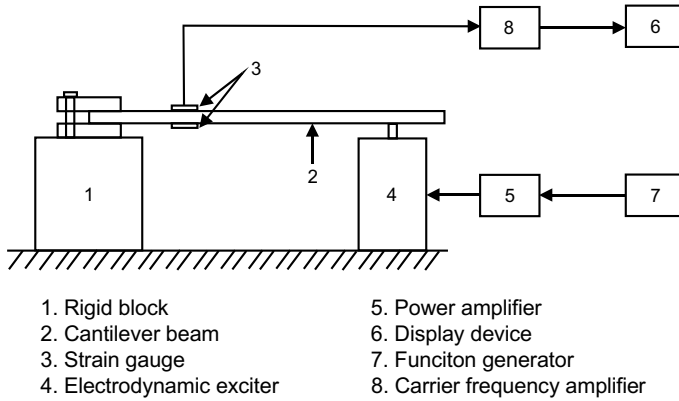
Fig. 12.21 Experimental setup for measurement of dynamic strain

Table 12.8 Record of dynamic strain measurements

Sl. No.	Frequency, Hz	Sine wave excitation V p-p	Strain V p-p	Ratio of o/p to i/p
1.				
2.				
3.				
4.				
5.				
6.				

are fed to a carrier frequency amplifier with AC bridge excitation. The output of the carrier frequency amplifier may be fed to a display device in the form of a storage oscilloscope or recorder. The beam is excited at various frequencies, and at each frequency, the excitation amplitude is changed and the dynamic strains (peak–peak) are noted as shown in Table 12.8.

Questions

(1) What do you mean by AC coupling?
(2) How do you create combined static and dynamic strains?
(3) How do you obtain the static component?
(4) How do you set the sampling frequency for dynamic strain measurement?
(5) What happens if the vibration amplitude is made very large?

12.18   Measurement of Strains Due to Combined Bending and Torsion

DID YOU KNOW that Leonardo Da Vinci, in 1493, was the first scientist to correctly identify the stress and strain distribution across the section of a beam in bending? Unfortunately, he did not have Hooke's law and calculus to complete his proposition. Galileo Galilei, more than a century later, attempted to develop a theory for beams, but unfortunately made the incorrect assumption that the cantilever beam under study rotated about the base at the point of support and that the tensile stress across the section was uniform and equal to the tensile strength of the material. Leonhard Euler and Daniel Bernoulli were the first to put together a useful beam theory around 1750. The Bernoulli–Euler beam theory is still used to date to analyse the behaviour of beams under axial and bending forces. The two primary assumptions made in this beam theory are that 'plane sections remain plane' and that deformed beam angles (slopes) are small. There are, of course, other more complex theories such as the Timoshenko beam theory, but the Bernoulli–Euler theory or classical beam theory, being a special case of the Timoshenko beam theory, is accurate enough for design in most cases.

Introduction: Many structural members are to be designed to resist more than one kind of loading or, in other words, combined loading. Examples of specimens loaded in bending and torsion include cylindrical pressure vessels supported as beams, aircraft wing spars, road signs on vertical poles along roads, distributed loads due to books on bookshelves, propeller shafts of ships, and power transmission shafting of all rotating machinery where a shaft is subjected to bending moment and torsion and direct thrust as well. There are also some types of levers, springs, brackets, screws for power transmission, etc. which are subjected to such combined loading. These structures can be analysed by doing a superposition of the stresses and strains caused by each load acting separately. This is permissible only if (i) the stresses and the strains are a linear function of the applied loads and (ii) the deformations are assumed to be small.

Theory: The theory discussed in Sect. 12.7 for bending of a beam with a half bridge as shown in Fig. 12.9 and that discussed in Sect. 12.11 for a beam in torsion with the resulting strain measured as shown in Fig. 12.13 may be used by doing a superposition of the stresses and strains caused by each load acting separately.

Aim of the Experiment: To measure the strains caused by combined bending and torsional loads.

Experimental Setup and Test Procedure: The experimental setup consists of a hollow cylindrical shaft AB (Fig. 12.22a) rigidly fixed at end A to prevent rotation
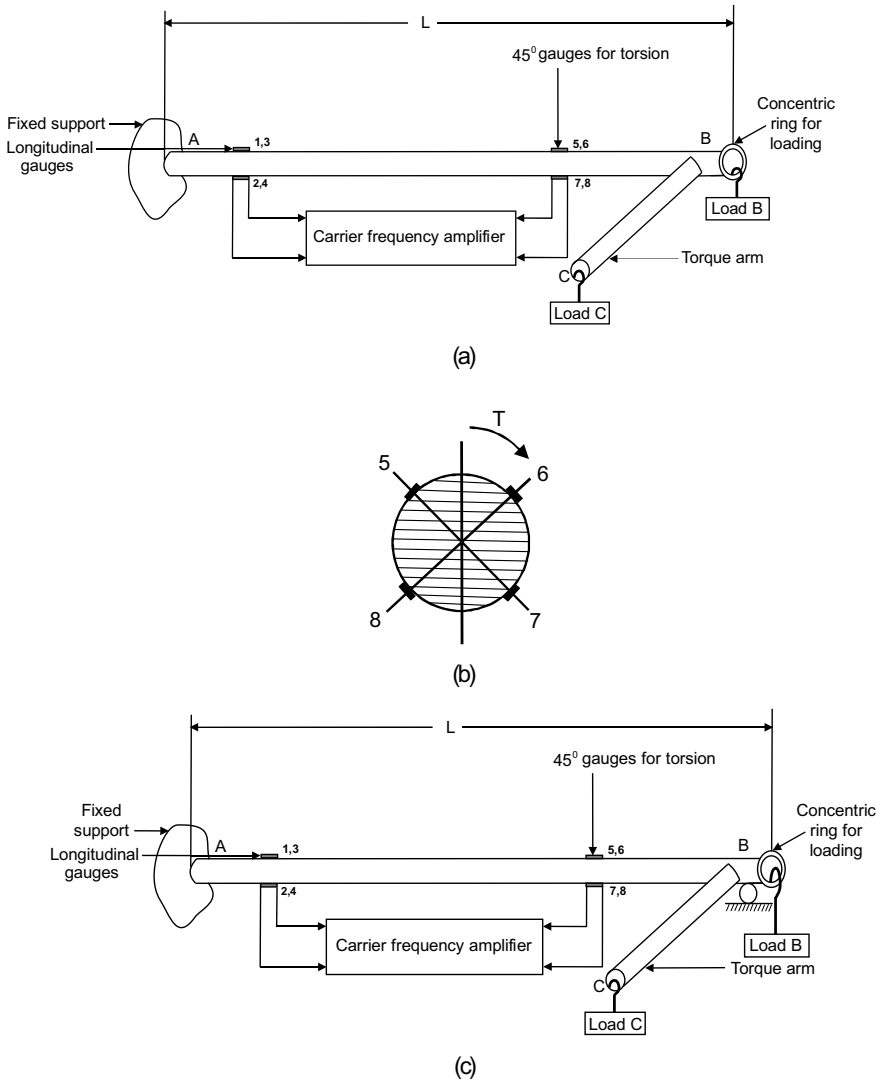
Fig. 12.22 Setup for measurement of combined bending and torsional strains: **a** without roller at B, **b** torsional strain gauge arrangement, and **c** with roller at B

and supported on a roller at end B. A ring concentric to the shaft with a hook for hanging weights is provided at end B. A torque arm BC is rigidly fixed to the shaft at B with a provision to load arm BC at C by hanging known weights from a hook at C. The shaft may be subjected to the following three types of loading:

(i) When a weight is hung from B with the roller support removed at B (Fig. 12.22a), shaft BC will be subjected to pure bending.

Table 12.9 Measurement of bending and torsional strains

Sl. No.	Bending load, N	Bending Strain, μm/m	Moment, N.m	Torsional Strain, μm/m	Combined Strain, μm/m
1.					
2.					
3.					

(ii) When a weight is hung from C, i.e., from the end of the torque arm BC, with the roller kept at B, the shaft will be loaded in pure torsion.

(iii) With loads B and C applied and with the roller support removed at B, the shaft will be subjected to combined bending and torsion (12.22c).

Four strain gauges (#1 to #4) are bonded along the length of shaft AB, #1 and #3 on the top surface symmetrically placed on either side of the centre line and very close to it as shown in Fig. 12.11, and gauges #2 and #4 on the bottom surface as mirror images of gauges #1 and #3, respectively, to sense the bending strain as described in Sect. 12.9. These four gauges are connected to a carrier frequency amplifier (Fig. 12.22a) to form four arms of a full Wheatstone bridge to sense bending with good sensitivity. Gauges #5 and #6 are connected at 45° on either side of the longitudinal axis on the top surface of shaft AB as described in Sect. 12.12. Gauges #7 and #8 are pasted on the bottom side of the shaft as mirror images of gauges #5 and #6 as shown in Fig. 12.2b. This arrangement ensures that bending does not affect the torsional strain readings. These four gauges are connected in full bridge configuration in another Wheatstone bridge to sense torsional strain as described in Sect. 12.11. Results may be tabulated as shown in Table 12.9.

Questions

(1) What are the basic assumptions in deriving the bending equation?
(2) What are the basic assumptions in deriving the torsion equation for a circular member?
(3) How do you decide on the limiting loads?
(4) What are the major sources of error in this experiment?

12.19   Study of Buckling Behaviour of a Column

DID YOU KNOW that stone columns were used in ancient Egypt as early as 2600 B.C.E.? Some of the columns in the ancient world with the most elaborate workmanship were those of Persia, many of which are still standing

today, some being more than 30 m tall. Single-piece heavy stone columns were extensively used in ancient architecture. Other stone columns, made of multiple sections of stone, were also used; they were generally fixed using mortar or dry-fit together. Many of the present-day columns are made of steel, poured or precast concrete, or brick. The San Jacinto Monument, a 172.92 m column located in Texas, United States, is the world's tallest masonry column. It is topped with a 200 metric tonne star.

Introduction: A column is a straight slender member subjected to an axial compressive load. If this member is short, it remains straight when loaded and its failure results from yielding of the material of the member by crushing or plastic squashing. If, on the other hand, the member is long, a different kind of failure is observed. When the compressive load is equal to the 'critical load', the member undergoes large deflection in the lateral direction due to its instability. This deflection becomes very large with even a very small increase in load and this phenomenon is called buckling. This mode of failure is quick, and hence dangerous. Buckling can occur even when the maximum stress in the column is less than the yield stress of its material. Buckling analysis involves finding the maximum load a column can support before it collapses. It is a unique field of structural mechanics in the sense that failure is not related to the strength of the material; it is instead dependent on the column's stiffness, both material and geometric.

Aim:

(1) To determine the load-carrying capacity of slender columns of varying lengths and study the effect of slenderness ratio.
(2) To compare the experimental buckling load and critical stress of the test column with those predicted by the Euler equation.

Theory: The vertical members of a structure which carry mainly compressive loads are said to be columns. Though both a column and a strut are subjected to compressive forces, a column is generally considered as one with an unsupported length which is at least 10 times its least lateral dimension. Ideal columns are those that are perfectly straight, loaded exactly axially, those which are free of any residual stress, and manufactured from a perfectly isotropic material; this is not true of real columns. However, theory regarding such ideal columns contributes greatly to our understanding of column behaviour. The four common boundary conditions (B.C.) for conducting studies on columns are (i) clamped–free (C-F), (ii) pinned–pinned (P-P), (iii) clamped–pinned (C-P), and (iv) clamped–clamped (C-C) as shown in Fig. 12.23. Buckling depends on the slenderness ratio, defined as the ratio of the effective length of a column (L_e) to its smallest radius of gyration (r_g) about the axis under consideration. Here, $L_e = KL$ where L is the actual unsupported length of the column and K is the effective length factor which depends on the boundary
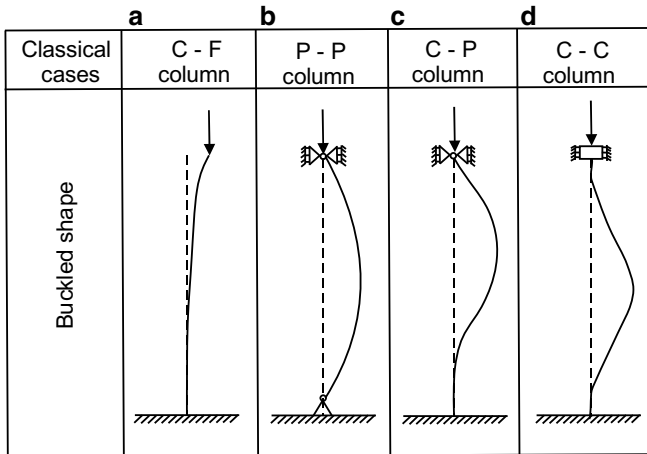
Fig. 12.23 Typical boundary conditions for columns: **a** Clamped–free (C-F), **b** Pinned–pinned (P-P), **c** Clamped–pinned (C-P), and **d** Clamped–clamped (C-C)

conditions. It was Leonard Euler, a Swiss mathematician (1707–1783), who first derived an expression for buckling loads. The reader is requested to read Sect. 6.8 for the requisite theory. The main assumptions in his theory are as follows:

(i) The material of the column is homogeneous, elastic, and isotropic, thus, obeying Hooke's law.
(ii) The column is straight and is loaded truly axially.
(iii) Its cross section is uniform throughout its length.
(iv) Failure occurs through buckling alone.
(v) The self-weight of the column is negligible.
(vi) The theory is applicable only to long slender columns and not to columns of small or intermediate lengths.
(vii) The shortening of the column due to direct compression is very small and negligible.
(viii) Euler's formula determines critical loads, not working loads. Hence, a suitable factor of safety, typically, 2 or 3, depending on the material, is to be used to obtain a practical allowable value.

Test Setup and Procedure: The test setup is as shown in Fig. 12.24. The dimensions of the specimen, a steel column, are to be measured using a steel scale and slide calipers. The column is then to be placed in the testing apparatus, typically a universal testing machine, with the appropriate supports for the required boundary conditions. The test setup has a height-adjustable load-carrying cross-arm and a spindle to apply a compressive force to the column. An axial bearing between the spindle and the column support prevents the torsional loading of the column. The axial compressive load is applied and is measured using a load cell. The lateral deflection of the bar is indicated on a dial gauge. The critical buckling load is to be noted. The test is to be repeated for all the required boundary conditions. The slenderness ratio and
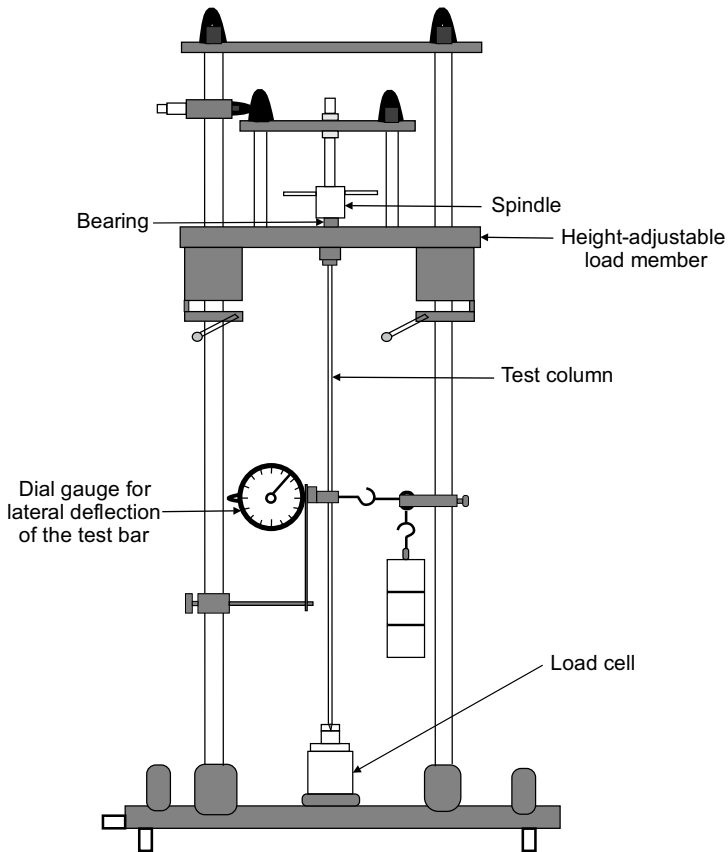
Fig. 12.24 Test setup for determination of column buckling load

critical loads for the different support conditions may be computed. Readings may be tabulated as shown in Table 12.10. The following plots may also be made:

  (i) Experimental P_{cr} and theoretical P_{cr} versus K.
 (ii) Experimental P_{cr} and theoretical P_{cr} versus $L_e = KL$.
(iii) Experimental critical stress σ_{cr} and theoretical critical stress σ_{cr} versus slenderness ratio KL/r_g.

Questions

(1) As a designer, what steps can be taken to reduce the tendency to buckle, (i) geometrically? (ii) material-wise?
(2) How well did the equations predict the actual critical compressive force?
(3) Were the observed discrepancies reasonable?
(4) What are the possible sources of error?

Table 12.10   Measurement of bending and torsional strains

B.C.	Effective length factor K	Length L, mm	Radius of gyration r_g, mm	Slenderness Ratio KL/r	Anal. P_{cr}, N	Expt. P_{cr}, N	Expt. σ_{cr}, N/m^2	Anal. σ_{cr}, N/m^2
C-F	2							
P-P	1.0							
C-P	0.7							
C-C	0.5							

Note: Anal.: Analytical, Expt.: Experimental

12.20   Measurement of Stress Concentration Factors

Have you heard of the Liberty ships, 2708 of which were constructed by the United States during the period from 1939 to 1945 for World War II? 1031 of these got damaged due to brittle fracture by April 1, 1946. More than 200 of them sank or got damaged beyond all repair. These Liberty ships were manufactured from steel that experienced a ductile-to-brittle transition when they were docked in the frigid North Atlantic, where temperatures dropped to below the transition temperature. The first few ships experienced major structural damage when cracks developed in their decks and hulls. Three of them catastrophically got split into half due to cracks that developed and grew to critical lengths and then rapidly propagated completely around the ships' girths. Though the exact origin of the fracture could not be determined, it was assumed that in 52% of all serious fractures, the growth started at the corners of the cargo hatches (doors) which were square and acted as points of stress concentration where cracks initiated.

Introduction: It is known that a change in the geometry of a part results in stress over and above the computed stress and this is known as stress concentration. The more abrupt the changes in geometry, the higher is the stress concentration effect. Because of this, a more intelligent design approach is required to reduce the stress concentration effect in critically stressed areas. Analysis of stress concentration was first done in 1898 by Ernst Gustav Kirsch who proposed a linear elastic solution for stresses around a hole in an infinite plate. Kirsch first solved the basic case of uniaxial tension. He proposed the well-known factor-of-three stress concentration at the hole under uniaxial loading, but we know that it can, in fact, vary from two to four for complex loading conditions, different from uniaxial tension. Superposition techniques may be used to develop solutions for the more complex cases, say, equibiaxial tension and shear. Some methods of reducing stress concentration that may be employed are as follows: using only rounded corners with large radii and avoiding sharp corners, polishing surfaces to remove any notches or defects that occur dur-

ing processing, reducing the stiffness of straight load-bearing segments, and placing notches and threads in low-stress areas.

Aim: To study stress concentration around a circular hole in a rectangular plate subjected to uniaxial tension.

Theory: The theory is as discussed in Sect. 6.9.

Test Setup and Procedure: An aluminium plate of length 60 cm, width 15 cm (2b), and thickness around 6 mm (h) with a circular hole of diameter 37.8 cm (2a) at its centre (Fig. 12.25) is to be subjected to a tensile load of magnitude P (within elastic limit). The resulting strain distributions are to be measured using strain gauges bonded to different locations on the plate. Local tangent strains are to be measured at five locations (gauges #1, 2, 3, 4, and 5 as shown in Fig. 12.25) along the horizontal centreline. Results may be tabulated as shown in Fig. 12.26 in terms of normalized stress versus a normalized coordinate which is defined as $(r - a)/(b - a)$ where r
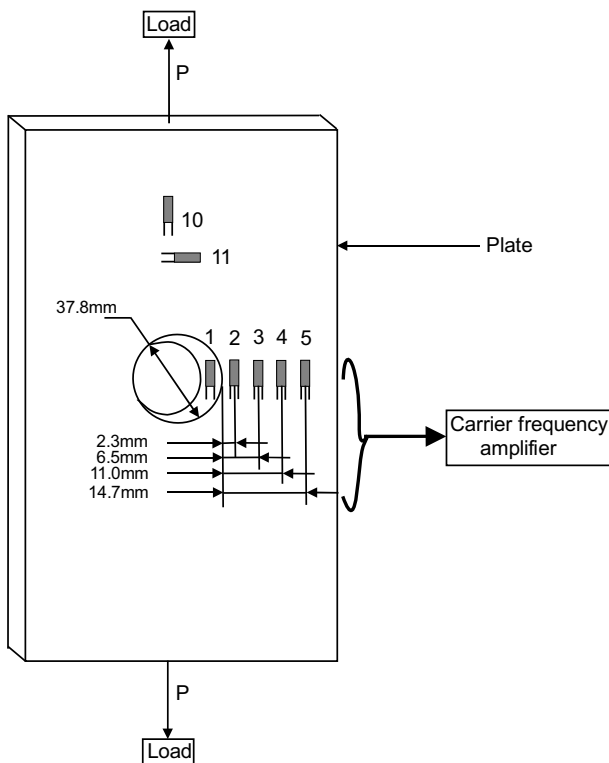
Fig. 12.25   Locations of strain gauges on the plate
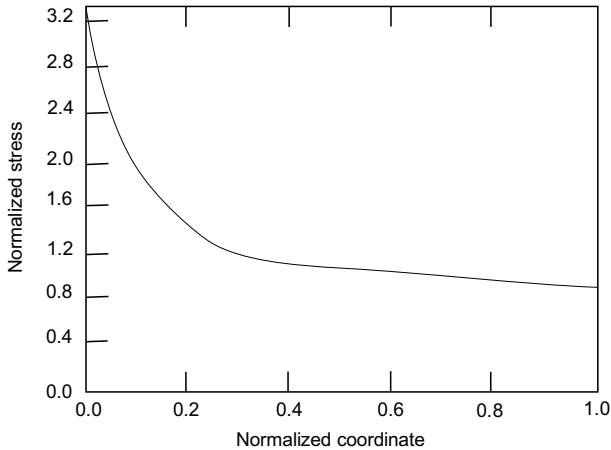
Fig. 12.26 Normalized stress distribution along horizontal centreline

is the distance along the horizontal centre line with $r = 0$ at the centre.

Questions

(1) What is stress concentration factor?
(2) How can stress concentration factor be found out from photo-elasticity?
(3) How can brittle coatings be used to find out stress concentration factor?
(4) Is fracture toughness related to stress concentration factor?
(5) How do you take stress concentration into consideration in the case of a component subjected to dynamic loading?

Bibliography

1. Boyes, W. (2010). *Instrumentation reference book*. UK: Elsevier.
2. Dally, J. W., & Riley, W. F. (1991). *Experimental stress analysis*. McGraw-Hill College.
3. Doebelin, E. O., & Manik, D. N. (2007). *Measurement systems*. Tata McGraw-Hill.
4. Hannah, R. L., & Reed, S. E. (1992). *Strain gage users' handbook*. Netherlands: Springer.
5. Holman, J. P. (2004). *Experimental methods for engineers*. Tata McGraw-Hill.
6. Jindal, U. C. (2012). *Experimental stress analysis*. Pearson Education India.
7. Kobayashi, A. S. (1987). *Handbook on experimental mechanics*. Prentice Hall.
8. Khan, A. S., & Wang, X. (2001). *Strain measurements and stress analysis*. Prentice Hall.
9. Oehlert, G. W. (2010). *A first course in design and analysis of experiments*. Library of Congress.
10. Perry, C. C., & Lissner, H. R. (1955). *The strain gage primer*. McGraw-Hill Book Company Inc.
11. Rangan, C. S., Sarma, G. R., & Mani, V. S. V. (1983). *Instrumentation devices and systems*. New Delhi: Tata McGraw Hill Publishing Company Ltd.
12. Restivo, M. T. (2014). *Fernando Gomes de Almeida and Maria de Fatima Chouzal*. Strain Measurement: Lulu.com.

13. Srinivas, J. (2011). *Stress analysis and experimental techniques: An introduction*. Alpha Science International Ltd.
14. Vaughan, J. A. M. (1975). *Application of B & K equipment to strain measurement*. Brüel and Kjær.
15. Window, A. L. (1993). *Strain gauge technology*. Netherlands: Springer.
16. Yakovlev, V. F., & Inyutin, I. S. (2014). *Measurements for stresses in machine components*. Pergamon.

Index