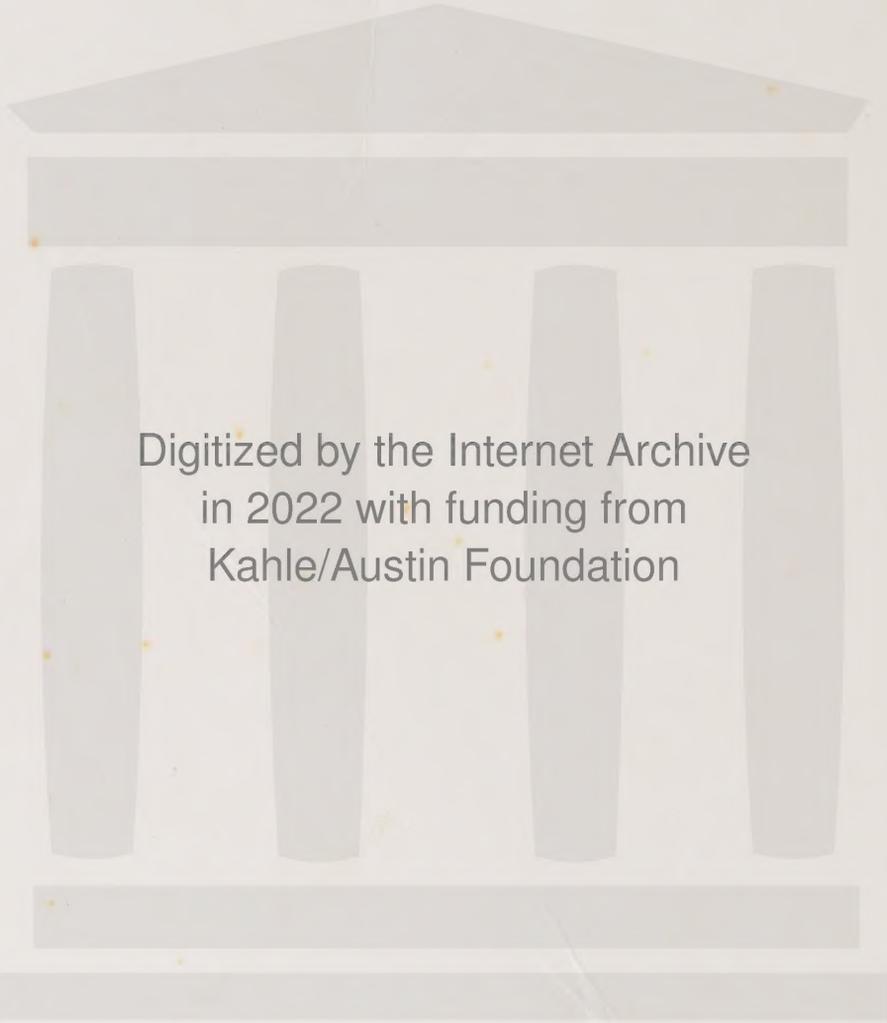


ESSENTIAL PRINCIPLES OF PHYSICS

SECOND EDITION



WHELAN & HODGSON



Digitized by the Internet Archive
in 2022 with funding from
Kahle/Austin Foundation

119.20
179.89

ESSENTIAL
PRINCIPLES OF
PHYSICS

Shareed Ali

Second edition

SCHOLAR'S BOOK W
Jubilee Cinema bu
Main Road, Chag

ESSENTIAL PRINCIPLES OF PHYSICS

Second edition

Shareed Ali

ESSENTIAL PRINCIPLES OF PHYSICS

Second edition

P M WHELAN MA • M J HODGSON MA

St. Martin's School, Northwood

Shareed Ali

#7

John Murray

By the same authors
Questions on Principles of Physics

First published as *Essential Pre-University Physics* © P. M. Whelan and M. J. Hodgson 1971

Essential Principles of Physics © P. M. Whelan and M. J. Hodgson 1978, reprinted 1979,
1982, 1985, 1987

Second edition © M. J. Hodgson and S. R. McLeod 1989, reprinted 1990

All rights reserved. Unauthorized duplication contravenes applicable laws.

Printed in Great Britain by BPC Hazell Books

British Library Cataloguing in Publication Data

Whelan, P. M. (Patrick Michael)

Essential principles of physics. — 2nd ed.

I. Title II. Hodgson, M. J. (Michael John)

530

ISBN 0-7195-4566-8

PREFACE

Aims

(a) To provide a rigorous one-volume reference book which discusses all the principles of physics that are needed by a student going on to read physics at university.

(b) To present an account of physics up to GCE S level in which both the choice of subject matter and its emphasis are in line with current teaching trends, and in which the presentation, nomenclature and symbols follow the most recent practice.

The layout of the book has been dictated by two considerations:

(i) We want to encourage the student to use the book continually, and to make it the basis of his/her course. Every item has been fully indexed so that it can be found quickly.

(ii) It is intended that the book should be used for *learning* and *revising* (rather than for general reading).

The text is deliberately fragmented, and has been presented so that it makes an immediate visual impact. Throughout, the verbal explanations and descriptions have been reduced to a minimum, but their function has to some extent been replaced by about 600 fully labelled diagrams.

Content

Deliberately, this book is *not* about modern physics: there is little or no treatment of topics such as fundamental particles, masers and lasers, holography or astrophysics. We feel that these developments contain ideas so difficult that they cannot be taught properly at this level, and certainly they merit no space in a book of this type.

Rather we have tried to stress those ideas on which the student can later build his/her understanding of modern physics. We have excluded material of secondary importance to make room for a thorough introduction of the *fundamental* principles.

The scope of the book is roughly that of the Special Paper of GCE A level. Those parts of the text that go beyond A level are marked with an arrow ► and are printed on a tinted background.

The book assumes an understanding of the ideas developed during a GCSE course.

Details

(a) **Physical quantities** are defined by equations throughout the book, and the appropriate SI unit is indicated for each quantity involved in the definition.

(b) Great emphasis has been put on the part played by **molecules**. Inter-molecular forces and energies are discussed at greater length than is usual at this level to pave the way for the study of materials science and the solid state.

(c) An attempt has been made to stress the wide application of the concept of **wave motion**. In Section III we develop the properties of waves in general, and then apply these ideas to electromagnetic waves and sound waves in Sections VII and VIII.

(d) In Section V on **thermal properties of matter** the physical quantity *amount of substance*, and its unit the *mole*, have both been used freely. We have both

here and elsewhere stressed the importance of being consistent about dimensions: for this reason we have avoided (as far as possible) the use of unit quantities such as 1 mol, or 1 kg. It is our experience that they often hide the true nature of a relationship.

Temperature is introduced through the thermodynamic scale, using the triple point of water and (by implication) absolute zero as the fixed points. The *Celsius* temperature is then developed from the thermodynamic scale.

PREFACE TO THIS NEW EDITION

Since the publication in 1978 of *Essential Principles of Physics*, the various Advanced level syllabuses have been changing rapidly and significantly. The new edition has been written to cater for these developments. Diagrams have been added wherever appropriate and, although some of the new material is relatively descriptive, the treatment has remained concise. It is increasingly important for students to be clear about their syllabus requirements and to use the book selectively. There has been no attempt to cover all the *optional topics*, such as *medical physics* and *telecommunications*, that are included in a number of syllabuses. However, some of them, such as *rotational dynamics* and *fluids and waves*, need only a small amount of reinforcement from other sources.

These are the principal changes:

(a) Expansion of the chapters in Section IV on **Structure of Matter and Elasticity**, to deal with the properties of materials.

(b) Extensive development of the chapter on **Electronics** in Section XI.

(c) A new chapter entitled **Energy Resources** in Section XII.

(d) Additional material in 22 other chapters, particularly those in Sections II, V and XI, often dealing with practical and engineering applications. (The companion volume, *Questions on Principles of Physics*, includes further applications.)

March 1988

M.J.H.

Publisher's note:

After a long illness, courageously borne, Paddy Whelan died in May 1980 at the age of 42.

A NOTE TO THE STUDENT

Learning Physics

Acquiring proficiency in physics involves each of the following:

(a) *Complete familiarity* with the small number of *fundamental concepts* that constitute the real basis of the subject.

(b) *Access to detailed textbooks* which provide the thorough explanations necessary for a proper understanding, and which describe experimental work in some detail.

(c) *Ability to apply the basic laws and principles to both familiar and unfamiliar problems*, an ability which is developed by constant practice.

(d) A laboratory course of practical work.

The purpose of this book is to satisfy the first requirement. You are strongly recommended to make your own *summary of the book* in about a hundred pages, which can easily be done by emphasizing symbols and labelled diagrams, and economizing on words. This summary will help you to become familiar with

the book, will demonstrate the *logical structure* of the subject, and will serve for last minute *revision*.

Revision

Well before the exam (say three months at least) draw up a *written timetable*, and keep to it.

- (1) Set aside for each day a definite length of time (say *not more* than 30 minutes).
- (2) Find a place where you will not be distracted.
- (3) Have a pencil and paper to hand.
- (4) Choose a particular topic to revise, and then devote *all* your attention to that topic.
- (5) Pay particular attention to diagrams and their labels.
- (6) Repeat to yourself the full meaning of a graph (e.g. emphasize which is the dependent, and which the independent variable, and whether some other quantity is by implication held constant).
- (7) When you think that a topic is mastered, test yourself by *re-writing the main ideas*, as though explaining them to someone who has no previous knowledge of that topic.

Revision should be very much an active (not a passive) *occupation*.

Examination technique

An examination is an exercise in efficient communication, and therefore also requires planning.

- (1) Read the *rubric* at the top of the paper very carefully to see exactly what you have to do.
- (2) See how much *time* can be spent on each question, and do not exceed it. If you run short of time, make your answer shorter, but make sure it includes the fundamentals.
- (3) *Read the question* that you are answering very carefully to see *exactly* what is required. Do not spend time giving details that are not asked for, and do not repeat yourself. Work out the order in which the different parts of the question are best answered: it is usually better to tackle a problem before a lengthy description.
- (4) Illustrate your answer by very large *clear labelled diagrams and graphs* wherever you can. Far more information can be conveyed by a quick sketch than in sentences of description. If you are asked to do a scale drawing, draw it accurately and as large as the paper will allow. State your scale.
- (5) The following procedure is recommended in *numerical problems*:
 - (i) Read the question carefully to see *what has to be calculated*.
 - (ii) Summarize the information, making the *units* explicitly clear. This can frequently be done most clearly and concisely with a diagram.
 - (iii) State clearly the *principle* on which your method depends. Often an equation is enough.
 - (iv) Do the numerical computation in such a way that the examiner can easily find mistakes. Use units as recommended on p. 10.
 - (v) State your answer with a complete sentence, and make sure that it is of a *sensible order of magnitude*. If you think that it is not, then say so.
- (6) When describing experiments, list events in the order in which they happen. Illustrate by symbols the measurements to be taken, and demonstrate how the result is to be calculated.
- (7) *Good English and legible handwriting are essential*: without them you cannot communicate.

However good your examination technique, there is no substitute for knowing your subject.

ACKNOWLEDGEMENTS

C.J. Millar generously made his own teaching notes available during the preparation of the original manuscript, and read part of the manuscript.

Many people gave helpful advice during the preparation of the new edition, in particular **R.O. Ridley** on the Bernoulli equation and **W.C. Thoburn** on surface tension.

J. Allen assisted in the expansion of the electronics chapter for the new edition.

J.W. Warren read the original typescript, correcting numerous minor errors and contributing constructive suggestions, particularly with regard to molecular behaviour. **P. Knight** meticulously checked the additional material for the new edition.

T. Robinson provided an unusual combination of artistic skill and understanding of physics to draw the original diagrams. **The Parkway Group** drew the additional diagrams for the first edition of *Essential Principles of Physics*, and **RDL Artset** those for the second edition, in a style which matches well those of the original book.

J. Macqueen expertly copy-edited the script of the new edition and checked the proofs. **J. Pickering** also helped immensely in reading the proofs.

Many current publications, particularly the books listed in the bibliography (p. 540) have influenced the writing of this book.

The cover photo is a representation of a Newton's cradle and was taken by **David Purdie**. The cover design is by **Peter Theodosiou**.

CONTENTS

I	GENERAL INTRODUCTION	1
II	PRINCIPLES OF MECHANICS	27
III	OSCILLATIONS AND WAVE MOTION	81
IV	STRUCTURE AND MECHANICAL PROPERTIES OF MATTER	125
V	THERMAL PROPERTIES OF MATTER	179
VI	GEOMETRICAL OPTICS	251
VII	WAVE PROPERTIES OF LIGHT	281
VIII	SOUND WAVES	321
IX	ELECTROSTATICS	335
X	CURRENT ELECTRICITY	371
XI	ELECTRONIC, ATOMIC AND NUCLEAR PHYSICS	467
XII	ENERGY RESOURCES	531
	<i>Appendix A. Special Names and Symbols for Derived SI Units</i>	538
	<i>Appendix B. Recommended Values of Selected Physical Constants</i>	539
	Bibliography	540
	Index	542

I GENERAL INTRODUCTION

1 GENERAL INTRODUCTION

1.1	PHYSICAL INTERACTION AND THE NATURE OF PHYSICS	2	1.7	ORDERS OF MAGNITUDE	12
1.2	PHYSICAL QUANTITIES	4	1.8	TREATMENT OF ERRORS	13
1.3	THE NATURE AND STATUS OF PHYSICAL EQUATIONS	5	1.9	SCALAR AND VECTOR QUANTITIES	17
1.4	THE ORDER IN WHICH PHYSICAL QUANTITIES ARE DEFINED	6	1.10	THE USES OF DIMENSIONS	19
1.5	UNITS AND THE REALIZATION OF STANDARDS	9	1.11	DIMENSIONAL ANALYSIS	20
1.6	MEASUREMENT	11	1.12	SOME USEFUL MATHEMATICS	21
			▶ 1.13	DIFFERENTIAL EQUATIONS IN PHYSICS	23
			1.14	GRAPHS	25

1.1 PHYSICAL INTERACTION AND THE NATURE OF PHYSICS

The structure of physics

Natural science gives descriptions of the natural world, but in the last resort it can only describe *what* happens rather than why it happens.

Some great themes persist throughout physics, and give the subject its unity. These include (for example):

- (a) the concept of *energy*,
- (b) the *conservation laws*,
- (c) the concept of *wave motion*,
- (d) the concept of *field*;

and it is possible to study the subject by grouping topics under those headings. However, it is more convenient (at pre-University level) to subdivide physics according to subject matter (Mechanics, Light, Electricity, etc.) and to coordinate during a degree course the ideas thus presented.

Some topics are too difficult to study at this elementary level. For example:

(a) *Mechanics*

(i) *Relativistic quantum mechanics* deals with the mechanics of all bodies.

(ii) *Quantum mechanics* treats the special cases where bodies have speeds that are very much less than that of light, but caters for all masses.

(iii) *Relativistic mechanics* is applied in the special cases where bodies have masses greater than (say) 10^{-27} kg, but caters for all speeds.

(iv) *Newtonian mechanics* (or *Classical mechanics*) is the only form discussed in detail in this book, and can be used in the double special case: (1) *relatively* small speeds, and (2) *relatively* large bodies. It is nevertheless applicable (as an excellent approximation) to all macroscopic bodies moving at these speeds. This is illustrated by fig. 1.1(a) and (b) opposite.

(b) *Statistical mechanics* (which can be *classical* or *quantum*) applies the techniques of statistics to very large numbers of interacting particles in an attempt to explain the (macroscopic) properties of matter in bulk. Classical statistical mechanics links *kinetic theory* and *thermodynamics* (chapter 26).

(c) *Electromagnetism*. *Maxwell's equations* describe the behaviour of electromagnetic fields, and at the end of the nineteenth century provided a wonderful link between the theory of *light* and that of *electricity and magnetism*. The linking process is beyond the scope of this book.

The beginning student must always be prepared for some change of emphasis when viewing the subject later with a more experienced eye.

Interactions

An **interaction** in physics is a process by which one particle is enabled to influence (exert forces on) another. At present it seems that there are four interactions, which are summarized in the table opposite.

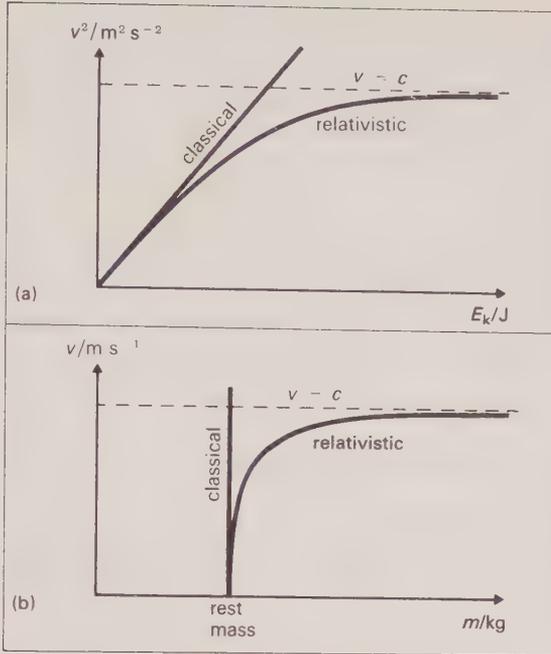


Fig. 1.1. Departure of classical from relativistic mechanics.

► There is one further effect which could be called an interaction. According to the **Pauli exclusion principle** no two electrons in an atom may be described by the same set of quantum numbers, which is equivalent to saying that the electrons can exert some sort of repulsive force on each other (quite apart from the *Coulomb* interaction).

Fields

In this book we are mainly concerned with the electromagnetic and gravitational interactions. We visualize the following to happen:

(1) Body A establishes round itself a region in which it has the ability (or potential) to exert forces. This region is the **field**.

(2) If body B enters the field, then it experiences a **force** (as does A which is now in the field set up by B).

One can imagine the field to be the *cause*, and the force to be the *effect*.

In the past there has been a tendency to emphasize the field aspect to develop the theory of the electromagnetic interaction, and the force aspect for gravitation. In this book field ideas are used for both interactions—see chapters 20, 43 and 52.

Physical laws

The aim of physics is to find some understanding of the physical world through the formulation of general principles. We look for statements that summarize ideas of importance with great precision and yet simplicity: these statements should relate observations from widely different experiments, and should be capable of general application. *Such a summary is frequently called a law.* It is often in the form of an equation connecting symbols that represent operations which we carry out in the laboratory. We would like the equation to relate cause and effect, but laws are *not necessarily* obeyed. A law may represent the relation

Name	Relative effectiveness	Comment	Example of interaction
Strong or nuclear	1	large short-range forces between nuclear particles	the holding together of nucleons in a nucleus
Electro-magnetic	$\sim 10^{-2}$	account for the majority of the forces we meet in everyday life	repulsion between two solid bodies whose surfaces are brought into contact
Weak	$\sim 10^{-14}$	little understood at present	β -decay of nucleus (p. 526)
Gravitational	$\sim 10^{-40}$	only become significant when very large masses are involved (and then the force is dominant)	pull of Earth on Moon and Moon on Earth

between cause and *probable* effect, particularly when quantum effects are significant.

In time it becomes apparent that some laws (such as *Newton's Law of Gravitation*) are inadequate: they do not represent *exactly* what happens, and may have to be modified before they can be applied in certain situations.

Some other terms used in physics

Unlike other terms in physics, these are *not* always used with exactly the same meaning. This list is a rough guide as to what is usually intended.

(a) A **fact** is an experimental observation which can be reproduced by repeating an experiment.

(b) A **law** (see above). Examples include *Snell's Law* (p. 101) (to which there are exceptions), *Newton's Second Law* (p. 34) (which in the form $F = ma$ only applies exactly in classical mechanics), and the *Law of Conservation of Momentum*, to which there are no known exceptions.

(c) A **principle** is a general belief, and may often be called a law. *Heisenberg's Indeterminacy or Uncertainty Principle* is an example. The fact that the momentum of an isolated system remains constant is sometimes called a law, and sometimes a principle. The two terms are quite often interchanged.

(d) A **rule** is the name given to a series of working instructions, such as *Routh's Rules* (p. 59), and *Kirchhoff's Rules* (p. 389).

(e) A **theory** is a framework of ideas, based on experiment, which tries to relate previously disconnected observations by a single explanation. Examples include *Einstein's Theory of Relativity*, and the *Kinetic Theory of Matter* (p. 136). In the same way as a law, a theory may well be modified in the light of experience (e.g. *Newton's Corpuscular Theory* (p. 107)).

(f) An **hypothesis** is a more tentative idea proposed to explain observed phenomena, which has not yet been tested by a crucial experiment. If it survives the test it may then be called a theory.

Example. Avogadro's Law (p. 200) was for a long time referred to as *Avogadro's Hypothesis*.

Physical models

Progress in physical science occurs in two ways:

(1) by experiment (the characteristic feature of science), and

(2) by appeal to models.

A **model** is a concept (rather than a material structure) proposed by scientists as an aid to visualizing the possible processes within a particular physical situation. Such a model may give

a highly simplified (or even misleading) picture of reality. If they then apply established principles to the hypothetical model, they may be led to make predictions which themselves lead to further experimental tests of that model. (See also the development of the history of gravitation (p. 168).)

Models have proved themselves especially useful in

(a) visualizing atomic structure (p. 513),

(b) developing the kinetic theory of matter (p. 136), and

(c) explaining the propagation of electromagnetic waves (p. 282).

1.2 PHYSICAL QUANTITIES

We can classify physical quantities as

(a) **basic, primary or fundamental quantities**, which by agreement are treated as independent, and

(b) **derived or secondary quantities**.

(A) Basic quantities

For convenience, seven quantities have been selected:

<i>mass</i>	<i>electric current</i>
<i>length</i>	<i>temperature</i>
<i>time</i>	<i>luminous intensity</i>
<i>amount of substance</i>	

It is to be emphasized that these quantities are selected for *convenience*, and not through any necessity. Thus in mechanics one could equally well have chosen length, *force* and time, in which case mass would have been a derived quantity. Once these quantities are chosen, we can

(1) select an *arbitrary* unit, and then

(2) measure by direct comparison with that unit.

This is essentially a *counting* process, in which we follow a specified procedure.

(B) Derived quantities

Any other physical quantity can be derived from the basic quantities by following the procedure implied in the defining equation (see below).

Defining equations

A physical quantity is defined by a series of operations which are best summarized by an equation. The equation will relate the primary quantities by

that series of operations, each of which involves only multiplication, division, integration or differentiation.

Examples

$$v = \frac{ds}{dt} \quad (\text{p. 28})$$

$$\Phi = \int \mathbf{B} \cdot d\mathbf{A} \quad (\text{p. 401})$$

$$\frac{dQ}{dt} = -\lambda A \frac{d\theta}{dx} \quad (\text{p. 237})$$

These equations are more useful than verbal descriptions, since they can be applied immediately.

Notes

(a) Equations can only serve as complete definitions when each symbol used in the equation is explicitly defined.

(b) An equation involving only symbols for physical quantities does not specify or require any particular system of units.

(c) In this book each defining equation is repeated, and a recommended set of consistent units is indicated in *square* brackets after each term.

Example. If a normal force F acts on an area A , then the average pressure p is given by

$$p = F/A$$

$$p[\text{Pa}] = \frac{F[\text{N}]}{A[\text{m}^2]}$$

Strictly the units have no place in the equation.

Dimensions

The **dimensions** of a physical quantity indicate explicitly how the quantity is related, through its defining equation, to the basic quantities. If we write $[v] = [L T^{-1}]$ we indicate that speed is measured by dividing a length by a time.

Some quantities (such as refractive index) are described by a number which is independent of the units of measurement chosen for the primary quantities: such quantities are said to be **dimensionless**.

Some quantities have units which are part dimensional, and part dimensionless. The name or abbreviation of the dimensionless unit is not always explicitly stated.

Example. Angular frequency (or pulsance) ω is usually expressed in *radians per second* (rad s^{-1}), but this is often written *per second* (s^{-1}).

The uses of dimensions are discussed more fully in section 1.10 (p. 19).

Conventions for printing and writing

This book adopts the convention in which

(a) symbols for physical quantities are in *italics* (e.g. mass m),

(b) symbols for units are in roman (upright) type (e.g. kg).

The value of any physical quantity can be expressed by

$$(\text{physical quantity}) = (\text{numerical value})(\text{unit})$$

This means that ordinary algebraic rules can be applied to

(i) physical quantities,

(ii) numbers, *and also*

(iii) units (see p. 9).

The chosen symbol represents both a *number* and a *unit*.

Examples

' $V = 5.0$ volts' is correct.

'Let the p.d. be V volts' is wrong.

'Let the p.d. be V' ' is correct.

We might later find

$$V = (5 \times 10^{-3})V = 5 \text{ mV.}$$

V is *not* just a number.

A symbol divided by a unit, such as F/N , may represent a *number*: it is therefore the correct heading for tables and for labelling the axes of graphs (p. 25).

Example. This table summarizes the results of an experiment in which a force F stretched a spring through a displacement Δx .

F/N	$\Delta x/\text{mm}$
0.0	0.0
9.8	25
19.6	49
	etc.

1.3 THE NATURE AND STATUS OF PHYSICAL EQUATIONS

Since physics is primarily an exact (quantitative) discipline, much of our work is concerned with equations. We can distinguish the following types.

(a) *Defining equations.* These were identified in the previous section. It helps if we all use technical terms with the same meaning, and so to reduce confusion *defining equations must be learned*.

(b) *Laws.* Equations representing laws were discussed in section 1.1. These too *must be learned*, since they often represent the result of experiment, and so cannot be simply deduced.

(c) *Principles and useful results.* There are many equations in physics that summarize the application of laws and defining equations to common situations. The following are well-known examples:

$$p + \frac{1}{2}\rho v^2 + h\rho g = \text{constant}$$

$$p = \frac{1}{3}\rho\overline{c^2}$$

$$V = \left(\frac{1}{4\pi\epsilon_0}\right)\frac{Q}{r}$$

$$\mathcal{E} = I(R + r)$$

Since these equations can be derived from laws and definitions, it is *not* strictly *necessary* to learn any of them, but it is *useful* to learn them, as this will save a lot of time.

(d) *Special results.* You will often read in textbooks an equation which is the end-product of an analysis of a special situation which is not often encountered. *Such an equation should not be learned.* Rather, you should master the principles used to derive it.

Summary

When you meet a particular equation for the first time, decide whether it is

- (a) a definition,
- (b) a law,
- (c) an important *idea* which you will want to use repeatedly, or
- (d) a specialized result.

The equations in boxes in this book fall into the first three categories: they alone should be learned.

1.4 THE ORDER IN WHICH PHYSICAL QUANTITIES ARE DEFINED

The logical structure of physics, in some subject areas particularly, is based on a distinct hierarchy, and it is important to appreciate the order in which quantities have been defined. For example if we choose to define magnetic flux Φ by the scalar product $\Phi = \mathbf{B} \cdot \mathbf{A}$, then we cannot later define magnetic flux density \mathbf{B} by the quotient $\mathbf{B} = \Phi/\mathbf{A}$. The latter equation can be regarded as a satisfactory definition of \mathbf{B} only if both Φ and \mathbf{A} have already been defined by some other means. (There exists an alternative approach to electromagnetism in which one defines Φ from

$\mathcal{E} = -d\Phi/dt$, and then $\mathbf{B} = \Phi/\mathbf{A}$ becomes acceptable as the formal definition of \mathbf{B} .)

The schemes which follow show a logical order of defining physical quantities in five branches of physics. Every equation shown is a *definition*, and no attempt has been made to relate quantities by other equations. For simplicity the definitions are given in an elementary form. The boxes of the fundamental quantities (the logical starting points) are shown in grey.

Linear dynamics

(See scheme 1 opposite.)

Notes

(a) No arrows are shown for mass ratio. The equation $\mathbf{F} = m\mathbf{a}$ defines the measure of *force*, and is also used to measure *inertial mass* through $\mathbf{F} = m_1\mathbf{a}_1 = m_2\mathbf{a}_2$.

(b) The definitions of rotational dynamics are not included—they follow a similar scheme.

Thermal properties of matter

(See scheme 2 opposite.)

Electrostatics

(See scheme 3 on p. 8.)

Notes

(a) In this scheme quantities are introduced from mechanics only where they make a significant contribution to the understanding of the whole.

(b) Note carefully that electric charge is not a fundamental quantity, since it is derived from electric current through $Q = \int I dt$.

Current electricity

(See scheme 4 on p. 8.)

Notes

(a) Mechanical quantities are introduced as in the previous scheme.

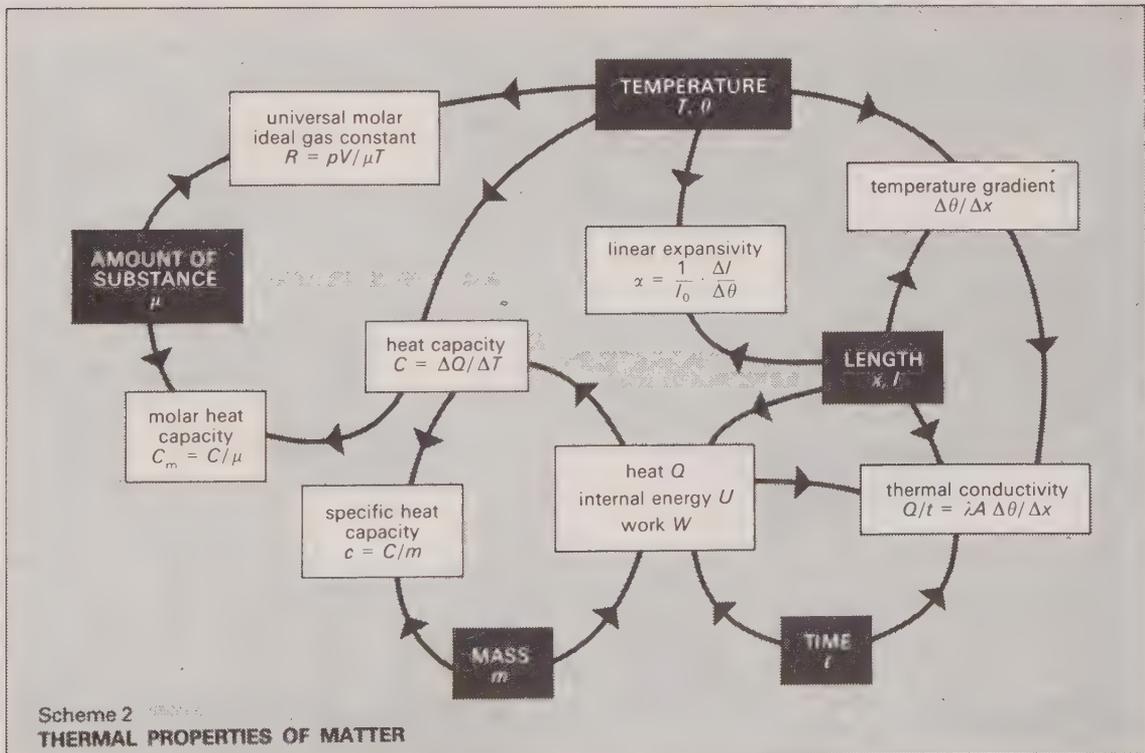
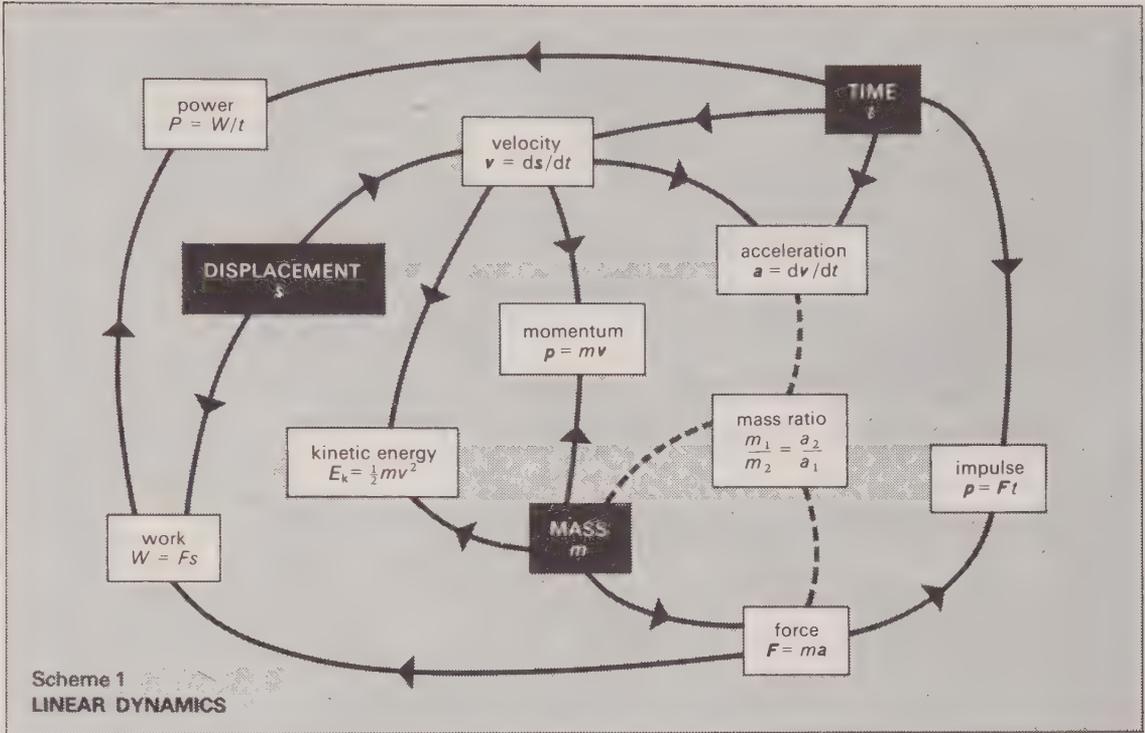
(b) The defining equations for \mathcal{E} and V are symbolically identical, and it is important to identify carefully the quantity represented by W in each definition.

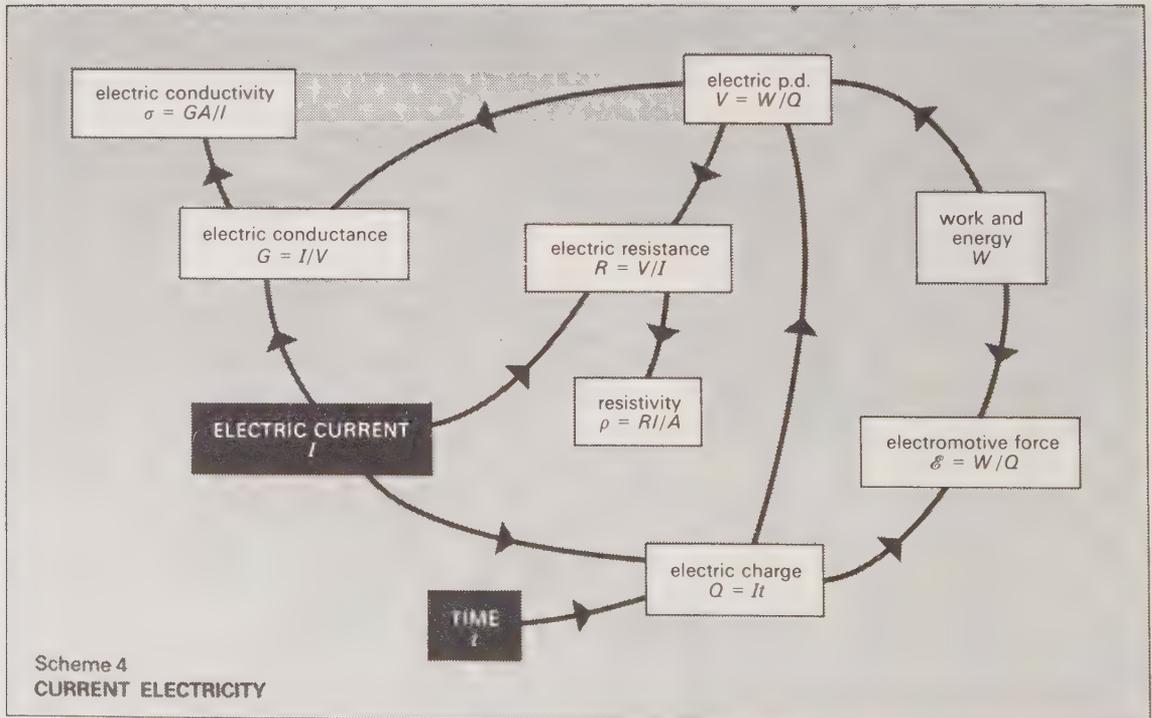
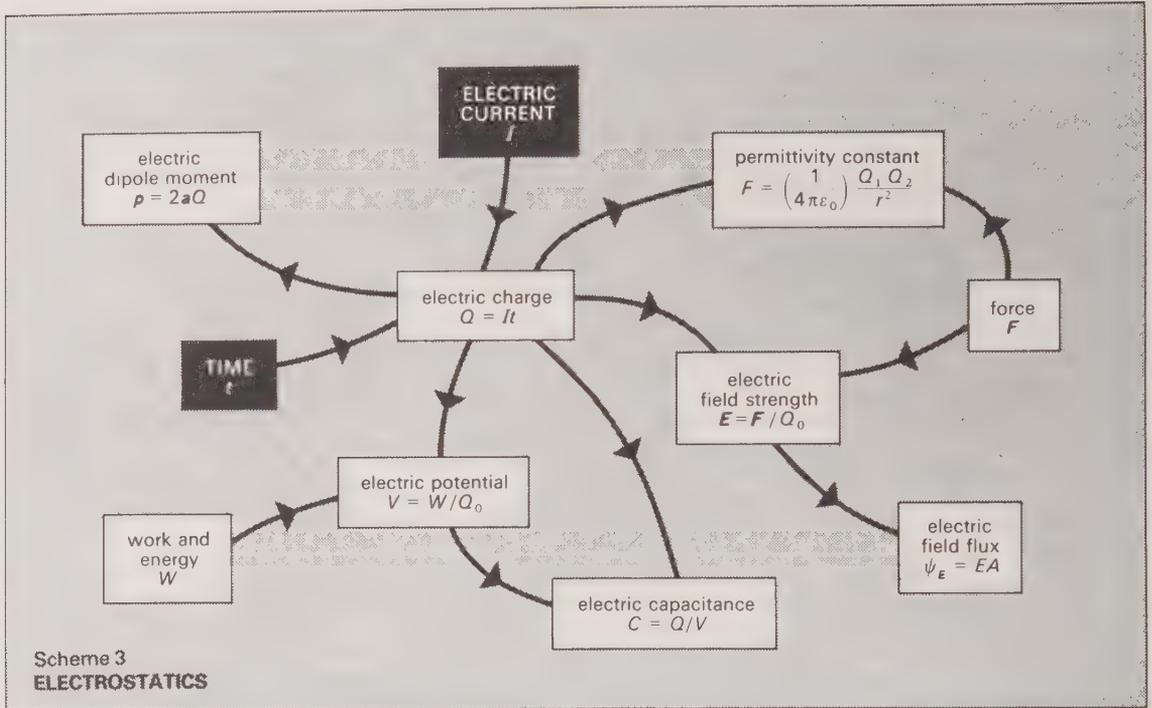
The magnetic field

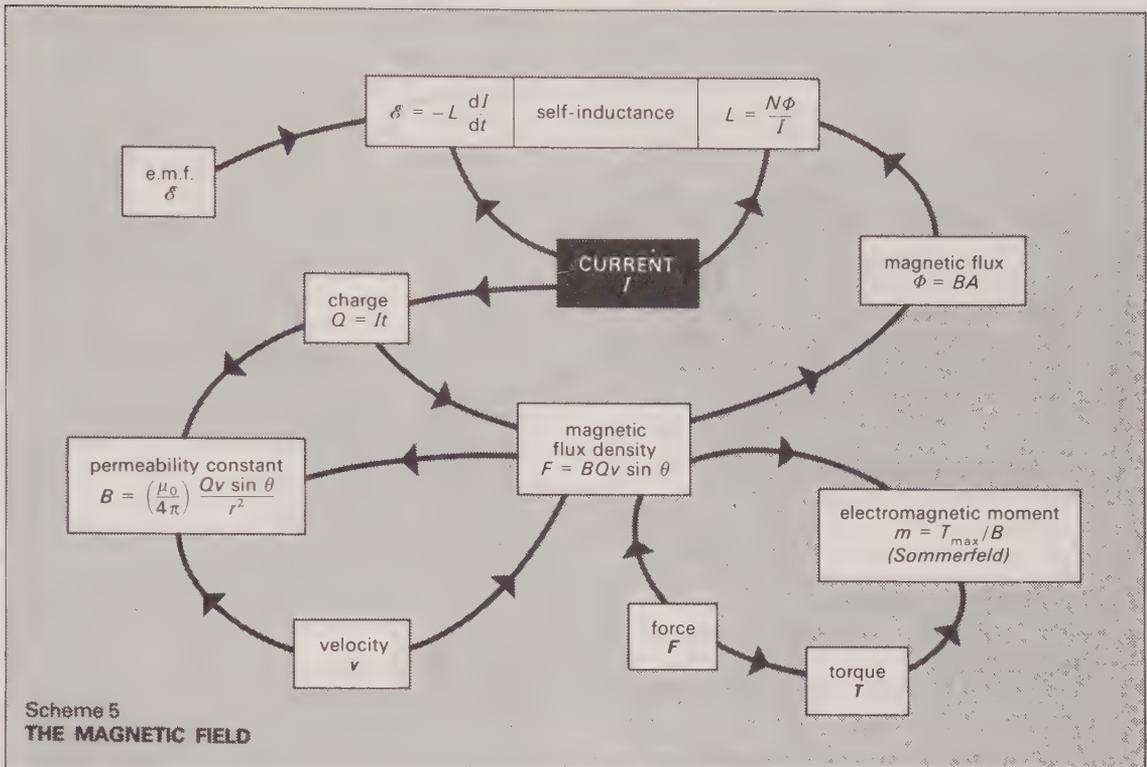
(See scheme 5 on p. 9.)

Notes

(a) There are several alternative approaches to electromagnetism. The one used here is that recommended by the *Association for Science Education*.







(b) The equations $F = BIl \sin \theta$ and

$$\delta B = \left(\frac{\mu_0}{4\pi}\right) \frac{I \delta l \sin \theta}{r^2}$$

are exactly equivalent to the two used above to define B and μ_0 respectively.

1.5 UNITS AND THE REALIZATION OF STANDARDS

The basic SI units

SI Units are those of *Le Système International d'Unités*, adopted in 1960 by the *Conférence Générale des Poids et Mesures*. The system

(a) is based on the metre, kilogram, second, ampere, kelvin, mole, and candela, and

(b) uses other units which can be obtained from the basic units by multiplication and division (i.e. without including any numerical factors).

Metre: the metre is the length equal to 1 650 763.73 wavelengths in a vacuum of a specified radiation from an atom of krypton-86.

Kilogram: the kilogram is the mass of the international platinum-iridium prototype.

Second: the second is the duration of 9 192 631 770 periods of a specified radiation from an atom of caesium-133.

Ampere: the ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 m apart in a vacuum, causes each to exert a force of 2×10^{-7} N on one metre length of the other (see p. 416).

Kelvin: the kelvin is the fraction $\frac{1}{273.16}$ of the thermodynamic temperature of the triple point of water.

Mole: the mole is the amount of substance of a system which contains as many elementary units as there are carbon atoms in 12×10^{-3} kg of carbon-12.

Thus a mole contains about $6.022\,045 \times 10^{23}$ specified elementary units, such as atoms, molecules, ions, electrons, etc.

Examples

(a) A mole of electrons has a charge of $(6.0 \times 10^{23}) \times (-1.6 \times 10^{-19})$ C, the size of which is called the *Faraday constant*.

(b) A mole of protons has a mass of $(6.0 \times 10^{23}) \times (1.7 \times 10^{-27})$ kg.

(c) A mole of any ideal gas at s.t.p. has a volume of 22.4×10^{-3} m³.

One would refer respectively to the molar charge, molar mass or molar volume of the substance.

The term **molar** means 'divided by amount of substance'.

(*Candela* is not used in this book: it is a unit of luminous intensity.)

Why these units have been chosen

Scientists continually exchange information, so their units must be understood with minimum explanation, and unambiguously. These ends are achieved by

- (a) a small number of units which are
- (b) internationally accepted, because they are
 - (i) easily compared
 - (ii) easily reproduced, and
 - (iii) invariant in time.

To satisfy (iii), standards are frequently based on the properties of atoms: for this reason the present kilogram may be replaced at some future date by a property of the proton.

SI units with special names

Some derived SI units are relatively complex; and are conveniently given a name whose abbreviation is simpler than the unit expressed in terms of the basic units.

Examples

The name *farad* (F) is given to $\text{m}^{-2} \text{kg}^{-1} \text{s}^4 \text{A}^2$ (p. 358).

The name *watt* (W) is given to $\text{m}^2 \text{kg} \text{s}^{-3}$ (p. 47).

The following units are defined at appropriate places in the text: hertz (Hz), newton (N), joule (J), pascal (Pa), coulomb (C), volt (V), ohm (Ω), siemens (S), weber (Wb), henry (H), tesla (T), becquerel (Bq) and gray (Gy). They are also listed in appendix A at the end of this book.

Standard prefixes

The prefixes attached to a unit abbreviation have the meanings shown in the table below.

Fraction	Prefix	Symbol
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a
Multiple	Prefix	Symbol
10^3	kilo	k
10^6	mega	M
10^9	giga	G
10^{12}	tera	T
10^{15}	peta	P
10^{18}	exa	E

Units in calculations

This book uses a **coherent** system of units, which means that any physical quantity has strictly only one correct unit. In using any equation it is possible to adopt one of two methods:

- (a) one can replace each symbol by the numerical measure for each quantity *expressed in the basic units*, and *assume* the unit for the calculated quantity (which will be unambiguous), or
- (b) one can replace each symbol by both its *number* and its *unit abbreviation*: since the unit abbreviation can be treated algebraically, one can *calculate* the *unit* of the required physical quantity, as well as the number. This method has two advantages.

- (i) It indicates at once (by a wrong unit) if an incorrect equation has been used.
- (ii) It causes the student to revise constantly the definitions of derived units, which leads to invaluable familiarity.

Method (b) is strongly recommended.

Laboratory measurements are frequently expressed in a non-SI unit (such as $^{\circ}\text{C}$). This does not matter provided one converts to the SI unit before replacing a symbol in an equation by its number and unit abbreviation.

Example. Calculate the electrical energy converted to thermal energy when a current of 2.0 mA passes through a resistor of resistance 3.0 k Ω for a time of 10 minutes.

$$\begin{aligned}
 W &= I^2 R t && \text{(p. 385)} \\
 &= (2.0 \times 10^{-3} \text{ A})^2 \times (3.0 \times 10^3 \Omega) \times (600 \text{ s}) \\
 &= 7.2 \text{ A}^2 \times \frac{\text{V}}{\text{A}} \times \text{s} \\
 &= 7.2 \frac{\text{C}}{\text{s}} \times \frac{\text{J}}{\text{C}} \times \text{s} \\
 &= 7.2 \text{ J}
 \end{aligned}$$

7.2 joules of energy are converted.

The realization of standards

Metre: Realization involves marking off a length containing the stated number of wavelengths. Accurate reproduction of the krypton-86 orange line at 606 nm requires carefully specified conditions such as type of lamp, suitable temperature of operation to reduce *Doppler* broadening, current density etc. Measurements of particular distances can be made using interferometers or lasers, which determine the number of waves covering the distance.

Kilogram: Copies of the prototype are sent periodically to *Sèvres* for comparison and these can be adjusted easily after carefully weighing on appropriate balances. Masses of the order of 1 kg can be compared to 1 part in 10^9 .

Second: The hyperfine transition of the caesium-133 atom is observed in an atomic beam apparatus. The magnetic deflection of the atoms is reversed when the frequency applied from a quartz oscillator equals that of the spectral line. Such equipment is required only rarely as many countries broadcast time signals which are accurate to 1 part in 10^{11} .

Ampere: The arrangement stated in the definition is not a practical one though it is consistent with electromagnetic theory. Current balances or dynamometers can be used to measure the forces exerted by currents very accurately. Good results can also be achieved using standard resistors and *Weston* cells, although such methods are not absolute. Nuclear magnetic resonance can also be utilized; one measures the frequency of precession of protons in a magnetic field produced by a current-carrying coil of suitable shape.

Kelvin: The International Practical Temperature Scale, I.P.T.S.-68, is a practical scale which is easily and accurately reproducible, and which gives thermodynamic temperatures. The defining fixed points are established by realizing specified equilibrium states between phases of pure substances (p. 182).

Mole: Although defined in terms of a number of entities, the mole is usually realized by weighing rather than counting. Volume comparison can be used for gases, and charge measurements can be used in electrolytic reactions.

1.6 MEASUREMENT

Measurement is that process by which we compare a physical quantity with the unit chosen to express that quantity. The statement 'The mass of the body is 2.0 kg' means that the body has a mass 2.0 times the mass of the standard kilogram: we arrive at the figure 2.0 by a specified procedure (detailed, in this instance, on p. 36).

To measure a particular physical quantity, we look for some physical phenomenon that depends on the quantity. Thus an electric current establishes a magnetic field; a change of temperature may cause a change of electrical resistance. We then choose an instrument whose design is based on a convenient phenomenon which

(a) can be applied in the range we are using, and

(b) gives the sensitivity that we require.

In this section we discuss the measurement of some fundamental quantities.

(A) Measurement of length

We choose an instrument appropriate to the order of magnitude, the nature of the length (e.g. it might be an internal tube diameter), and the sensitivity required.

(a) **Rulers** are used for lengths exceeding about 50 mm, and are probably accurate to about 0.5 mm.

(b) **Micrometer screw gauges** are frequently used for small lengths. If the screw has a pitch of 0.50 mm, and the sleeve has 50 divisions, each division corresponds to 0.01 mm (fig. 1.2).

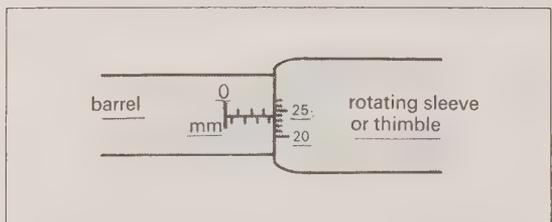


Fig. 1.2. A micrometer screw gauge indicating 3.74 mm.

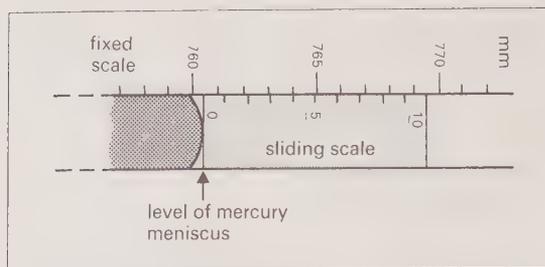


Fig. 1.3. A vernier scale indicating 760.4 mm (shown on its side for convenience).

(c) A **vernier scale** is a device which eliminates guesswork when we interpolate between scale divisions. Vernier scales are frequently used on calipers, barometers, travelling microscopes and spectrometers. Fig. 1.3 shows such a scale on a Fortin barometer.

(B) Measurement of mass

Bodies possess both *inertial* mass and *gravitational* mass: experiment shows that these masses are equivalent to 1 part in 10^{12} (p. 88).

(a) **Inertial mass** can be measured by methods which do not require the existence of a gravitational field, such as

- (i) the **inertial balance** of fig. 7.3 (p. 68),
- (ii) trolley experiments, involving the mutual interaction of two masses (p. 36).

In each method one compares the mass to be measured with that of (a replica of) the standard kilogram.

(b) **Gravitational mass** is deduced from a measurement of the *weight* of a body (p. 37).

It is highly inconvenient to measure inertial mass *directly*, but relatively easy to measure or compare weights. The method adopted to find inertial mass is (usually)

- (1) to measure (perhaps by direct comparison) a body's weight, then
- (2) to calculate its gravitational mass, and so
- (3) deduce its inertial mass, which is equivalent.

(C) Measurement of time

Time is measured by a *clock*, a device which has some repetitive or periodic property, and which has been calibrated by comparison with a standard. The second used to be defined as $1/31\,556\,925.974\,7$ of the duration of the solar year 1900, but this time interval is *not reproducible*, and has therefore been replaced by the definition on p. 9.

One chooses a clock consistent with the requirements of the measurement to be made. Clocks have been based on the following properties:

- (a) radioactive decay processes (p. 530),
- (b) the rotation of the Earth on its axis,
- (c) the oscillations of pendulums and balance wheels (as in the wrist watch),
- (d) the oscillation of electrons in a circuit which carries an a.c. (as in the ticker-tape timer),
- (e) the oscillation of a nitrogen atom within an ammonia molecule (an **atomic clock**).

The *quality* of a clock is determined by how well it keeps time with other clocks, and with the standard. An atomic clock is very little affected by any external influence (such as change of temperature), and is therefore very reliable.

(D) Measurement of weight

Measurement of weight implies the comparison of a body's weight with the standard force unit (the newton).

(a) A **spring balance** is calibrated by known *forces*, and can then be graduated in *newtons*. Such a balance will then record the weight of a body to which it is attached provided the body and balance are in relative equilibrium. The balance registers the apparent weight of the body (not corrected for the Earth's rotation) *wherever it is used*.

(b) A **beam balance** or a **lever balance** compares the pull of the Earth G_u on the **unknown** mass m_u with that G_k on known masses m_k (colloquially called 'weights'). Both masses are at the same place, so for a beam balance whose arms are of equal length

$$G_u = G_k$$

$$m_u g = m_k g$$

$$m_u = m_k$$

These balances compare unknown masses with standard masses by comparing their weights. If the value of g is known *at that place*, then the weight can be calculated from $G = mg$.

(The **ballistic balance** (p. 52) compares inertial masses, but utilizes the gravitational field for the measurement of horizontal velocity.)

1.7 ORDERS OF MAGNITUDE

Developing a feeling for orders of magnitude

In physics the phrase 'correct to an order of magnitude' means that the value quoted is reliable to within a factor of ten or so, and for many purposes

the information may be sufficiently precise in this form. It is most important for physicists to acquire a working knowledge of the size of typical quantities. This enables them

- (a) to judge the plausibility of any given quantity, and
- (b) to estimate the possible sizes of further quantities.

Even the best physicists make elementary errors in their working, but only a bad physicist is happy to present a solution whose size is patently absurd. It is vital to develop an awareness of reliable information sources from which orders of magnitude can be obtained.

Orders-of-magnitude tables

The tables below and overleaf show the number of particles to be found in various bodies, and, separately, the ranges of distances, masses and times found in the Universe. In both tables the numbers are presented on a logarithmic scale.

Parent body	Number of particles
Universe	10^{80}
Sun	10^{70}
Earth	10^{60}
Earth's atmosphere	10^{50}
human breath	10^{40}
elementary particle	10^{30}
	10^{20}
	10^{10}
	10^0

Energy

One of the principal interests of physicists is energy, and its conversion from one form to another. One reason for the concept of energy holding such an important place is that we can often use it to decide whether a particular physical change is likely to occur. A physical system achieves equilibrium by making its p.e. a minimum. Thus a liquid surface contracts, an electron in an atom moves as close as possible to the nucleus, and a radioactive nucleus disintegrates, each making an attempt to reduce the p.e. of the system.

In the SI all energies are expressed in *joules*, and this has the enormous advantage of revealing the true comparative values of energies previously expressed relative to other quantities (such as the electronvolt). The energy table on p. 15 should be

studied closely, since many useful conclusions may be drawn from it. For example one can predict that photons of violet light will cause photoelectric emission from caesium, but not from tungsten; or that air is ionized by photons from the middle ultraviolet, but not by those of lower frequency.

1.8 TREATMENT OF ERRORS

We do not expect the result of a measurement to be the *true* value of the measured quantity. There always exists some uncertainty, which is usually referred to as **experimental error**. An experiment is successful if it is reliable.

Example. The results of two experiments to measure the specific heat capacity of water were quoted as

- (1) $c = (4.0 \pm 0.2) \times 10^3 \text{ J kg}^{-1} \text{ K}^{-1}$,
- (2) $c = (4.10 \pm 0.01) \times 10^3 \text{ J kg}^{-1} \text{ K}^{-1}$.

The second experiment is less successful, since the claim that the error was $10 \text{ J kg}^{-1} \text{ K}^{-1}$ is obviously not justified (c for water = $4.18 \times 10^3 \text{ J kg}^{-1} \text{ K}^{-1}$).

In this section we discuss (non-rigorously) causes of experimental error, and ways of estimating its size.

(A) How errors arise

The term *error* excludes *mistakes* which can be put down to the individual. Mistakes can arise through

- (i) misreading scales,
- (ii) faulty arithmetic,
- (iii) trying to apply a theory where it does not hold,
- (iv) faulty transcription, etc.

Experimental errors are of two main kinds (fig. 1.4, p. 16).

(a) A **random error** is one that has an equal chance of being negative or positive. It can be caused by

- (i) lack of perfection in the observer,
- (ii) the sensitivity of the measuring instrument,
- (iii) a particular measurement not being reproducible.

The random error is revealed by repeated observation of a particular quantity: this procedure also helps to reduce its effect.

(b) A **systematic error** causes a random set of measurements to be spread about a value other than the true value. This effect can result from (for example)

Number representing $l/m, m/kg, \Delta t/s$	Distance l	Body whose mass m is considered	Time interval Δt	
10^{28}	distance to furthest photographed galaxy	Sun	age of Earth \approx mean life of ${}^{238}_{92}\text{U}$	
10^{24}		Earth		
10^{20}	radius of our galaxy	large ship		
10^{16}	distance to nearest star			
10^{12}	radius of solar system	year		
10^8	radius of Moon	car		day
10^4		man's height		time between heart beats
$1 = 10^0$	thickness of Al foil	oil-drop from atomizer		time period of highest audible note
10^{-4}	wavelength of visible light			
10^{-8}	diameter of molecule	blood corpuscle		duration of photon emission
10^{-12}	diameter of nucleus	protein molecule	time period of visible light vibration	
10^{-16}				
10^{-20}		proton	time for light to cross an elementary particle	
10^{-24}				
10^{-28}		electron		
10^{-32}				

(i) an instrument having a **zero error** (in which case the systematic error may be constant),

(ii) an instrument being incorrectly calibrated (such as a slow-running stopclock),

(iii) the observer *persistently* carrying out a mis-timed action (e.g. in starting and stopping a clock).

Such systematic errors are *not* revealed by repeated measurement (fig. 1.4, p. 16), but they can be eliminated or corrected by

(i) varying the conditions of measurement, or

(ii) suitable treatment of the observations (see the graph on p. 302).

Description of energy W	W/J	Description of energy W	W/J
binding energy of Earth–Sun system	10^{33}	maximum energy of proton in CERN synchrotron	$\frac{1 \mu J}{10^{-8}}$
translational k.e. of Moon	10^{28}	rest-mass-energy of 1 unified atomic mass unit	1 nJ
energy radiated by the Sun in 1 second	10^{26}	energy released by fission of 1 uranium nucleus	1.6×10^{-10}
energy received per day on Earth from Sun	10^{22}	typical binding energy per nucleon	3×10^{-11}
estimated human energy requirement per year (1980)	10^{20}	minimum energy of γ -ray photon for ion-pair production	1.3×10^{-12}
energy associated with a strong earthquake	10^{20}	rest-mass-energy of 1 electron	1 pJ
energy released by annihilation of 1 kg of matter	9×10^{16}	photon energy of typical X-ray	1.6×10^{-13}
energy released by fission of 1 mol of ^{235}U	10^{13}	photon energy of middle u.v.	1 fJ
	1 TJ	average energy to create 1 ion-pair in air	10^{-17}
energy dissipated by a lightning discharge	10^{10}	ionization energy of hydrogen atom	5×10^{-18}
energy released by combustion of 1 kg of petrol	1 GJ	width of forbidden region between energy bands in diamond crystal	2×10^{-18}
energy converted by a 1 kW heater operating for 1 hour	10^8	change of binding energy per atom in a typical chemical change	} 1 aJ
energy required to charge a car battery	3.6×10^6	work function energy of tungsten	
energy provided by a slice of bread	1 MJ	photon energy of visible light (violet)	5×10^{-19}
k.e. of a bowled cricket ball	10^5	work function energy of caesium	3×10^{-19}
energy stored in the magnetic field of a 2 H inductor carrying a current of 1 A	10^1	photon energy of visible light (red)	2.5×10^{-19}
energy stored in the electric field of a 1 μF capacitor charged to 200 V	1	width of forbidden region between energy bands in germanium crystal	1×10^{-19}
	2×10^{-2}	photon energy of middle i.r.	2×10^{-20}
recommended maximum ionizing radiation dose to be absorbed per year by 1 kg of living tissue	1 mJ	translational k.e. of ideal-gas molecule at 300 K	6×10^{-21}
	4×10^{-4}	photon energy of microwaves	10^{-23}
	.	photon energy of medium band broadcast radio waves	10^{-27}

In the remainder of this section we assume that this has been done.

An **accurate** experiment is one for which the **systematic error** is relatively small. A **precise** experiment has a small **random error**. These points are illustrated by fig. 1.5 (p. 16).

(B) Estimating the size of an error

Single measurement

The error is guessed using commonsense.

(a) In experiments where a *setting* is required (e.g. using a jockey or an optical bench), the true

position can be located by bracketing techniques. One might find, for an image distance,

$$v = +(316 \pm 10) \text{ mm}$$

(b) Where a scale has to be read, there will be a **sensitivity limit**. Thus we would write:

for an ammeter

$$I = (25.4 \pm 0.2) \text{ mA}$$

or, for a metre rule

$$l = (371.0 \pm 0.5) \text{ mm}$$

The *largest* error (which could be that resulting from the uncertainty of calibration) is the most important.

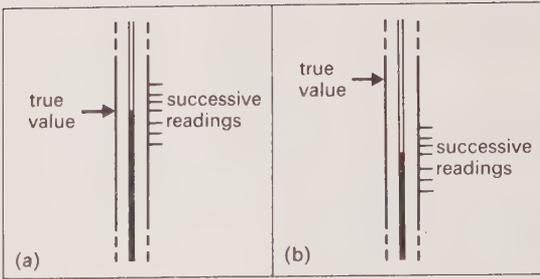


Fig. 1.4. Each bar represents a measurement taken with a thermometer, in which
 (a) only random errors are present,
 (b) a systematic error is superimposed on the random errors.

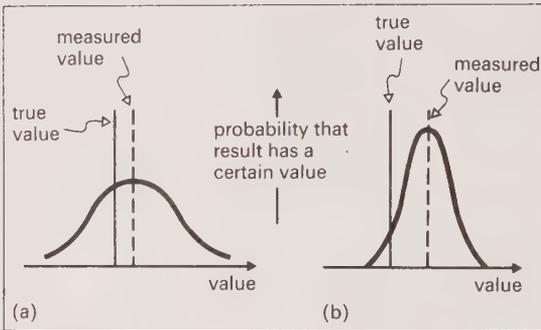


Fig. 1.5. Two examples illustrating the nature of experimental results:
 (a) an accurate experiment of low precision,
 (b) a less accurate but more precise experiment.

Several repeated measurements

When n observations are taken of a single quantity, the likely difference between the mean value and the true value is reduced. When n is large, statistical techniques are used to find a *probable range of error*. They are not considered in this book, which deals only with the *maximum error* calculated in a non-rigorous way.

(C) Percentage errors

Suppose a reading x is obtained of a quantity whose true value is X . Then

$$x = X \pm e$$

where e is the **absolute error**.

The fractional error

$$f = \frac{e}{X}$$

and the **percentage error**

$$p = \frac{e}{X} \times 100$$

Usually $f \ll 1$, and $p \ll 100$.

To find the *maximum error in compound quantities*, we proceed as follows:

(a) Sum and difference

We add the *absolute errors*.

If $S = a + b$

then $\Delta S = \Delta a + \Delta b$

where ΔS refers to the absolute error of the sum.

If $D = a - b$

then $\Delta D = \Delta a + \Delta b$

The error, which could be either side of the true value, is still *added*.

(b) Product and quotient

We add the *fractional or percentage errors*.

If $P = ab$

$$\frac{\Delta P}{P} = \frac{\Delta a}{a} + \frac{\Delta b}{b}$$

$$p_P = p_a + p_b$$

where p_P refers to the percentage error of the product.

If $Q = \frac{a}{b}$

$$p_Q = p_a + p_b$$

Examples

(a) $P = a^n$

$$p_P = np_a$$

(b) $Q = \frac{x^2}{y}$, and $p_x = 2\%$
 $p_y = 3\%$

$$p_Q = (2 \times 2 + 3)\% = 7\%$$

(D) Estimating the uncertainty in an experiment

To investigate whether an experiment is successful, we can proceed as follows:

(1) Repeat the experiment several times, and calculate the required quantity independently for each set of readings.

For example, use

$$\eta = \frac{\pi r^4 \Delta p t}{8 l V} \quad (\text{p. 167})$$

several times, and obtain a set of values for η . Analyse them to obtain a maximum (or a probable) error.

(2) Estimate (as in (B)) the maximum errors in the individually measured quantities r , Δp , t , l and V , and calculate their percentage errors.

Since

$$\begin{aligned} \ln \eta &= \ln \left(\frac{\pi}{8} \right) + 4 \ln r + \ln \Delta p + \ln t \\ &\quad - \ln l - \ln V \\ \frac{\Delta \eta}{\eta} &= 0 + \frac{4 \Delta r}{r} + \frac{\Delta(\Delta p)}{\Delta p} + \frac{\Delta t}{t} - \frac{\Delta l}{l} - \frac{\Delta V}{V} \end{aligned}$$

To find p_η , we *add* all the percentage errors, noting factors such as the 4 for p_r . (In fact the measurement of r , in this experiment, has more effect on the uncertainty in η than all the other readings put together.)

(3) Compare the errors calculated in (1) and (2). If they disagree significantly, then the experiment is not satisfactory.

(E) Treatment of graphs

One purpose of a graph is to display results: it can also show an estimated error range for individual points (fig. 1.6).

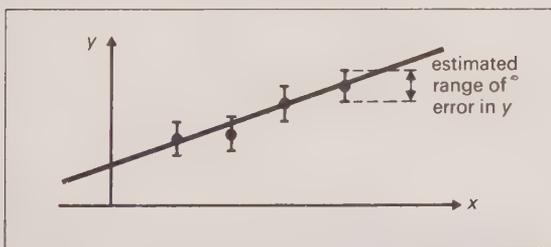


Fig. 1.6. Display of range of error on a graph.

Any line which fits the experimental points within their estimated ranges of error is in agreement with the experiment.

If a point is displaced significantly from the line, we can choose between

(a) the estimated limits of error being too optimistic,

(b) there being a *mistake* (in which case the point can be disregarded), or

(c) the theory under test not being applicable under the conditions of the experiment.

Estimating an error from a graph

(a) **Intercept.** See how far the line can be displaced without going seriously outside the error ranges.

(b) **Slope.** Use the same techniques to measure the maximum and minimum slopes. If the graph is insufficiently sensitive, then we can pair the points, and evaluate the slopes analytically. Thus for 6 points we would calculate slopes:

m_1	from points	1 and 4
m_2		2 5
m_3		3 6

and then work out a mean slope.

It is important to keep the treatment of errors in proportion. For example it is futile to carry out a lengthy calculation of probable error, when the time could be better spent in reducing its main cause.

1.9 SCALAR AND VECTOR QUANTITIES

Scalars and vectors

Scalar quantities are those which are completely specified by a physical magnitude (expressed by the product of a positive or negative number and a unit) alone. They can be added *algebraically*.

Vector quantities require, in addition, a direction to be specified. The magnitude of a vector is inherently positive. In this book the symbol is printed in *bold italic*, unless we refer to magnitude only.

Examples in mechanics

Scalar quantity	Vector quantity
length l	displacement s
mass m	area A
time t	force F
volume V	velocity v
speed v	acceleration a
energy W	momentum p
pressure p	torque T
density ρ	angular momentum L

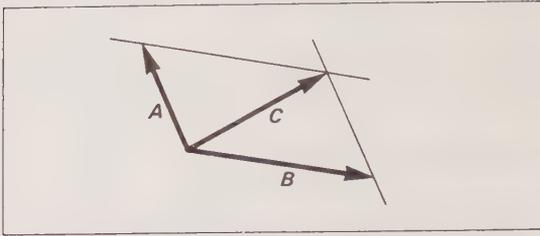


Fig. 1.7. Vector addition: $A + B = C$.

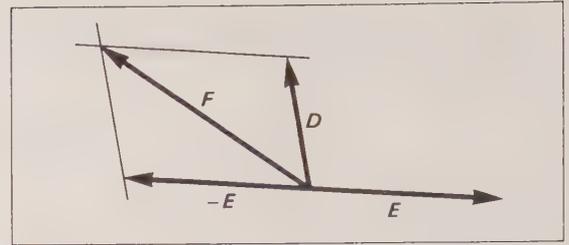


Fig. 1.8. Vector subtraction: $F = D + (-E) = D - E$.

For equations in this book we adopt the following procedure:

(i) Vector quantities are printed in bold type if the *directions* of the quantities involved are *along the same straight line*.

(ii) They are represented by their magnitudes only (in italic) if, perhaps as a consequence of multiplication, the vector quantities involved lie along different directions.

Examples

(a) $F = ma$ (p. 36)

(b) $F = BQv$ (p. 401)

A **vector** is a line drawn to represent a vector quantity in

- (a) *direction*, by its direction, and
- (b) *magnitude*, by its length.

When the answer to a problem is a vector quantity, its *direction* (not just its magnitude) must always be given, even if not asked for explicitly. Usually this is done best by a labelled diagram.

Adding and subtracting vectors

(a) Vectors must be *added* by the **parallelogram law** (fig. 1.7) which ensures that their directions, as well as their physical magnitudes, are taken into account. They may *not* be added algebraically. Vectors may be added in any order.

(b) The *negative* of a vector is that vector with its direction reversed. So to *subtract* vector E from vector D , we reverse the direction of E before adding by the parallelogram law (fig. 1.8). (For an example see p. 30.)

Resolving vectors

In fig. 1.7 the vectors A and B have been compounded, or added, to give their **resultant** C .

(a) The reverse process, in which C is replaced by two vectors, is called finding the **components** of the vector. There is an infinite number of paral-

lelograms for which C could be a diagonal: there is an infinite number of ways in which C can be split into two components.

(b) When we find components X and Y which are *perpendicular*, then they are called **resolved parts** of the vector Z (fig. 1.9), and the process is called **resolving**.

The effect of Y along the direction Ox is zero, so X represents the *total effective value* of Z in that direction, where

$$X = Z \cos \theta$$

Similarly

$$Y = Z \sin \theta$$

This process is particularly useful for forces (p. 35).

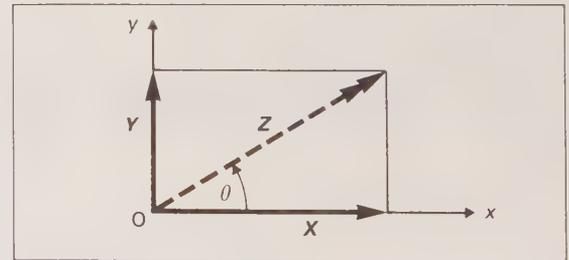


Fig. 1.9. X and Y are the resolved parts of Z .

► Multiplying vectors

Multiplication of a vector by a scalar

The result of multiplying the vector V by the scalar S is a *vector* of

- (a) magnitude SV , and
- (b) the same direction as V .

Multiplication of a vector by a vector

When a vector multiplies a vector, the product can be

- (a) a scalar, if the dot product ($A \cdot B$) is used, or
 (b) a vector, if the cross product ($A \times B$) is used.

In this case the order of multiplication (e.g. $A \times B$ or $B \times A$) determines the *sense* of the product. More properly such quantities are called **pseudovectors**, or **axial vectors**. The direction associated with them is that of an axis of rotation, and this enables them to be distinguished from ordinary (or **polar**) vectors.

We can choose (by convention) one of the two directions of the axis to represent the direction of the pseudovector—thus a pseudovector has a **sense** associated with it.

Example. The product of a displacement r and a force F can result in

(a) *work* (p. 42), in which we are interested in the resolved parts of r and F along the *same* direction, or

(b) *torque* (p. 65), in which we are interested in their resolved parts along *perpendicular* directions.

Work is a scalar, but torque is a vector whose direction is perpendicular to the plane which contains r and F , and whose sense is found conventionally by using a right-hand corkscrew rule.

1.10 THE USES OF DIMENSIONS

In a correct physical equation we can equate both number *and* unit for each term that appears. If we could not equate the unit, then a change of the system of units might result in the numbers changing by different factors, which would invalidate an equation whose numbers were previously equal.

Each term in a correct physical equation must have the same dimensions. Use of this fact is called the **method of dimensions**.

(A) Conversion of units

When several systems of units were in common use, the method of dimensions gave a quick way of converting units for complex derived quantities (such as thermal conductivity) from one system to another. This procedure should no longer be necessary for the beginning student.

(B) To check equations

Since physical equations are dimensionally homogeneous, terms which are incorrect (other than by a dimensionless factor) can quickly be detected.

Examples

(a) $T = \frac{1}{5} \sqrt{l/10}$ is an approximate equation for the period T of a simple pendulum of length l . It is not dimensionally homogeneous, and can only be applied when T is in seconds and l in mm, i.e. for a *specified set of units*.

(b) Suppose we are told that the speed c of transverse waves along a wire of tension F and mass m is given by

$$c = \sqrt{\frac{F}{m}}$$

$$[c] = [LT^{-1}]$$

$$[F^{1/2}m^{-1/2}] = [MLT^{-2}]^{1/2}[M]^{-1/2} \\ = [L^{1/2}T^{-1}]$$

Because the left- and right-hand sides have different dimensions, we detect a mistake: m is not mass, but mass *per unit length*, and has dimensions $[ML^{-1}]$.

(c) Consider the equation

$$\left(p + \frac{a}{V_m^2}\right)(V_m - b) = RT \quad (\text{p. 234})$$

Problems: (i) Is it dimensionally correct?

(ii) If so, what is the unit for a ?

(i) If the dimensions are right, each side of the equation will have the same unit. Taking the first term in each bracket, the left side has unit

$$(\text{N m}^{-2})(\text{m}^3 \text{ mol}^{-1}) = \text{J mol}^{-1}$$

The right side has unit

$$(\text{J mol}^{-1} \text{ K}^{-1})(\text{K}) = \text{J mol}^{-1}$$

The equation is dimensionally correct.

(ii) The unit for a/V_m^2 must be the same as that for p , since they occur in the same bracket. Therefore a has the unit of pV_m^2

$$= (\text{N m}^{-2})(\text{m}^3 \text{ mol}^{-1})^2 \\ = \text{N m}^4 \text{ mol}^{-2}$$

(d) Consider the equation

$$F' = \left(\frac{\mu_0}{2\pi}\right) \cdot \frac{I_1 I_2}{a} \quad (\text{p. 415})$$

What is the unit for F' ?

It has the same unit as the right side, namely

$$\left(\frac{\text{H}}{\text{m}}\right) \cdot \frac{\text{A}^2}{\text{m}} = \left(\frac{\text{N}}{\text{A}^2}\right) \cdot \frac{\text{A}^2}{\text{m}} \\ = \text{N m}^{-1}$$

F' tells us the force *per unit length* of the wire.

(C) Dimensional analysis

This enables us to predict how physical quantities may be related. Some examples are given in the next section.

1.11 DIMENSIONAL ANALYSIS**Limitations**

(a) The method cannot find the value of any dimensionless constants (such as 2π).

(b) Since in mechanics we have chosen to use only *three* fundamental quantities, [M], [L] and [T], we cannot in general determine how a given quantity depends upon more than *three* other quantities. Nevertheless problems in thermodynamics and electricity, for example, may need the introduction of further fundamental quantities, such as temperature [Θ], amount of substance [N] or electric current [I].

It will then be possible to relate a given quantity to more than three others. (See examples (C) and (D) below.)

(c) Even in mechanics it may be possible to relate a given quantity to more than three others if it is already known before the analysis how some of them *must* be related to one another. (See p. 164.)

(d) We must make a reasonable guess (based on physical intuition or experiment) as to all the dimensional quantities which may be involved in the equation (see p. 165).

(e) Dimensionless combinations of quantities will not be revealed.

(f) We can only discover relationships which involve the product of powers: no information, for example, can be obtained about exponential functions.

Applications

Dimensional analysis is particularly useful at this level for viscosity problems (p. 164), but can be applied in all branches of physics. Examples are found in the text, and below.

(A) Speed of ocean waves

Experiment indicates that the speed c is effectively independent of amplitude, and for long wavelengths is independent of surface tension.

$$\text{Suppose } c = kg^x \lambda^y \rho^z$$

where k is a dimensionless constant.

$$[LT^{-1}] = k[LT^{-2}]^x [L]^y [ML^{-3}]^z$$

By the method of dimensions, the powers of L, M

and T must be the same on each side of the equation.

$$\begin{aligned} (1) [L] & 1 = x + y - 3z \\ (2) [T] & -1 = -2x \\ (3) [M] & 0 = z \end{aligned}$$

We conclude $c = k\sqrt{(g\lambda)}$, and is independent of ρ . (We expect this from physical intuition: the restoring force is gravitational pull, which is proportional to inertial mass.)

We cannot evaluate k by this method.

(B) Oscillation of a drop

Suppose that a *small* liquid drop is disturbed from its spherical shape, and set oscillating. We assume that gravitational effects are negligible (p. 154). We consider the frequency f of oscillation to be controlled by the drop radius r , and the liquid surface tension γ and density ρ .

$$\begin{aligned} \text{Put } f &= k\gamma^x r^y \rho^z \\ [T^{-1}] &= k[MT^{-2}]^x [L]^y [ML^{-3}]^z \end{aligned}$$

Comparing indices

$$\begin{aligned} (1) [L] & 0 = y - 3z \\ (2) [M] & 0 = x + z \\ (3) [T] & -1 = -2x \end{aligned}$$

$$\text{We conclude } f = k \sqrt{\frac{\gamma}{\rho r^3}},$$

a result which would be very difficult to obtain by any other method.

Physically, it is of the form

$$f \propto \sqrt{\left(\frac{\text{elasticity factor}}{\text{inertial factor (mass)}} \right)}$$

See the discussion on p. 82.

(C) Thermal conductivity of an ideal gas

It is thought that the thermal conductivity k of an ideal gas will depend upon the mean molecular speed \bar{c} , the density ρ , the specific heat capacity at constant volume c_v , and the molecular mean free path λ . Find a possible relationship between these five quantities.

$$\text{Suppose } k = (\text{constant})(\bar{c})^w \rho^x c_v^y \lambda^z$$

$$[MLT^{-3}\Theta^{-1}] = [LT^{-1}]^w [ML^{-3}]^x [L^2T^{-2}\Theta^{-1}]^y [L]^z$$

Comparing indices

$$\begin{aligned} (1) [M] & 1 = x \\ (2) [\Theta] & -1 = -y \\ (3) [T] & -3 = -w - 2y \\ (4) [L] & 1 = w - 3x + 2y + z \end{aligned}$$

These equations have the solution

$$w = x = y = z = 1,$$

and the constant can be shown by other means to be $1/3$. We conclude

$$k = \frac{1}{3} \bar{c} \rho c v \lambda$$

The result is important for the correlation of the transport phenomena (p. 203).

Note that, since the problem involved *four* fundamental quantities, we were able to relate k to *four* other variables.

(D) Energy transfer by electromagnetic waves

It is thought that the rate of flow of electromagnetic energy P through an area A placed normal to an electromagnetic wave will depend upon μ_0 , and upon the instantaneous values of the electric and magnetic fields, E and B . Find a possible relationship between these five quantities.

$$\text{Suppose } P = k \mu_0^w E^x B^y A^z$$

$$[\text{ML}^2\text{T}^{-3}] = k [\text{MLT}^{-2}\text{I}^{-2}]^w [\text{MLT}^{-3}\text{I}^{-1}]^x \\ \cdot [\text{MT}^{-2}\text{I}^{-1}]^y [\text{L}^2]^z$$

Comparing indices

$$\begin{aligned} (1) [\text{M}] \quad & 1 = w + x + y \\ (2) [\text{L}] \quad & 2 = w + x + 2z \\ (3) [\text{T}] \quad & -3 = -2w - 3x - 2y \\ (4) [\text{I}] \quad & 0 = -2w - x - y \end{aligned}$$

These equations have the solution

$$w = -1, \text{ and } x = y = z = 1.$$

We conclude

$$P = kEBA/\mu_0,$$

or that the intensity

$$\frac{P}{A} = k \frac{EB}{\mu_0}$$

The dimensionless constant can be shown by other means to be 1, and the vector intensity (P/A) is referred to as the **Poynting vector**, $(\mathbf{E} \times \mathbf{B})/\mu_0$.

1.12 SOME USEFUL MATHEMATICS

Page numbers are references to places in the text where a particular point of mathematics is used.

(A) Algebra

Binomial Theorem

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{1 \times 2} x^2 \\ + \frac{n(n-1)(n-2)}{1 \times 2 \times 3} x^3 + \dots$$

If $x \ll 1$, then

$$\begin{aligned} (1+x)^n &\approx 1 + nx \\ (1+x)^{-n} &\approx 1 - nx \end{aligned} \quad (\text{p. 122})$$

These useful approximations are valid when x^2 is negligible. They lead to

$$\begin{aligned} \frac{1+\alpha}{1+\beta} &\approx (1+\alpha)(1-\beta) \\ &\approx 1 + (\alpha - \beta) \end{aligned}$$

for small α and β .

Quadratic equations

The equation

$$ax^2 + bx + c = 0$$

has the solution

$$x = \frac{-b \pm \sqrt{(b^2 - 4ac)}}{2a} \quad (\text{p. 263})$$

(B) Logarithms

Common logarithms are those to base 10.

$$\text{Thus if } x = 10^y$$

$$\text{then } y = \lg x$$

where \lg implies \log_{10} .

Natural logarithms are to base e .

$$\text{Thus if } x = e^y$$

$$\text{then } y = \ln x$$

where \ln implies \log_e . e is defined by

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = 2.718\dots$$

$$\begin{aligned} \ln(1+x) &\approx x - \frac{1}{2}x^2 \\ &\approx x \end{aligned}$$

for very small x .

► **Logarithmic and exponential relationships**

$$\log_b(x \times y) = \log_b x + \log_b y$$

$$\log_b x^n = n \log_b x$$

$$\log_b\left(\frac{x}{y}\right) = \log_b x - \log_b y$$

If $p = a^x = b^y$

$$x = \log_a p, \text{ and } y = \log_b p$$

Then $\log_a p = y \log_a b$
 $= \log_a b \times \log_b p$

This shows $\log_e x = (\log_e 10) \log_{10} x$

or $\ln x = (2.303) \lg x$

and $\lg x = (0.434) \ln x$

(C) Trigonometry

Angles

Refer to fig. 1.10.

Angle θ is defined by

$$\theta = s/r$$

$$\theta[\text{rad}] = \frac{s[\text{m}]}{r[\text{m}]} \quad (\text{p. 55})$$

(a) When $s = r$, $\theta = 1$ **radian** (rad).

(b) When $s = 2\pi r$ (one revolution)

$$\theta = \frac{2\pi r}{r} \text{ rad} = 2\pi \text{ rad}$$

$$= 360^\circ$$

(Note that the degree ($^\circ$) is *not* an SI unit.)

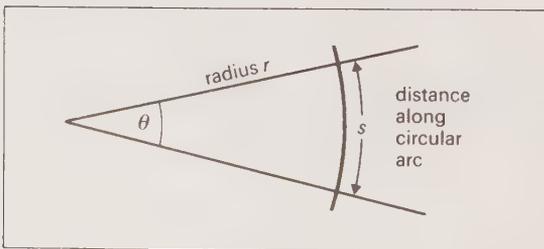


Fig. 1.10. Definition of angle θ .

$$\pi \text{ rad} = 180^\circ$$

Thus $1 \text{ rad} \approx 57.3^\circ$.

If $x = \sin \theta$,

then θ is the angle whose sine is x , and this would be abbreviated to

$$\theta = \arcsin x$$

Useful relationships

The quadrants in which particular trig. functions are positive are shown in fig. 1.11.

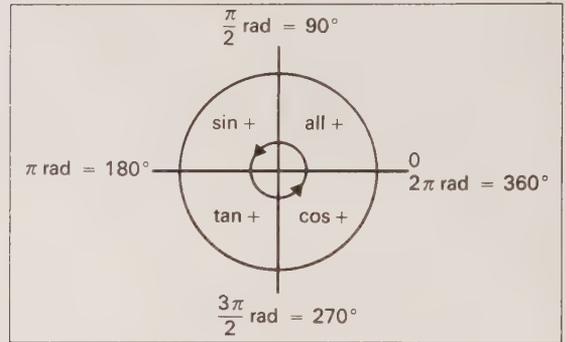


Fig. 1.11. Signs of trigonometric functions.

$$\tan \theta = \frac{\sin \theta}{\cos \theta}$$

$$\sin^2 \theta + \cos^2 \theta = 1 \quad (\text{p. 84})$$

$$\sin A + \sin B = 2 \sin\left(\frac{A+B}{2}\right) \cos\left(\frac{A-B}{2}\right) \quad (\text{p. 111})$$

Sine Law. $\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}$

Cosine Law. For any triangle
 $a^2 = b^2 + c^2 - 2bc \cos A$

It can be shown that

$$\sin \theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \dots$$

So when $\theta \ll 1$ (say $\theta \sim 0.1$ rad)

$$\sin \theta \approx \theta$$

Similarly $\tan \theta \approx \theta$ but

$$\cos \theta \approx 1$$

These approximations are used frequently in s.h.m. problems (p. 88) and in optics (p. 253).

(D) Calculus

Derivatives and integrals

$y = f(x)$	$\frac{dy}{dx} = f'(x)$	$\int y \, dx$
$y = x^n$	nx^{n-1}	$\frac{x^{n+1}}{n+1} + C$ (except when $n = -1$)
$y = \frac{1}{x}$	$-\frac{1}{x^2}$	$\ln x + C$
$y = \ln x$	$\frac{1}{x}$	$x \ln x - x + C$
$y = \sin x$	$\cos x$	$-\cos x + C$
$y = \cos x$	$-\sin x$	$\sin x + C$

C is the constant of integration, whose value will be determined by the limits of integration.

Logarithms and exponentials

If $\frac{dx}{dt} = cx$
 $\ln x = ct + C$
 giving $x = Ae^{ct}$

$$\frac{d}{dx}(\exp x) = \exp x$$

$$\int e^x \cdot dx = e^x + C$$

Average value of a function

The mean or average value $\langle y \rangle$ of $y = f(x)$ over the interval $x = a$ to $x = b$ is given by

$$\langle y \rangle = \frac{1}{b-a} \int_a^b y \, dx \quad (\text{p. 442})$$

Similarly

$$\langle y^2 \rangle = \frac{1}{b-a} \int_a^b y^2 \, dx$$

(pp. 198 and 442)

The root mean square (r.m.s.) value of y

$$y_{\text{r.m.s.}} = \sqrt{\langle y^2 \rangle}$$

For a periodic function, the interval $(b - a)$ is understood to be taken over an integral number of periods or half periods.

(E) Geometry

In fig. 1.12 the intersecting chords theorem gives

$$(2R - x)x = r^2$$

When x is small

$$x \approx r^2/2R \quad (\text{p. 301})$$

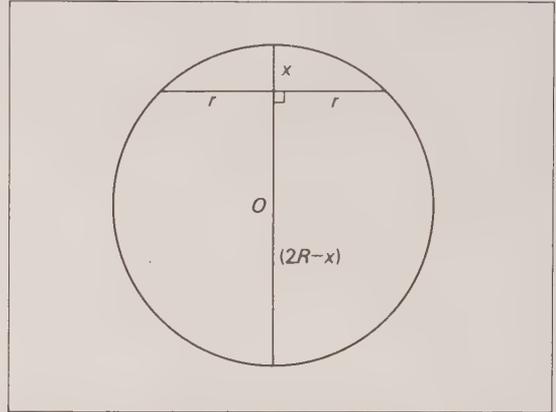


Fig. 1.12. The sagitta relationship.

Use of symbols

Δx represents some finite change in the value of x . The change might be positive or negative: thus $\Delta V = -3.0 \text{ m}^3$ would signify a volume reduction. (We do not write $-\Delta V$, and in so doing anticipate the negative sign.)

δx represents a variation in the value of x , a small increment whose size is later to be considered as vanishingly small. (Refer to p. 42 for an example.)

► 1.13 DIFFERENTIAL EQUATIONS IN PHYSICS

Differential equations arise in many branches of physics, as the following examples show.

(a) *Mechanics.* The solution of many problems in dynamics depends upon the equation

$$F = ma = m(d^2x/dt^2)$$

F may be a function of t , x and/or \dot{x} , and so we might have to integrate to find x as a function of t .

(b) *Oscillations and waves.* When F can be written in the form $-kx$, the equation becomes

$$m(d^2x/dt^2) + kx = 0$$

which is referred to as the differential equation of

s.h.m. Its solution is discussed below. Wave motions have a similar important differential equation: in one dimension

$$\frac{\partial^2 y}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 y}{\partial t^2}$$

(c) **Heat flow.** In one dimension the rate of heat flow through a material is described by the differential equation

$$dQ/dt = -\lambda A(d\theta/dx)$$

If, in a given situation, dQ/dt is constant, and both λ and A depend upon x , then we would have to integrate to find the variation of θ with x .

(d) **D.C. circuits.** The growth of current in a d.c. LR circuit is described by the equation

$$\mathcal{E} = L(dI/dt) + IR$$

The decay of charge on the plates of a capacitor in a d.c. CR circuit is described by the equation

$$IR + Q/C = 0$$

$$R(dQ/dt) + Q/C = 0$$

(e) **A.C. circuits.** The application of the law of energy conservation in an a.c. circuit leads to the differential equation

$$\mathcal{E}_0 \cos \omega t = L(dI/dt) + Q/C + IR$$

Since $I = dQ/dt$, we may write

$$\mathcal{E}_0 \cos \omega t = L \frac{d^2 Q}{dt^2} + R \frac{dQ}{dt} + \frac{1}{C} Q$$

a differential equation which can be solved to find Q (and I) in terms of t .

(f) **Radioactive decay.** The disintegration of a radioactive nucleus is an entirely *random* process. This means that $-dN/dt$, the rate of decay of a particular sample, is proportional to the (large) number N of active nuclei that it contains. We write

$$dN/dt = -\lambda N$$

If a radioisotope is being created in a nuclear reactor at the constant rate C , but at the same time the new nuclide disintegrates at the rate $-\lambda N$, then we can calculate the net rate of increase of nuclei from

$$dN/dt = C - \lambda N$$

The number of nuclei present after a given time interval must be found by solving this differential equation. (Note the formal similarity between

$$\begin{aligned} dN/dt + \lambda N - C &= 0, \text{ and} \\ dI/dt + (R/L)I - (\mathcal{E}/L) &= 0 \quad (\text{above}) \end{aligned}$$

Similarities of this kind enable us to solve many differential equations by comparing them with others whose solutions are already known.)

The solution of differential equations

There exists a considerable scheme for classifying differential equations as to their order and degree, and also a formal set of rules for their solution. Most of the differential equations which we meet can be solved by one of two methods.

(a) **By separation of variables.** If the variables can be separated onto the two sides of the equation, then the solution will be obtained by direct integration. Several examples are given below.

(b) **By inspection.** Because the equations that we deal with tend to treat a small number of distinct physical situations, previous experience and physical intuition will often enable us to *guess the form of the solution*. We can then test our proposed solution by substituting it into the original differential equation.

Example: the LC circuit. Experience teaches us that an equation of the form

$$\frac{d^2 Q}{dt^2} = -\left(\frac{1}{LC}\right) Q$$

describes an oscillatory situation, since it closely resembles

$$\frac{d^2 x}{dt^2} = -\left(\frac{k}{m}\right) x$$

Suppose we write $(1/LC) = \omega^2$, and then suggest

$$Q = A \sin \omega t + B \cos \omega t$$

as our solution of

$$d^2 Q/dt^2 = -\omega^2 Q$$

A and B are arbitrary constants. Differentiating twice, we have

$$(1) \quad dQ/dt = A\omega \cos \omega t - B\omega \sin \omega t$$

$$(2) \quad \begin{aligned} d^2 Q/dt^2 &= -A\omega^2 \sin \omega t - B\omega^2 \cos \omega t \\ &= -\omega^2 Q \end{aligned}$$

Since this is the original equation, we have verified that

$$Q = A \sin \omega t + B \cos \omega t$$

is a solution.

Constants

The above solution can also be expressed in the form

$$Q = C \sin(\omega t + \delta)$$

where C and δ are two further arbitrary constants related to A and B . To eliminate *two* such constants, and hence arrive back at the original equation, it was necessary to differentiate *twice*. This illustrates a general rule:

The general solution of an n th order equation (involving $d^n y/dx^n$) will have n constants.

A **particular solution** is one in which the arbitrary constants have been given specified values. Thus to obtain the *particular solution* of a differential equation describing s.h.m. we need to have *two* separate items of information, or **boundary conditions**. For example we might be told that $\dot{x} = 0$ and $x = +a$ when $t = 0$, and this would enable us to write a particular solution involving no arbitrary constants.

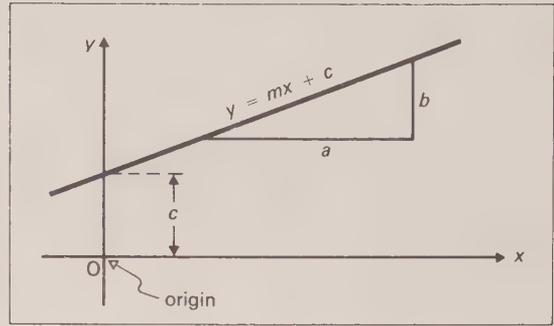


Fig. 1.13. A typical straight line.

1.14 GRAPHS

The usefulness of graphs

(a) They demonstrate pictorially how a pair of varying quantities are related. They can indicate, for example, the point at which a relationship is no longer linear. Graphs used like this are visual aids.

(b) If we know how two quantities are related, we can determine graphically the constants of the equation which relates them. (We use the graph as a way of judging the best line through the points.)

(c) A graph is frequently an effective substitute for a table of values when an instrument is calibrated. For a thermocouple thermometer one could plot a graph of galvanometer deflection in divisions against temperature in kelvins.

Straight and curved graphs

(a) For pictorial purposes a curved graph may be acceptable, or even preferable to a straight line. The equation

$$y = ax^2 + bx + c$$

represents a **parabola**, of which examples are given on p. 30 and p. 341.

(b) **Equation of a straight line** (fig. 1.13). In the equation $y = mx + c$, $m = b/a$ is the **slope** or **gradient** of the line.

When $x = 0$, $y = c$, which is the **intercept** on the y -axis. Similarly when $y = 0$, $x = -c/m$.

x is the **independent** variable, and y the **dependent** variable.

Choosing what to plot

If a graph is to be used for calculation, one nearly always chooses to plot a *straight line* from which the slope m and/or the intercept c can be measured.

Examples of straight line plots

Equation	Plot along		m	c
	y -axis	x -axis		
$T^2h = Bh^2 + A$	T^2h	h^2	B	A
$T = AI^{1/2}$	T	$I^{1/2}$	A	zero
$1/n = Bx + A$	$1/n$	x	B	A

Log graphs

Suppose we assume that two quantities y and x are related by

$$y = Ax^B$$

where A and B are constants.

$$\begin{aligned} \text{Then } \lg y &= \lg A + \lg(x^B) \\ &= \lg A + B \lg x \end{aligned}$$

If we plot $\lg y$ along the vertical axis, and $\lg x$ along the horizontal axis, then

(1) a *straight line graph* would confirm that the original assumption was correct, and

(2) its slope would be B , and intercept $\lg A$.

Hence A and B could be evaluated.

How to plot the graph

(a) (i) Plot the quantity that is controlled during the experiment (the *cause*) along the x -axis (as **abscissa**).

(ii) Plot the quantity which changes as a result (the *effect*) along the y -axis (as **ordinate**).

(b) Essentially a graph relates *numbers* (rather than physical quantities). The axes should be labelled in the same way as the headings for a column in a table.

(c) The scale chosen is inevitably a compromise. Consider the following.

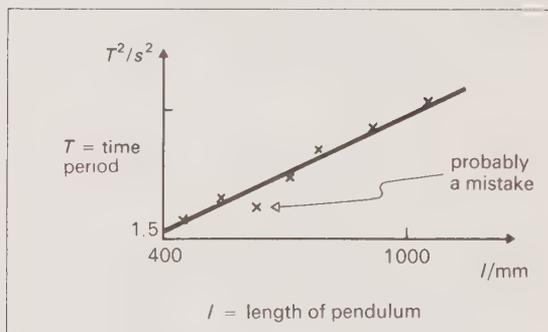


Fig. 1.14. A graph drawn to reach a conclusion from a single pendulum experiment.

(i) The scale should be simple to prevent mistakes.

(ii) The paper should be filled, and experimental points distributed as widely as possible.

(iii) If an intercept is required, the origin may have to be included.

(iv) If the slope of a straight line is to be measured, it should be inclined to the x -axis at an angle between 30° and 60° .

(d) Points can be marked \times , \odot or \ominus if it is required to show the range of likely error. (An isolated dot is easily lost.)

(e) A smooth curve or straight line should be drawn, even if this does emphasize a random error.

(f) Straight line graphs for which $c = 0$ is expected should *not* be drawn through the origin automatically. Non-coincidence may reveal a systematic error (see p. 302 for an example).

Fig. 1.14 illustrates some of these points.

The graph should be plotted before apparatus is dismantled. This has two advantages.

(i) Mistakes will become obvious, and can be checked.

(ii) Further measurements can be taken for crucial regions of the graph. These can then be re-plotted on a larger scale.

II PRINCIPLES OF MECHANICS

2	KINEMATICS	28
3	NEWTON'S LAWS	34
4	WORK AND ENERGY	42
5	MOMENTUM	48
6	ROTATIONAL DYNAMICS	55
7	EQUILIBRIUM	65
8	FLUIDS AT REST	71
9	THE BERNOULLI EQUATION	74
	SUMMARY OF MECHANICS	79

2 KINEMATICS

2.1 DEFINITIONS	28	2.4 PARABOLIC MOTION	30
2.2 LINEAR MOTION	29	2.5 ANGULAR MOTION	30
2.3 RELATIVE VELOCITY	29	2.6 UNIFORM MOTION IN A CIRCLE	32

2.1 DEFINITIONS

Kinematics is the study of the motion of points (or massless particles): we do not consider the agent that causes the motion. The following types of motion occur frequently and can be described by simple mathematics:

- (a) linear motion
- (b) circular motion
- (c) parabolic motion
- (d) simple harmonic motion.

The **displacement** s of a particle is the length and direction of the line drawn to the particle from some fixed point (frequently called the **origin**). Displacement is a *vector* quantity.

Velocity

Suppose that during some motion, the displacement s of a particle changes by Δs during a time interval Δt . The **average velocity** v_{av} during the interval is defined by the equation

$$v_{av} = \frac{\Delta s}{\Delta t}$$

$$v_{av} \left[\frac{\text{m}}{\text{s}} \right] = \frac{\Delta s[\text{m}]}{\Delta t[\text{s}]}$$

Average velocity is a *vector* quantity. The vector Δs is found by subtracting the initial displacement from the final displacement. Suppose that Δs and Δt become vanishingly small, and are written δs and δt .

The **instantaneous velocity** v (at the instant contained by δt) is defined by the equation

$$v = \lim_{\delta t \rightarrow 0} \left(\frac{\delta s}{\delta t} \right) = \frac{ds}{dt}$$

Instantaneous velocity is a *vector* quantity which may be measured in m s^{-1} .

Speed

The **average speed** of a particle is defined by

$$\frac{\text{distance travelled along actual path}}{\text{time taken } \Delta t}$$

The **instantaneous speed** is the value of this expression when the time interval Δt becomes vanishingly small. From these definitions it follows that

(a) the instantaneous speed equals the magnitude of the instantaneous velocity (speed being a scalar quantity), but

(b) the average speed is *not* necessarily equal to the magnitude of the average velocity (although it can be).

Acceleration

Suppose that the velocity changes by Δv during some time interval Δt . Then we define the **average acceleration** a_{av} by

$$a_{av} = \frac{\Delta v}{\Delta t}$$

and the **instantaneous acceleration** a by

$$a = \lim_{\delta t \rightarrow 0} \frac{\delta v}{\delta t} = \frac{dv}{dt}$$

$$a \left[\frac{\text{m}}{\text{s}^2} \right] = \frac{\delta v [\text{m/s}]}{\delta t [\text{s}]}$$

Acceleration is a *vector* quantity, since Δv is calculated by vector subtraction. A body moving with uniform velocity has zero acceleration: this means that neither its speed nor its direction of motion is changing with time.

2.2 LINEAR MOTION

(A) With constant acceleration

Suppose that $v = u$

and $s = 0$

when $t = 0$.

Then after time t

$$a = \frac{\Delta v}{\Delta t} = \frac{v - u}{t}$$

so $v = u + at$ (1)

Because

displacement = (average velocity)(time)

we write $s = \frac{u + v}{2} \cdot t$ (2)

(1) and (2) may be combined so as to eliminate v , u or t . The results are listed for reference.

$$\begin{aligned} v &= u + at \\ s &= \frac{u + v}{2} \cdot t \\ s &= ut + \frac{1}{2}at^2 \\ s &= vt - \frac{1}{2}at^2 \\ v^2 &= u^2 + 2as \end{aligned}$$

*Uniformly
accelerated
linear
motion*

Comments

(a) If we know any *three* of v , u , a , s and t , then the remainder can be found.

(b) s represents *displacement* (not distance travelled by particle).

(c) These results can be found quickly by integration.

Integration of

$$\frac{dv}{dt} = a$$

gives $v = u + at$

(since a is to be constant).

$$a = \frac{dv}{dt} = \frac{ds}{dt} \cdot \frac{dv}{ds} = v \cdot \frac{dv}{ds}$$

which gives $\frac{1}{2}v^2 = \frac{1}{2}u^2 + as$

The constants of integration have been inserted.

(B) With variable acceleration

If a depends on t in some simple way, then calculus can be used. If a varies in an irregular manner, the problem is best solved by plotting a graph.

The displacement–time graph

Since $v = ds/dt$, the *slope* of the graph represents the *instantaneous velocity*.

The area under the graph has no particular significance.

The velocity–time graph

(a) Since $a = dv/dt$, the *slope* of the graph represents the *instantaneous acceleration*.

(b) Since $v = \frac{ds}{dt}$

$$s = \int v \cdot dt$$

the *area* under the graph represents the *displacement*.

Areas above the time-axis are counted as positive, to ensure that a velocity in a given direction corresponds to a displacement in the same direction. The area under a graph can be determined by counting squares on graph paper.

A point moving with simple harmonic motion has a variable acceleration: graphs which describe the motion are plotted on p. 85.

2.3 RELATIVE VELOCITY

The velocity of a point A relative to a point B is the velocity which A *appears* to have to an observer who is moving with B. The velocity of B relative to A is a vector of equal magnitude in the opposite direction.

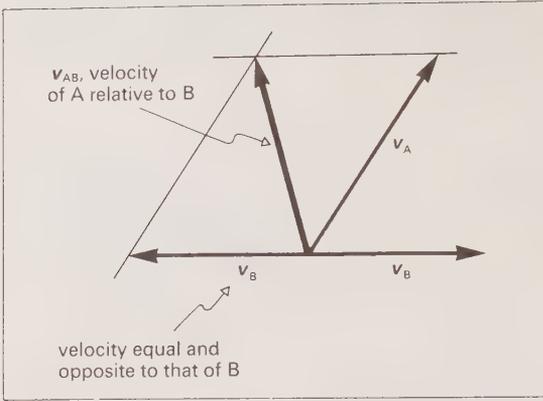


Fig. 2.1. Calculation of relative velocity.

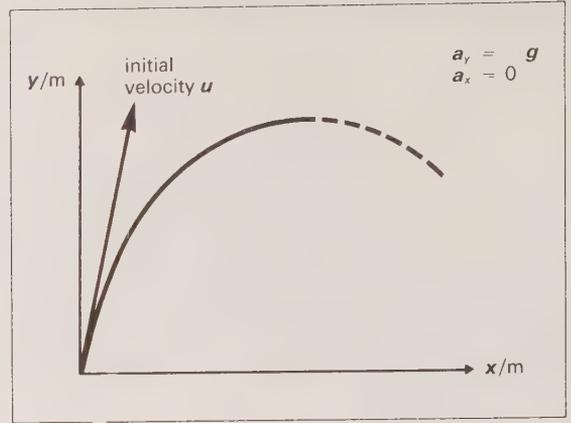


Fig. 2.2. Parabolic motion of a point in the x - y plane.

To calculate the velocity of A relative to B we adopt this procedure (fig. 2.1).

A velocity ($-v_B$) equal and opposite to that of B has been superimposed on both A and B, so that B has been brought to rest. v_{AB} is A's new velocity, which is found by vector addition.

$$\begin{aligned} v_{AB} &= v_A + (-v_B) \\ &= v_A - v_B \end{aligned}$$

Similarly $v_{BA} = v_B - v_A = -v_{AB}$

Thus to find the velocity of A relative to B, we subtract (vectorially) the velocity of B from that of A.

2.4 PARABOLIC MOTION

Suppose a point is made to move in two dimensions as shown in fig. 2.2.

Suppose that for this motion

- acceleration in the y -direction $a_y = -g$
- x -direction $a_x = 0$
- initial velocity in the y -direction $= u_y$
- x -direction $= u_x$

It is convenient to treat the x - and y -motions independently. Thus using

- (a) $v^2 = u^2 + 2as$, we have
- $v_y^2 = u_y^2 - 2gy$ and
- $v_x^2 = u_x^2$.

- (b) $s = ut + \frac{1}{2}at^2$, we have
- $y = u_y t - \frac{1}{2}gt^2$ and
- $x = u_x t$.

We can put $t = x/u_x$ to eliminate t from the last two equations, giving

$$y = u_y \frac{x}{u_x} - \frac{1}{2}g \left(\frac{x}{u_x} \right)^2$$

This relation between x and y is called the **trajectory equation**, and shows that the point describes a **parabolic path**. The motion is described by the graphs of fig. 2.3 (opposite).

The graphs should be read together with the comments on p. 29.

Examples of parabolic motion

A point describes a parabolic trajectory when it experiences a uniform acceleration in one direction while having a component of velocity in some other direction. This may occur

- (a) in motion under gravity, as implied in this section,
- (b) when an electrically charged particle is subject to a *uniform electric field* (p. 341).

2.5 ANGULAR MOTION

Suppose that a line is changing its orientation with time, and that it rotates through an angle $\Delta\theta$ in a time interval Δt .

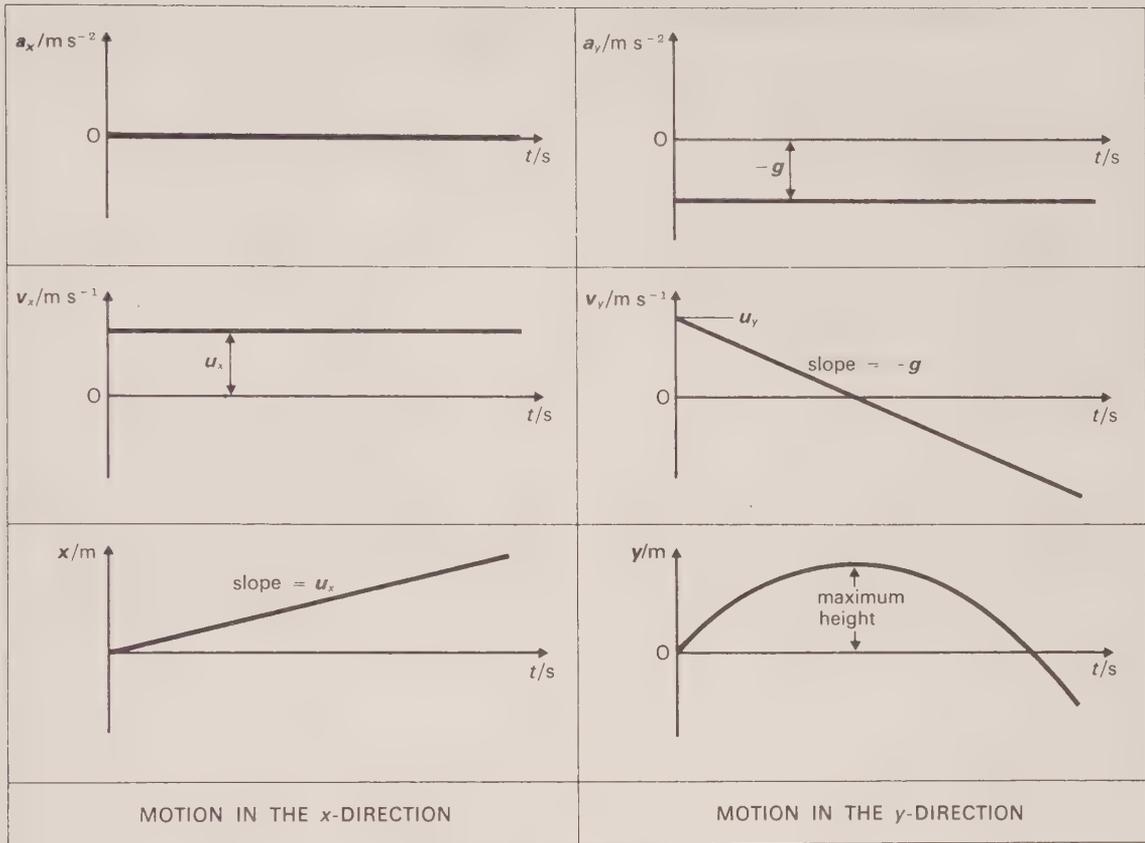


Fig. 2.3. $a-t$, $v-t$ and $s-t$ graphs for a parabolic motion.

The **average angular velocity** ω_{av} during the interval is defined by the equation

$$\omega_{av} = \frac{\Delta\theta}{\Delta t}$$

$$\omega_{av} \left[\frac{\text{rad}}{\text{s}} \right] = \frac{\Delta\theta[\text{rad}]}{\Delta t[\text{s}]}$$

The **instantaneous angular velocity** ω is defined by the equation

$$\omega = \lim_{\delta t \rightarrow 0} \left(\frac{\delta\theta}{\delta t} \right) = \frac{d\theta}{dt}$$

Instantaneous angular velocity is a *pseudovector** which is usually measured in rad s^{-1} , and whose dimensions are $[\text{T}^{-1}]$.

Frequently we are concerned with points moving at a fixed speed in a circular path (fig. 2.4). When we refer to the angular velocity of P, we mean the angular velocity of the radius vector OP drawn to P. If the linear speed v is fixed, then so is the angular velocity ω ; we can write

$$\omega = \frac{d\theta}{dt} = \frac{\Delta\theta}{\Delta t}$$

The **period** T for one revolution is the time taken to describe an angle 2π radians, thus using

* In fig. 2.4 the angular velocity of P is (by convention) directed into the paper, so as to obey a right-hand corkscrew rule.

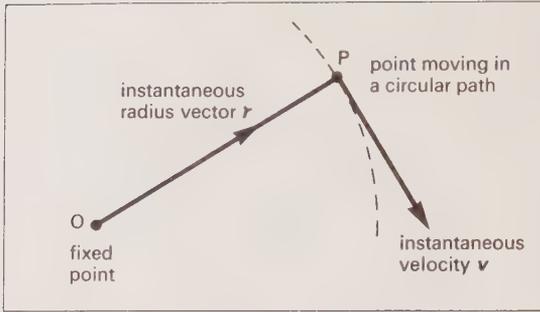


Fig. 2.4. The angular velocity of a point.

$$\Delta t = \frac{\Delta\theta}{\omega}$$

We have

$$T = \frac{2\pi}{\omega}$$

Since the point P covers a distance $2\pi r$ in time T at speed v

$$T = \frac{2\pi r}{v}$$

from which

$$v = \omega r$$

$$v \left[\frac{\text{m}}{\text{s}} \right] = \omega \left[\frac{1}{\text{s}} \right] r [\text{m}]$$

2.6 UNIFORM MOTION IN A CIRCLE

Fig. 2.5 shows a point moving with instantaneous velocity v and constant speed v in a circle of radius r . The angular speed $\omega = v/r$. Since the direction of motion changes with time, the point is accelerated. We will calculate its acceleration.

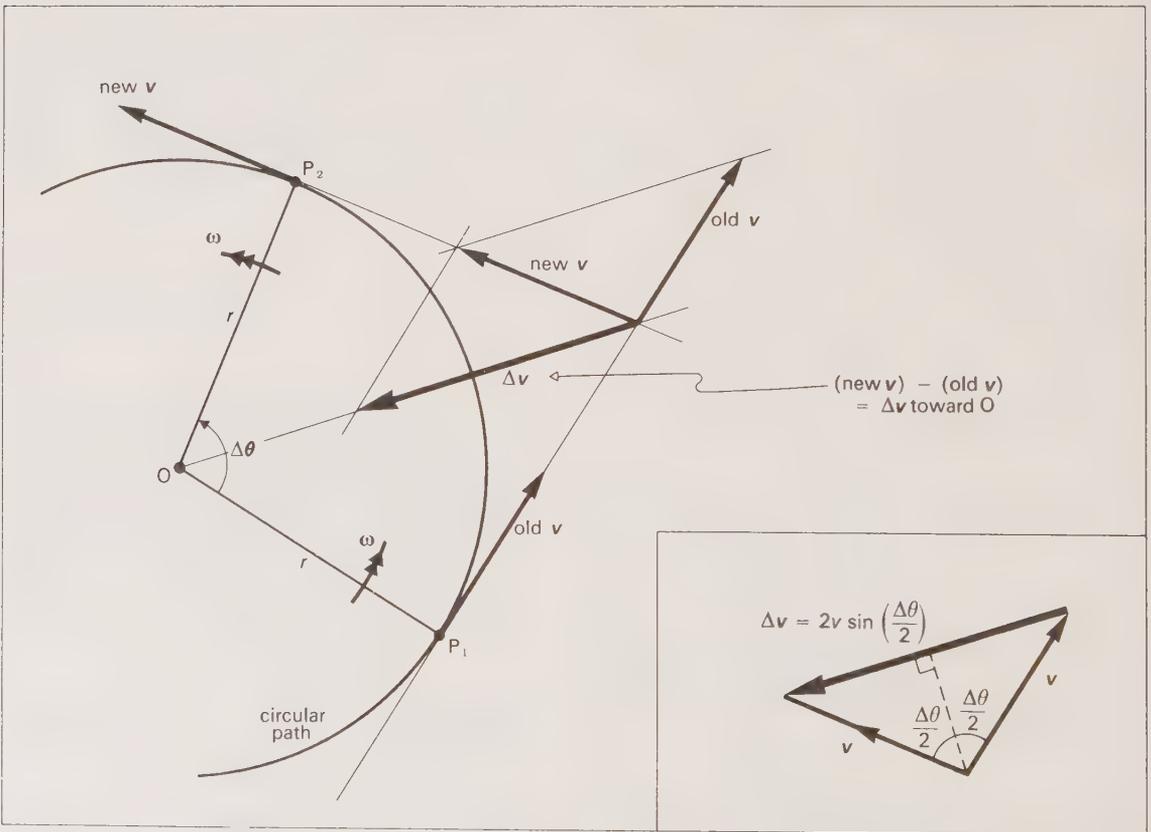


Fig. 2.5. Calculation of the change of velocity Δv over the time interval Δt .

Suppose the successive positions P_1 and P_2 are separated by an angular displacement $\Delta\theta$. This corresponds to a time interval

$$\Delta t = \frac{\text{length of path}}{\text{speed}} = \frac{r \Delta\theta}{v}$$

(A) Average acceleration a_{av}

Since (from the diagram)

$$\Delta v = 2v \sin\left(\frac{\Delta\theta}{2}\right)$$

towards the centre of the circle,

$$\begin{aligned} a_{av} &= \frac{\Delta v}{\Delta t} = \frac{2v \sin(\Delta\theta/2)}{r \Delta\theta/v} \\ &= \frac{v^2}{r} \cdot \frac{\sin(\Delta\theta/2)}{(\Delta\theta/2)} \end{aligned}$$

and is also directed towards the centre of the circle, i.e. is **centripetal**.

(B) Instantaneous acceleration a

$$\begin{aligned} a &= \lim_{\Delta t \rightarrow 0} \left(\frac{\Delta v}{\Delta t} \right) \\ &= \frac{v^2}{r} \cdot \lim_{\Delta\theta \rightarrow 0} \frac{\sin(\Delta\theta/2)}{(\Delta\theta/2)} \\ &= \frac{v^2}{r} \end{aligned}$$

towards the centre of the circle.

a is the *limiting value* of the average acceleration.

$$a = \frac{v^2}{r} = \omega^2 r = v\omega$$

directed towards the
centre of the circle

This is not uniformly accelerated motion, because the value of a is fixed only in magnitude, and not in direction. It is shown by a particle moving under a centrally directed force of constant size (see, for example, p. 39).

3 NEWTON'S LAWS

3.1 INTRODUCTION	34
3.2 NEWTON'S LAWS OF MOTION	34
3.3 FORCES	35
3.4 DISCUSSION OF NEWTON'S SECOND LAW	36

3.5 MASS AND WEIGHT	37
3.6 EXAMPLES OF THE USE OF $F = ma$	38
3.7 SKIDDING AND TOPPLING	40
▶ 3.8 THE TENSION IN A ROTATING HOOP	41

3.1 INTRODUCTION

Dynamics is the branch of mechanics which relates the properties of a body (such as its *mass*) and the effects of the body's environment (through an applied *force*) to describe the changes of motion of the body.

Classical mechanics confines itself to situations where the speeds of bodies relative to one another are small compared to c , the speed of electromagnetic radiation. Under these conditions (which are usually assumed in this book) **Newtonian mechanics**, as it is called, is known to be a very good approximation to **relativistic mechanics**. (See p. 2.)

A **particle** is an object without extent. In practice large bodies are treated as particles when they are small relative to their environment: thus when we analyse the dynamics of a car rounding a corner, we frequently ignore the fact that the mass of the car is distributed over a finite volume.

Momentum p

Suppose a body of mass m has an instantaneous velocity v .

The **momentum** p of the body is defined by the equation

$$p = mv$$

$$p \left[\frac{\text{kg m}}{\text{s}} \right] = m[\text{kg}]v \left[\frac{\text{m}}{\text{s}} \right]$$

Notes

(a) 1 N s is a unit equivalent to 1 kg m s^{-1} , since $1 \text{ N} = 1 \text{ kg m s}^{-2}$ (p. 36).

(b) Because mass is a scalar and velocity is a vector, momentum is a *vector* along the direction of the velocity.

(c) Changes of momentum are calculated by *vector subtraction*, thus:

$$\Delta p = \left(\begin{array}{c} \text{final} \\ \text{momentum} \end{array} \right) - \left(\begin{array}{c} \text{initial} \\ \text{momentum} \end{array} \right)$$

The concept of momentum is used in stating *Newton's Second Law*.

3.2 NEWTON'S LAWS OF MOTION

Law I Unless a resultant force acts on a body, its velocity will not change.

Law II The rate of change of momentum of a body is proportional to the resultant force that acts on it.

Law III If body A exerts a force F on body B, then body B exerts a force F on body A of the same size and along the same line, but in the opposite direction.

In symbols

Law I	If $F = 0$, $\Delta v = 0$
Law II	$F \propto \frac{d}{dt}(mv)$
Law III	$F_{AB} = -F_{BA}$

Comments on Newton's laws

(a) Law I gives us an intuitive meaning of force.

A **resultant force** is that agent which changes the velocity (and momentum) of a body.

In real situations, the resultant force is often *frictional* so that experimental verification of Law I requires effectively *frictionless* conditions. Some laboratory demonstrations include the stroboscopically photographed motions of

- (i) a CO₂ puck on a smooth, level glass plate,
- (ii) a small vehicle on a smooth, level air track, and
- (iii) a bullet travelling through air.

(b) Laws I and II together enable us to define the meanings of, and measures for, both mass and force (p. 36).

(c) Law I is a special case of Law II.

(d) Law III refers to a pair of forces which must always act on two *different* bodies. These two forces have the same size at every instant of time.

(e) A pair of forces acting on the same body *cannot* be the 'action–reaction' pair referred to in Law III. Possible difficulty is avoided by referring to any force in full detail, such as 'the upward push exerted *by* the table *on* the book'.

(f) We cannot *prove* Newton's Laws, because they are assertions. Newton himself used the words '*Axiomata Sive Leges Motus*'—'The axioms or laws of motion'. Nevertheless we are justified in using them because (with the qualifications mentioned on p. 2) everything which they predict agrees with experimental observation.

3.3 FORCES

Force is a *vector* quantity: this is an experimental result obtained by the following procedure.

(a) Apply several different forces in turn to a given body, and note the acceleration produced by each.

(b) Apply the same forces simultaneously. Their resultant produces the same acceleration as the vector sum of the independent accelerations.

This leads to the **superposition principle** for forces:

When several forces act on a body simultaneously, their single equivalent resultant can be found by vector addition.

The origin of forces

The *interactions* of nature are discussed on p. 3. A force experienced by one body shows us that it is interacting with another. In mechanics we are concerned with two types of force.

(a) **Gravitational forces** are so weak that we usually ignore them unless a massive body (such

as the Earth) is part of the environment. The gravitational forces we discuss are always attractive.

(b) **Electromagnetic forces** are responsible for most of the effects that we observe (e.g. the push of a table on a book). They may be attractive or repulsive.

We say that bodies are *separated* when they do not exert appreciable contact forces on each other: these forces are electromagnetic in origin. When the separation becomes large, the interaction between two bodies takes place through a *long-ranged force*, such as a gravitational or electrostatic force obeying an inverse square law. (The *range* of a force is discussed fully on p. 128.)

Description of forces

A force is described by specifying the following:

- (i) its magnitude and direction.
- (ii) the body on which it acts, and which part of the body,
- (iii) the body that exerts the force, and
- (iv) the nature of the force.

The gravitational forces that we discuss are always pulls, and are usually referred to as *weights*. The electromagnetic forces can be pushes or pulls, and are referred to variously as normal *contact forces*, *frictional forces*, fluid *upthrusts*, *lift* and *drag forces*, *electrostatic* and *magnetic forces*.

(a) **Weights** are discussed on p. 37.

(b) **Normal contact forces** are brought into play when the molecules of one body are so close to those of another that their respective electrons and atomic nuclei repel each other. The bodies are then said to **touch**. When the force is always normal to the bodies' common surface, then one of the surfaces is said to be **smooth**. If the bodies are **rough**, the resultant force is the vector sum of the normal force and the frictional force.

(c) **Tension** is the condition of a body subjected to equal but opposite forces which attempt to increase its linear dimensions along their line of action. Tension forces are then the *pulls* exerted by stretched strings or rods on the bodies to which they are attached, and act along the direction of the string or rod. If frictional forces are negligible the magnitude of a tension force is not altered by a change of direction of a string.

Compression is the opposite condition: a rod, but not a string, can exert a compressive force, which we think of as a *push*.

(d) *Frictional forces* act along the common surface between two bodies in such a direction as to oppose the *relative* movement of the two bodies.

Thus a car is accelerated by the frictional push of the road on the tyres, a force which prevents the tyres from sliding on the road. The *moving* car experiences a drag force from the air through which it moves.

As with all forces, frictional forces occur in pairs, and one must be careful to specify which force one is considering.

The coefficient of friction

Fig. 3.1 is a free-body diagram (p. 39) for a book resting on a table.

It is found that

- (1) F and T are in opposite directions,
- (2) $F \leq T$,
- (3) $F \leq F_{\max}$, the largest value that F can take,
- (4) $F_{\max} = \mu R$, where μ is a number (without unit) called the **coefficient of friction** for that pair of surfaces.

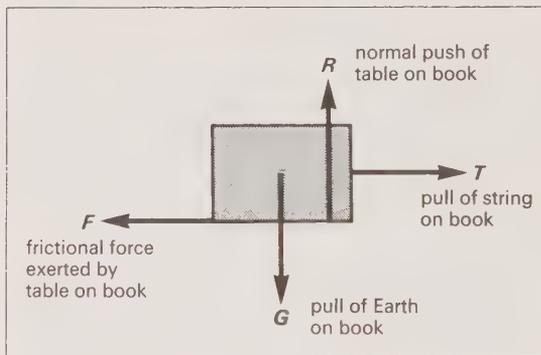


Fig. 3.1. Frictional forces.

Comments

(a) F_{\max} is controlled by the material and cleanliness of the surfaces in contact, but not by the common surface area.

(b) The value of F is only as large as is necessary to maintain equilibrium (when this is possible), and will frequently be less than μR .

(c) When sliding occurs the frictional force may become smaller, but we usually approximate by setting it equal to μR .

(d) If either surface is *smooth*, then $\mu = 0$ and $F = 0$.

3.4 DISCUSSION OF NEWTON'S SECOND LAW

On p. 34 Newton's Second Law was stated in the form

$$F \propto \frac{d}{dt}(mv)$$

or

$$F = k \frac{d}{dt}(mv)$$

where k is a non-dimensional constant. In classical mechanics, if we have a *body* (of constant mass)

$$\begin{aligned} F &= km \frac{dv}{dt} + kv \frac{dm}{dt} \\ &= km \frac{dv}{dt} \quad \left(\text{since } \frac{dm}{dt} = 0 \right) \\ &= kma \end{aligned}$$

We have units for m [kg] and a [m s^{-2}] but not for F . We then *choose* $k = 1$, and in so doing we also *define our unit for force*.

$$F[\text{N}] = m[\text{kg}] a \left[\frac{\text{m}}{\text{s}^2} \right]$$

1 newton (N) is that force which accelerates a mass of 1 kg at 1 m s^{-2} .

$$1 \text{ N} = 1 \text{ kg m s}^{-2}$$

The newton is an **absolute** force unit.

$$F = ma$$

One form of Newton's Second Law

Notes

(a) F is the *resultant* force experienced by the body, i.e. the vector sum of all the forces acting on the body.

(b) Writing $a = F/m$, in which a and F are vectors, we see

- (i) $a \propto F$ for a given m ,
- (ii) a occurs along the same direction as F , and
- (iii) $a \propto 1/m$ for a given F .

(c) $F = ma$ defines the procedure for the *measurement of force*: we measure the acceleration it produces in a known mass (such as the standard kilogram).

(d) $F = ma$ can be verified experimentally by studying the motion of trolleys on compensated runways or pucks on smooth glass plates. Values of F can be deduced from the extensions of stretched springs, and values of a can be calculated from ticker tape or from stroboscopic photographs.

Name	Physical quantity	Unit for measurement	Is a measure of	Measured by	Variation
mass	mass (scalar)	kilogram	inertia	comparison with standard mass	does not vary
weight	force (vector)	newton	gravitational attraction of the Earth	calibrated spring balance	variable

The principle of the measurement of mass

The **inertia** of a body is its ability to resist a change of motion.

The **mass** of a body is the quantitative measure of that inertia (for linear motion).

Suppose the same force F acts on two masses, m_0 , the standard kilogram, and m , a mass to be measured. (F can be produced by extending a given spring by a fixed amount.)

We define the ratio of the two masses by

$$\frac{m_0}{m} = \frac{a}{a_0}$$

where a and a_0 are the accelerations produced in the unknown and standard masses respectively.

Since a and a_0 can be measured directly, and m_0 is known (by definition), the value of m can be found.

Notes

(a) If a different common force is used, and gives accelerations a' and a'_0 , then *experiment* shows

$$\frac{m_0}{m} = \frac{a}{a_0} = \frac{a'}{a'_0}$$

The ratio of the masses does not depend on the common force.

(b) The procedure can be used to compare any (non-standard) masses:

$$\frac{m_1}{m_2} = \frac{a_2}{a_1}$$

(c) The experiment can be used to show that **mass** is a *scalar* quantity. Two bodies of masses m_1 and m_2 , when fastened together behave as would a mass ($m_1 + m_2$).

(d) In this section we show how masses can be

compared *in principle*; *in practice* (p. 12) different methods are simpler.

3.5 MASS AND WEIGHT

In this book the **weight** of a body is defined to be the gravitational attraction of a massive body (such as the Earth) for that body. *Weight* is thus the name given to a particular type of force, and must be distinguished from *mass* (see table above). (Note that the term *weight* is not always defined in this way: some authors define it to be the downward force exerted by the body on its support.)

The pull of the Earth on a body is controlled by the *gravitational mass* (p. 169): the resistance to a change of motion depends on the *inertial mass*. Because the gravitational and inertial masses are proportional (p. 88), inertial mass is closely related to weight, *but is a different physical quantity*.

Variation of weight and constancy of mass

(a) To obtain an intuitive feeling for the *mass* of a body, we frequently apply by hand a force to it, and sense the resistance it offers when undergoing a particular acceleration. This force *can* be applied *horizontally*.

(b) To measure the *weight* of a body, we frequently apply by hand a force to it which maintains it in approximate equilibrium. This force *must* be applied *vertically*.

Suppose we do this for a body of mass 6 kg, both at the surface of the Earth and at the surface of the Moon.

(a) To accelerate the body at 10 m s^{-2} horizontally, we apply a force

$$\begin{aligned} F &= ma \\ &= (6 \text{ kg}) \times (10 \text{ m s}^{-2}) \\ &= 60 \text{ N} \end{aligned}$$

both on the Earth, and on the Moon.

The inertia (mass) is the same in both cases.

(b) To hold the body in approximate equilibrium, we apply an upward vertical force.

(i) On Earth $F = 60 \text{ N}$.

(ii) On the Moon we would find by experiment that $F = 10 \text{ N}$.

The weight on the Moon is 1/6 its value on Earth.

This difference between weight and mass would be emphasized if one were to kick a massive body at a place where the gravitational field was vanishingly small.

Measurement of weight

Suppose that g is the acceleration, relative to the surface of the Earth, of a body in a state of free-fall. g can be determined directly by timing accurately the fall from rest of an object, such as a ball bearing, through a carefully measured vertical distance h , and then using $h = \frac{1}{2}gt^2$. Then if we ignore the effect of the Earth's rotation about its axis, we can in principle use

$$F = ma$$

in the form

$$G = mg$$

Measurement of m and g would enable us to calculate G . In practice G can be measured directly by using a calibrated spring balance. Nevertheless the interpretation of the balance reading is complicated, because

(a) g varies from point to point on the Earth, and

(b) the surface of the Earth is an accelerated frame of reference.

Fig. 3.2 shows that observed values for g do not give an exact measure for the pull of the Earth on a body: mg and the true value for G differ in both magnitude and direction.

Gravitational force units

These are force units related to the *apparent* pull of the Earth on a standard mass.

1 kilogram weight (kgwt) is the apparent pull of the Earth on a mass of 1 kg.

1 kilogram force (kgf) is the apparent pull of the Earth on a mass of 1 kg at a place where $g_0 = 9.80665 \text{ m s}^{-2}$ (standard gravitational acceleration).

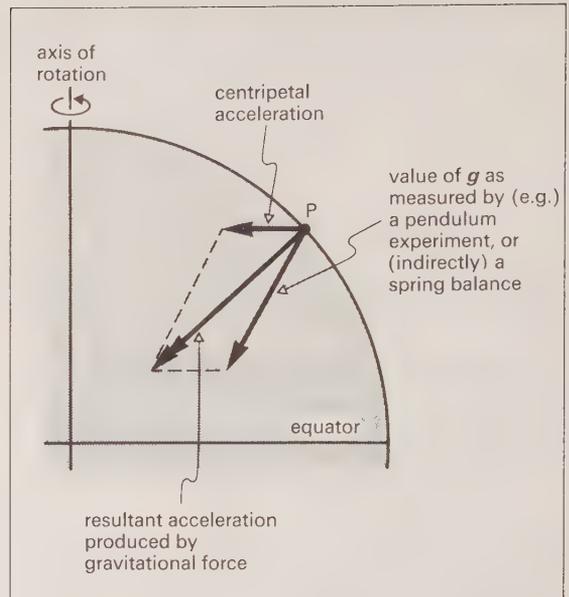


Fig. 3.2. The effect of the Earth's rotation on the measured value of g (greatly exaggerated).

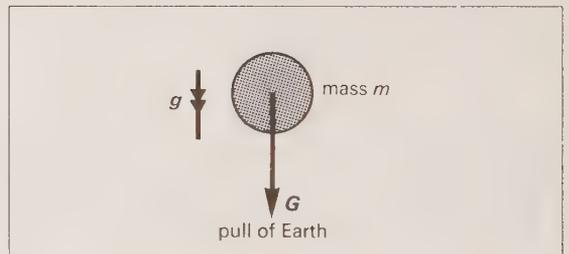


Fig. 3.3. A body in a state of free-fall in a vacuum.

It is seen that the kilogram weight has a variable magnitude, so a physicist chooses to use a newton, which has a constant magnitude.

Refer to fig. 3.3, in which G can be expressed in two ways:

(a) using $F = ma$

$$G[\text{N}] = m[\text{kg}] g \left[\frac{\text{N}}{\text{kg}} \right], \text{ or}$$

(b) G is numerically equal to m , if G is expressed in kgwt, and m in kg.

Thus we can relate gravitational and absolute force units:

1 kgwt is about 9.8 N, whereas
1 kgf = 9.80665 N.

It should be noted that *neither* of these units is an SI unit, and their use should be avoided.

3.6 EXAMPLES OF THE USE OF $F = ma$

In $F = ma$

F is the sum of the resolved parts of the forces in a particular direction, and a is the acceleration of the body in that direction.

To apply the law, adopt the following procedure.

(1) Draw the diagram that represents the general situation, marking all bodies in the environment.

(2) Select *one body* from the situation whose motion is to be analysed, and draw a **free-body diagram** for that body. For this the body is removed from its environment, together with *all the forces exerted on it* by the bodies with which it interacts. These forces should be marked in absolute units.

(3) Select and orientate convenient coordinate axes, marking the acceleration a in the direction to be considered.

(4) Calculate the value of F defined above, and apply the law.

One common mistake is to mark the forces exerted by the chosen body: these are not relevant.

(Learning these rules is no guarantee of facility in applying them: practice is essential.)

(A) Linear motion

For the situations (b) and (c) of fig. 3.4, *Newton's Second Law* is written

$$(R - w) = ma \quad \text{for load}$$

$$(T - R - W) = Ma \quad \text{for lift}$$

where m, M are masses, and w, W weights.

(B) Uniform circular motion

Every body moving in a circular path is accelerated (p. 32). For the special case of a body moving at speed v in a circle of radius r , the centripetal acceleration is always perpendicular to the instantaneous velocity, and has magnitude v^2/r . Using

$$F = ma$$

we have

$$F = \frac{mv^2}{r}$$

as the constant magnitude of the inward force that acts on such a body of mass m .

$$F = m \frac{v^2}{r} = m\omega^2 r = mv\omega \quad \text{centripetal force}$$

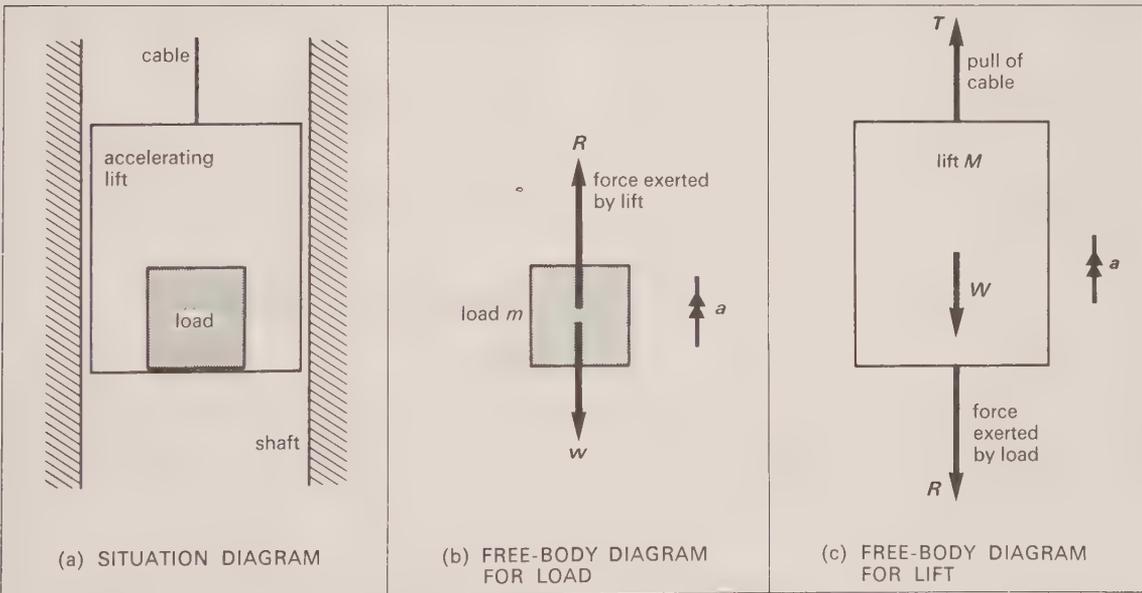


Fig. 3.4. *Newton's Second Law applied to linear motion.*

Notes

(a) *Newton's Third Law* indicates that there is an action–reaction pair in such a situation. The body under study *experiences a centripetal force* (towards the centre)—in some situations it may *exert a centrifugal force* (away from the centre) on some body in the environment. Since this centrifugal force is not acting *on* the body under study, it is not relevant to our analysis.

(Note that in general there is *no centrifugal force*. Two bodies such as the Sun and a planet attract mutually through a field of force: the force that each experiences and exerts is centrally directed.)

(b) The centripetal force is perpendicular to the instantaneous velocity. Sudden removal of this force would allow the body to assume uniform motion, which would be along the tangent to the circle.

(c) *Newton's Laws* are correctly applied by an observer in a *non-accelerated* (or **inertial**) *frame of reference*. A passenger in a car turning a corner is *not* such an observer, whereas (if we ignore the Earth's rotation) an observer standing on the pavement *is*. The passenger *seems* to experience an outward force. The passenger *actually experiences* a resultant *inward force from* the car upholstery: he *exerts* an equal *outward force on* the upholstery.

Examples of centripetal forces

Situation	Description of centripetal force
(1) Conker being whirled in a horizontal circle	(1) Force exerted by string (which is in tension)
(2) Earth in orbit around the Sun	(2) Gravitational force exerted by Sun
(3) Car turning a corner	(3) Frictional force exerted by road on tyres
(4) Aircraft banking	(4) Normal push of air
(5) Electron in hydrogen atom	(5) <i>Coulomb</i> attraction exerted by proton in nucleus
(6) Electron describing a circular path in a magnetic field	(6) Magnetic force exerted by agent that sets up magnetic field

Problem: at what angle should a road be banked at a bend to avoid the need for a centripetal frictional force? (fig. 3.5).

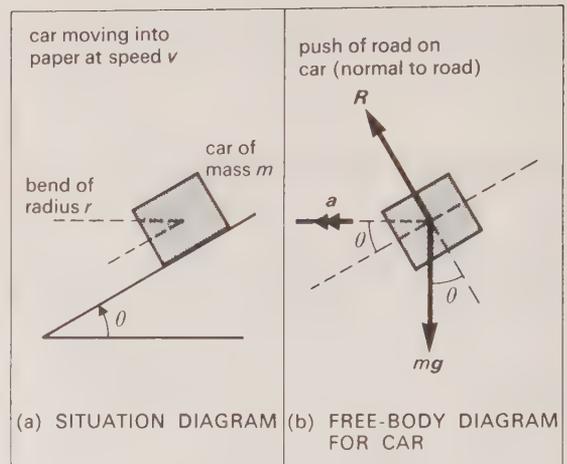


Fig. 3.5. The banking of a road on a bend.

Treat the car as a particle (p. 34) and resolve parallel to the line of greatest slope.

$$F = ma$$

$$\text{becomes } mg \sin \theta = m \left(\frac{v^2}{r} \cos \theta \right)$$

$$\tan \theta = \frac{v^2}{rg}$$

The equation (which should *not* be learned) shows that a bend of given radius r can only be banked in this way for cars of a given speed v .

(C) Simple harmonic motion

Further examples of the technique of using $F = ma$ are given in the section on s.h.m. (p. 82). In these examples a body is subjected to a force which varies with displacement and with time.

3.7 SKIDDING AND TOPPLING

Skidding

If a car of mass m travelling at speed v is to round a bend of radius r , it must experience a centripetal force mv^2/r . The *maximum* value of the centripetal frictional push F_{\max} exerted by the road on the tyres is given by

$$F_{\max} = \mu R = \mu mg \quad (\text{p. 36})$$

If the necessary centripetal force cannot be provided by this frictional push, then the car will skid.

So skidding occurs if

$$\frac{mv^2}{r} > \mu mg$$

or $v > \sqrt{(\mu gr)}$

Toppling

Under certain conditions the car will topple before it skids. A quantitative treatment of this situation is complicated by the spin angular momentum (p. 62). (We usually ignore the rotation of the car about its own centre of mass by treating it as a *particle*, but here we must apply forces to *different parts* of the car.)

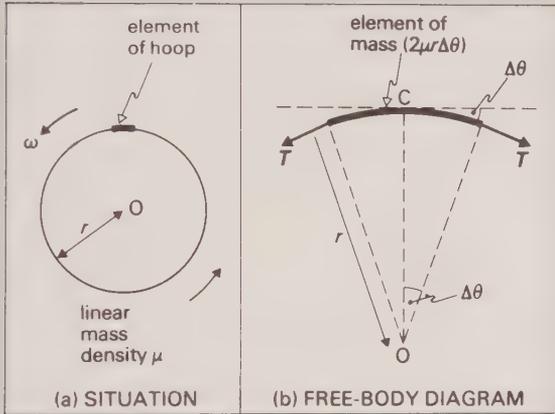


Fig. 3.6. A rotating hoop is in a state of tension.

The tendency to topple can be reduced by

- (a) lowering the centre of mass of the car, and
- (b) increasing the width of the axle.

Thus a racing car holds the road better if it has large tyres, and a low-slung frame with widely-spaced wheels.

► 3.8 THE TENSION IN A ROTATING HOOP

Fig. 3.6 shows a hoop of radius r rotating at angular speed ω about an axis through its own centre, and perpendicular to its plane.

The net centripetal force acting on the element of mass shown is

$$F = 2T \cos(90^\circ - \Delta\theta) \approx 2T\Delta\theta$$

if $\Delta\theta$ is small.

Apply $F = ma$ parallel to CO:

$$\begin{aligned} 2T\Delta\theta &= m\omega^2 r \\ &= (2\mu r \Delta\theta)\omega^2 r \\ \therefore T &= \mu r^2 \omega^2 \end{aligned}$$

The tension T in (for example) a car tyre is therefore proportional to the square of the speed. The stretching associated with it may cause the tyre to leave the rim of the wheel.

4 WORK AND ENERGY

4.1 WORK	42
4.2 ENERGY	43
4.3 FORMS OF ENERGY	44

4.4 ENERGY INTERCHANGES	45
4.5 CIRCULAR MOTION IN A VERTICAL PLANE	46
4.6 POWER P	47

4.1 WORK

In physics the term **work** always has this meaning:

Work is done when a force moves its point of application so that some resolved part of the displacement lies along the direction of the force.

(A) Work done by a constant force

Refer to fig. 4.1, which shows a constant force F . We define the **work** W done by the force F by the equation

$$W = Fs \cos \theta$$

$$W[\text{N m}] = F[\text{N}]s[\text{m}] \cos \theta$$

Notes

(a) W may be visualized as

(i) (force) \cdot (resolved part of displacement in direction of force), or

(ii) (resolved part of force in direction of displacement) \cdot (displacement).

In each case we obtain the same result.

(b) Work is a *scalar* quantity (obtained by a particular way of multiplying two vectors).

(c) The unit *newton metre* (N m) occurs so frequently it is conveniently called a **joule** (J): thus

$$1 \text{ joule} = 1 \text{ newton metre}$$

► (B) Work done by a variable force

The equation

$$W = Fs \cos \theta$$

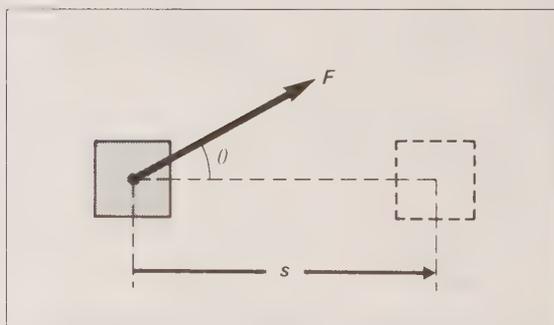


Fig. 4.1. Definition of work.

is more correctly written

$$\Delta W = F \Delta s \cos \theta$$

If the magnitude of F varies with s , the work δW done over a very small displacement δs is given by

$$\delta W = F \cos \theta \delta s$$

The work done over a finite displacement can be calculated from

$$W = \int_{s_1}^{s_2} F \cos \theta \, ds$$

Examples are given on pp. 174 and 353.

Examples of work done by forces

(a) If the *force does not move*, the work done is *zero*; e.g. a man supporting a beam. (There will,

nevertheless, be some conversion of chemical energy to thermal energy by muscular action.)

(b) If the force is perpendicular to the movement of the body, the work done is zero; e.g. the centripetal push of a road on a car, or the magnetic force acting on a charged particle made to move in a circular path. (The force does not change the body's speed.)

(c) If the force has a resolved part in the same direction as the displacement, then by our definition, the work is positive; e.g. the work done by a mass of gas when it increases its volume by pushing back its surroundings (p. 205). (This is consistent with the convention in thermodynamics which agrees to call the work done by the substance positive.)

(d) If the force has a resolved part along the line of the displacement which is oppositely directed, then the work done by that force is negative. (Since $\theta > 90^\circ$, $\cos \theta$ is negative.) Thus when sliding occurs between two surfaces in contact, the work done by the frictional forces is always negative, since the forces oppose relative motion.

4.2 ENERGY

Suppose body A exerts a contact force on body B, and does positive work: it follows that body B does negative work. Body A, having done positive work, is said to lose energy, while body B, having done negative work, is said to gain energy. The work done measures the interchange of energy between A and B.

A body which can do work has energy.

Energy and work are both scalars, and have the same dimensions and unit (joule). Our concept of energy is closely tied to our concept of force. In this book we are mainly concerned with gravitational and electromagnetic forces, and most of the energies we consider are derived from these two interactions.

Nevertheless, as with forces (p. 35) it is usually convenient to discuss energies in a descriptive way, and the following terms are used:

- (a) kinetic energy
- (b) potential energy—gravitational
—elastic
—electrostatic
- (c) thermal and internal energy
- (d) radiant energy
- (e) chemical energy
- (f) nuclear energy
- (g) mass energy.

On a microscopic scale, all forms of energy can be classified as either kinetic or potential.

Changes occur between different forms of energy, and the amounts possessed by different bodies, but, if we take all forms into account, we find

*There is no change in the total energy in the Universe. This is the **Law of the Conservation of Energy.***

If we appear to lose energy in some process, then rather than abandon this law, we look for a different form of energy.

Orders of magnitude

Description	Amount of energy W/J
binding energy of Earth–Sun system	10^{33}
energy radiated by Sun during 1 second	10^{26}
energy released by annihilation of 1 kg of matter	10^{17}
energy released by fission of 1 mole of $^{235}_{92}\text{U}$	10^{13}
electrical energy dissipated in a lightning discharge	10^{10}
energy released by combustion of 1 kg of petrol	10^8
energy required to charge a car battery	10^6
energy provided by a slice of bread	10^5
kinetic energy of a bowled cricket ball	10^1
energy stored in the electric field of a $10\ \mu\text{F}$ capacitor, charged to 20 V	10^{-3}
maximum energy of proton in CERN synchrotron	10^{-8}
work function of tungsten	10^{-18}
energy of visible light photon	10^{-19}

4.3 FORMS OF ENERGY

(A) Kinetic energy E_k

Suppose that a number of forces acting on a body of mass m have a resultant F , and that by acting over a displacement s (in the direction of F), F does work on the body, and thereby changes its velocity from u to v .

The acceleration a produced by F is uniform, and is therefore related to u , v and s by

$$v^2 = u^2 + 2as \quad (\text{p. 29})$$

Using $F = ma$

and $W = Fs$

we have $W = mas$

$$\begin{aligned} &= m\left(\frac{v^2}{2} - \frac{u^2}{2}\right) \\ &= \frac{1}{2}mv^2 - \frac{1}{2}mu^2 \end{aligned}$$

The work done by the force equals the change of the quantity $\frac{1}{2}m$ (velocity)².

The **kinetic energy** E_k of a body is defined by the equation

$$E_k = \frac{1}{2}mv^2$$

$$E_k[\text{J}] = \frac{1}{2}m[\text{kg}] v^2 \left[\frac{\text{m}^2}{\text{s}^2} \right]$$

The kinetic energy (k.e.) is a positive *scalar* quantity that represents the energy associated with the body *because of its motion*. It is equal to either

(a) the work done *by* the resultant force in accelerating the body from rest to an instantaneous speed v , or

(b) the work done *by* the body on some external agent which brings it to rest.

(a) and (b) are equivalent.

The relationship

$$W = \frac{1}{2}mv^2 - \frac{1}{2}mu^2$$

is sometimes called the **work–energy theorem**.

In words

$$\left(\begin{array}{l} \text{the work done by} \\ \text{the forces acting on} \\ \text{the body} \end{array} \right) = \left(\begin{array}{l} \text{the change of the} \\ \text{kinetic energy of the} \\ \text{body} \end{array} \right)$$

It is a very powerful principle, and is frequently used in solving problems.

(B) Potential energy

Potential energy (p.e.) is the energy possessed by a system by virtue of the relative positions of its component parts.

Examples

(a) *Gravitational p.e.*

(For a fuller discussion refer to p. 173.) Suppose we exert forces on a body of mass m and on the Earth, and thereby push the body to a rest position a vertical distance h above its initial position.

$$\text{We do work} \quad W = F \cdot s$$

$$E_p = mgh,$$

and yet there is no gain of k.e. The pulls of the Earth on the body and the body on the Earth have done *negative* work: we say the system has gained *gravitational p.e.* mgh .

When the system is released the two gravitational forces both do positive work on the body and on the Earth. Both, in principle, acquire k.e., but that gained by the Earth is negligible. The p.e. is associated with the relative positions (i.e. separation) of the two masses making up the system.

(b) *Elastic p.e. (stretching)*

Suppose we stretch a spring of force constant k from its natural length until the extension is x . Then we do work W given by

$$W = (\text{average force}) \cdot (\text{distance})$$

$$= \left(\frac{1}{2}kx\right) \cdot (x)$$

$$E_p = \frac{1}{2}kx^2$$

The spring also exerts a force of equal magnitude which is oppositely directed: this force does *negative* work, and we say the spring has gained *elastic p.e.* $\frac{1}{2}kx^2$.

If there were a body attached to the spring, and the spring released, the spring force would do positive work on the body, which would cause a gain of k.e. The spring force does this work by virtue of the change in the spring's *shape* during deformation. The components of this system are the molecules that make up the spring, and the shape of the spring determines their relative positions.

(c) *Elastic p.e. (twisting)*

Suppose that a pair of equal but oppositely directed torques T cause a rod (or wire) to be twisted through an angular displacement of θ . Then the equation

$$T = c\theta$$

defines the torsion constant c of that system (p. 148). The external agent that caused the twist has done work

$$W = (\text{average torque}) \cdot (\text{angle}) \\ = \left(\frac{1}{2}c\theta\right) \cdot (\theta) \quad (\text{p. 60})$$

$$\therefore E_p = \frac{1}{2}c\theta^2$$

This equals the negative work done by the rod. There is no gain of k.e., and so it represents the gain of the elastic p.e. of the system.

The components of the system are again the molecules that make up the rod. (The energy is electromagnetic in origin.)

(C) Internal energy

A frictional force always opposes relative motion, and when surfaces slide over one another, such a force always does negative work. This work represents energy being transferred to random molecular potential and kinetic energy (**internal energy**). (Frequently a negligible quantity of sound energy is produced, and this is eventually dissipated as internal energy.)

In fig. 4.2, B exerts a frictional force on A to the right, which moves its point of application to the left, and so does negative work. Macroscopically we see that A experiences a force which reduces its speed. Microscopically work is being done on a

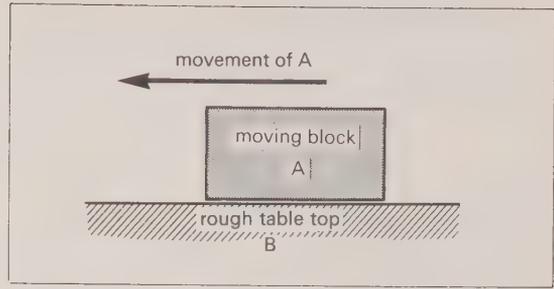


Fig. 4.2. Conversion of ordered energy to internal energy.

molecular scale that results in an increase of the random kinetic and potential energies of individual molecules. We observe a temperature increase along the common surface.

4.4 ENERGY INTERCHANGES

Consider a block sliding down a rough surface in the Earth's gravitational field. We can view the problem in three ways:

(1) The *total* energy of the system is constant (energy conservation). This is not helpful for problem solving because of the difficulty of keeping account of the internal energy.

Some examples of energy interchange

Situation	Energy converted		Is mechanical energy conserved?
	From	Into	
Bullet strikes a block and becomes embedded	k.e. of bullet	internal energy of block and bullet	no
Ball thrown into the air	k.e. of ball	p.e. of ball	yes, if we ignore air resistance
Pendulum swings in a vacuum	alternately k.e. of bob and p.e. of bob	alternately p.e. of bob and k.e. of bob	yes
Loudspeaker cone vibrates	(1) k.e. of cone (2) ordered energy of sound wave	(1) ordered energy of sound wave (2) internal energy of environment	(1) yes, but difficult to account for (2) no

Suppose now that frictional forces are absent.

(2) We can apply the work–energy theorem to the block

$$W = \text{change of body's k.e.}$$

where W is the work done by the environment on the body.

(3) We can view the *system* (Earth and block) as a whole, and say that any change of k.e. is accounted for by a reduction of the p.e. of the Earth–block system. Thus

$$(\text{gain of k.e.}) = (\text{loss of p.e.})$$

The definition of gravitational potential energy has led to a useful conservation law.

Law of conservation of mechanical energy

In the absence of forces other than gravitational forces the sum of the kinetic energies and mutual potential energy of the bodies in an isolated system remains constant.

This law is true by the definition of p.e.

► Potential energy and conservative forces

Potential energy can be defined for those forces which are **conservative**, such as the gravitational and electrostatic forces (p. 350).

A conservative force is one such that the work done in taking a body (or charge) from one point to another does *not* depend on the route taken (fig. 4.3). This implies

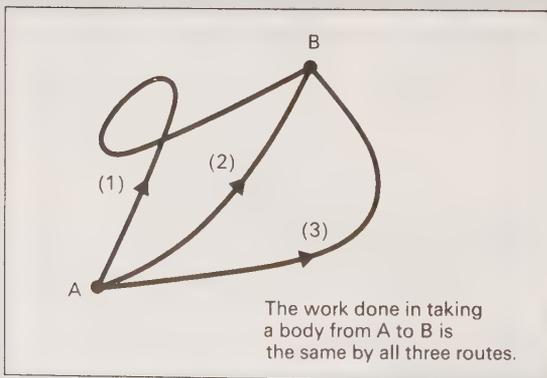


Fig. 4.3. A conservative force.

(a) that the difference of potential between two points depends only on the location of those points, and

(b) if a conservative force takes a body round a closed path (i.e. back to the starting point) then the total work done is zero.

The force of kinetic friction depends on the direction of a body's velocity, not just its position, and is therefore a **non-conservative** force. We cannot define a potential energy for such a force.

Notes

(a) Potential energy is really a property of a system of interacting particles: it depends on their relative positions.

(b) Because we are nearly always concerned with *differences* of potential energy, we are free to choose our own *zero*. For gravitational p.e. we usually choose the Earth's surface, or that of a body at infinity. (See also p. 350.)

4.5 CIRCULAR MOTION IN A VERTICAL PLANE

Fig. 4.4 shows a small mass m describing a circle in a vertical plane. The only forces acting on it are the pull mg of the Earth, and the centripetal pull F of the string; these forces are conservative, and so we may apply the law of conservation of mechanical energy.

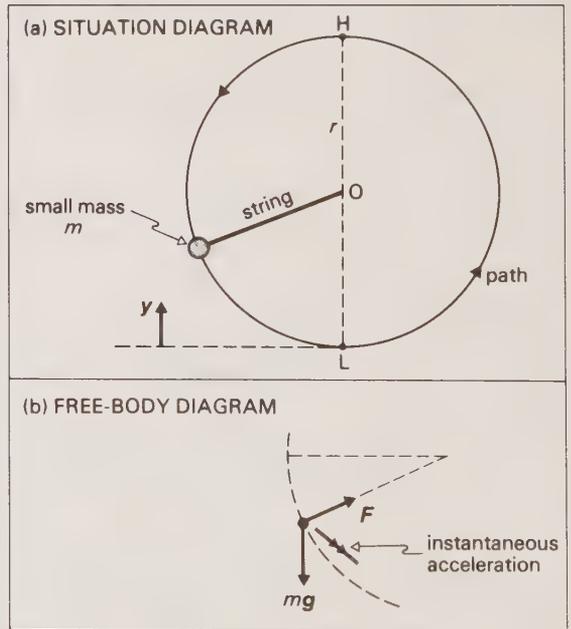


Fig. 4.4. Motion in a vertical circle.

Since $E_p + E_k = \text{constant}$
 $mg y + \frac{1}{2} m v^2 = \text{constant}$

At any instant the pull of the Earth may have a resolved part along the tangent to the circle, and this will cause a gain or loss of speed. The instantaneous speed can be found from the equation above, in which the constant will be determined by the starting conditions.

Minimum speed for a complete circle

When the body reaches H the net downward force on it is

$$F + mg$$

Newton's second law applied along the line HO gives

$$F + mg = m \left(\frac{v^2}{r} \right)$$

The minimum speed the body may have at this point if it is to describe a complete circle is obtained by putting $F = 0$.

Then $m \frac{v_{\min}^2}{r} = mg$

If $v < \sqrt{rg}$ at this point then the path would (temporarily) become parabolic.

4.6 POWER P

The power of a machine, or of the force exerted by a machine, tells us how fast it can transfer energy from itself to some other body.

Suppose a machine transfers energy ΔW over a time interval Δt . Then the **average power** P_{av} over the time interval Δt is defined by

$$P_{\text{av}} = \frac{\Delta W}{\Delta t}$$

$$P_{\text{av}} \left[\frac{\text{J}}{\text{s}} \right] = \frac{\Delta W [\text{J}]}{\Delta t [\text{s}]}$$

The **instantaneous power** P is defined by

$$P = \lim_{\delta t \rightarrow 0} \frac{\delta W}{\delta t} = \frac{dW}{dt}$$

Notes

(a) Power is a *scalar* quantity.

(b) The unit *joule (second)⁻¹* occurs so often it is called a **watt (W)**.

Thus $1 \text{ J s}^{-1} = 1 \text{ W}$

and it follows that

$$1 \text{ J} = 1 \text{ W s}$$

The *watt second* is a unit of *energy*.

(c) Power has dimensions $[\text{ML}^2\text{T}^{-3}]$.

Calculation of instantaneous power

Suppose a machine moves its point of application of a force F at velocity v in the direction of F .

Then since $\delta W = F \delta s$

and $P = \lim_{\delta t \rightarrow 0} \frac{\delta W}{\delta t}$

its instantaneous rate of working is given by

$$\begin{aligned} P &= \lim_{\delta t \rightarrow 0} F \frac{\delta s}{\delta t} \\ &= F \frac{ds}{dt} \\ &= Fv \end{aligned}$$

Thus the instantaneous power of a motor car depends not only on the force applied to it, but also on its instantaneous velocity.

5 MOMENTUM

5.1 IMPULSE AND MOMENTUM	48
5.2 CONSERVATION OF MOMENTUM	49
► 5.3 THE ROCKET	50
5.4 FORCE EQUALS RATE OF CHANGE OF MOMENTUM	50

5.5 COLLISIONS	52
5.6 THE BALLISTIC BALANCE	52
5.7 CONSERVATION OF MOMENTUM IN TWO AND THREE DIMENSIONS	53

5.1 IMPULSE AND MOMENTUM

On p. 34 the **momentum** p of a body was defined by

$$p = mv$$

and it was seen that we could measure **force** F from the equation

$$F = m \frac{dv}{dt}$$

when it acted on a body of constant mass.

(A) Constant force applied

Suppose a constant force F acts over a time interval Δt .

$$\begin{aligned} \text{Then} \quad F &= m \frac{\Delta v}{\Delta t} \\ F \Delta t &= m \Delta v \\ &= \Delta p \end{aligned}$$

where Δv and Δp are the velocity change and momentum change, respectively.

The **impulse of a constant force** is defined by the equation

$$\text{impulse} = F \Delta t$$

$$\text{impulse [N s]} = F[\text{N}] \cdot \Delta t[\text{s}]$$

Notes

(a) Impulse and momentum have the same dimensions $[\text{MLT}^{-1}]$. Both may be measured in N s or kg m s^{-1} .

(b) Impulse is a *vector* quantity, whose direction is that of F .

(c) The impulse of a steady force is illustrated by fig. 5.1(a).

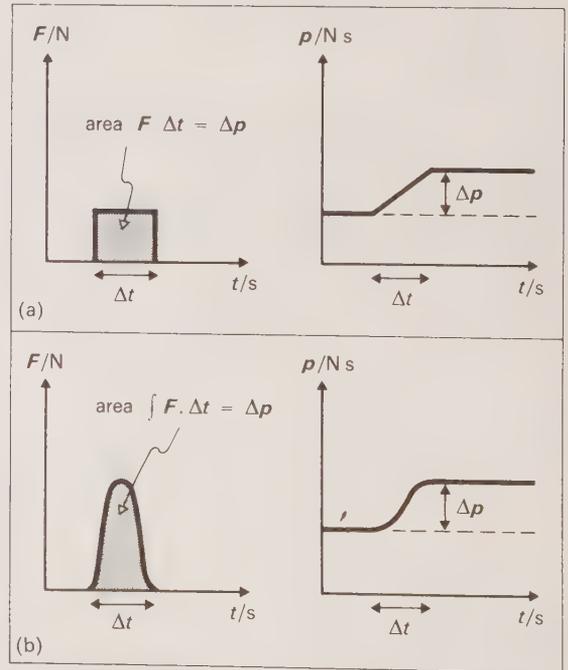


Fig. 5.1. Impulse and momentum change for (a) constant, and (b) variable forces.

(B) Variable force applied

From $F = m(dv/dt)$, integrating

$$\int F dt = m \int dv \\ = \Delta p$$

The **impulse of a variable force** is defined by

$$\text{impulse} = \int F dt$$

The impulse of a variable force is illustrated by fig. 5.1(b).

The impulse–momentum theorem

$$\int F dt = \Delta p \\ = mv - mu$$

where the impulse changes the velocity of a body of mass m from u to v .

The relationship

$$\int F dt = mv - mu$$

is sometimes called the **impulse–momentum theorem**.

In words

$$\left(\begin{array}{l} \text{impulse applied} \\ \text{to a body} \end{array} \right) = \left(\begin{array}{l} \text{change of momentum} \\ \text{experienced by a body} \end{array} \right)$$

This should be compared with the *work–energy theorem* (p. 44).

The term **impulsive force** is applied to a force which is variable in size, and which acts for a very short time only, such as that applied by a golf club to a ball. The impulse of the force *can* be measured from the vector change of momentum of the ball, even though the values of Δt , and that of F at each instant, are uncertain.

5.2 CONSERVATION OF MOMENTUM**Effects of external forces****(A) Single particle**

Fig. 5.2 shows a mass m acted upon by forces F_1 , F_2 , etc.

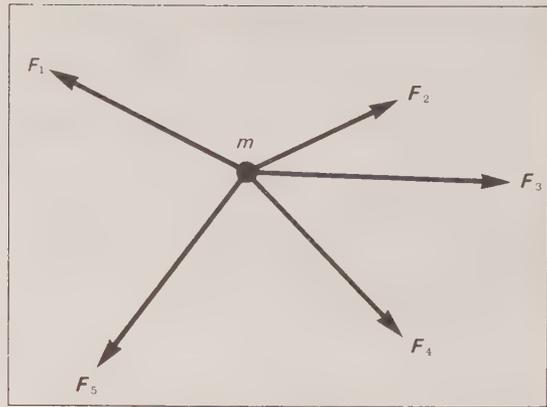


Fig. 5.2. Momentum conservation for a single particle.

According to *Newton's Second Law*, if there is a direction along which the sum of the resolved parts of the forces F is zero, then because there can be no acceleration in that direction, neither can there be any change of velocity or momentum.

(B) System of particles

As far as changes of momentum are concerned, the centre of mass of a system of interacting particles (defined on p. 68) behaves as though the whole mass of the system were concentrated at the centre of mass, and as though all the external forces were applied at the centre of mass. Thus the argument above, applied to a single particle, can be extended (as far as external forces are concerned) to systems of particles and bodies.

Effects of internal forces

These are forces exerted by one particle within the system on another *within the system*. (Such forces might be contact forces, or we could imagine separated masses to be connected by springs, etc.) *At every instant*, the instantaneous forces are an action–reaction pair of *Newton's Third Law*. Since change of momentum = $\int F dt$, a change of momentum experienced by one particle is accompanied by an equal but oppositely directed change of momentum experienced by the other particle. Since the *net* momentum change is found by *vectorial* addition, it is zero. This argument can be extended to all the pairs of *internal forces*, which thus produce *zero momentum change for the system*.

The law of conservation of linear momentum

This is a fundamental law of physics, and is applicable even in non-Newtonian mechanics.

The momentum of an isolated system is conserved.

An isolated system is one which experiences zero interaction with its environment, but of course in practice it is not easy to isolate a system. The following statement is a special case of the more general law which will be found very useful for solving problems.

If there is a direction in which no external forces act on a system, the total vector momentum of that system in that direction is constant (even if the bodies making up the system act on each other).

Using the law

- (1) Select the *system* to which it is to be applied.
- (2) Draw *before* and *after* diagrams of the situation, marking the masses and velocities of the bodies that constitute the system.
- (3) Marking a line along which the law is applicable, choose one direction as positive, and equate the total momentum before the interaction to that after.

Examples of conservation of momentum

(a) Before a gun is fired, the total momentum of the system, consisting of gun plus bullet, is zero. Immediately after firing, the total momentum must remain at zero and so the gun recoils with a momentum equal and opposite to that of the bullet.

(b) When a bomb explodes in mid-air, the total momentum of all the fragments just after the explosion must equal the total momentum of the complete bomb just before exploding.

(c) When a stationary nucleus disintegrates, the total momentum of the resulting nucleus and the emitted particles must add up to zero. Apparent discrepancies have led to the discovery of new fundamental particles.

The rocket and the ballistic balance are given as further examples.

► 5.3 THE ROCKET

Refer to fig. 5.3. The exhaust gases have a fixed speed u relative to the rocket, and Δm is the change in the mass of the rocket (which will, of course, be negative).

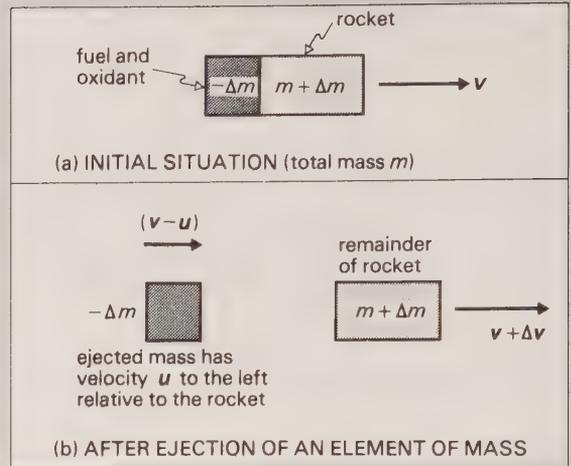


Fig. 5.3. The law of conservation of linear momentum applied to the rocket.

In (a) the linear momentum p to the right is given by

$$p = mv$$

In (b) it may be written

$$p = (m + \Delta m)(v + \Delta v) + (-\Delta m)(v - u)$$

Applying the law of conservation of linear momentum

$$mv = (m + \Delta m)(v + \Delta v) + (-\Delta m)(v - u)$$

Neglecting the term $\Delta m \Delta v$, this gives

$$m \Delta v = -u \Delta m$$

or

$$dv = -u \left(\frac{dm}{m} \right)$$

Suppose the rocket starts from rest with an initial mass m_0 . If the speed has become v by the time the mass has been reduced to m ,

$$\int_0^v dv = -u \int_{m_0}^m \frac{dm}{m}$$

$$\therefore v = -u \ln(m/m_0)$$

$$\therefore m = m_0 \exp(-v/u)$$

This result can also be deduced from *Newton's second law* (p. 34).

5.4 FORCE EQUALS RATE OF CHANGE OF MOMENTUM

Suppose we write *Newton's Second Law* in the equation

$$F = \frac{d}{dt}(mv) \\ = m \frac{dv}{dt} + v \frac{dm}{dt}$$

We can use the equation conveniently in three forms.

(A) Constant mass, variable velocity

Since $\frac{dm}{dt} = 0$

we use $F = m \frac{dv}{dt} = ma$

This is the commonest use: an example is given on p. 39.

(B) Constant velocity, variable mass

Since $\frac{dv}{dt} = 0$

we use $F = v \frac{dm}{dt}$

dm/dt is the rate at which a body changes its mass while moving at velocity v relative to the observer.

Examples

(a) *The conveyor belt.* If sand falls vertically at a mass rate of flow of dm/dt onto a conveyor belt which moves at a horizontal velocity v , the horizontal resolved part of the force exerted by the belt on the sand is $v(dm/dt)$.

(b) *The hovering helicopter.* Suppose the helicopter imparts a downward velocity v to air of density ρ which it collects over an area A .

In time Δt

$$\Delta m = \rho \times (\text{volume}) \\ = \rho Av \Delta t$$

$$\Delta m / \Delta t = \rho Av$$

The downward force exerted by the helicopter on the air

$$F = v(\Delta m / \Delta t) \\ = \rho Av^2$$

The air exerts an equal but opposite force on the helicopter. v can be adjusted for this to equal the pull of the Earth on the helicopter.

(c) *A molecular beam.* Consider a molecular beam of cross-sectional area A , containing a number density n of molecules, and moving at velocity c , to be incident on a wall which absorbs the

molecules without rebound. The mass of molecules hitting the wall in time Δt will be

$$\Delta M = nmAc \Delta t$$

where m is the mass of each molecule.

The force exerted on the wall will be to the right, and of size

$$F = c \frac{dM}{dt} \\ = c(nmAc) \\ = nmAc^2$$

The pressure $p = F/A = nmc^2$.

The reader should consider the effect on the pressure of

- (i) the molecules rebounding elastically,
- (ii) the molecular beam being directed obliquely, or
- (iii) the molecules having random velocities (p. 198).

It should be noted that the viewpoints (1) and (2) (below) are equivalent. Consider an *element* of sand from situation (B)(a), whose mass is Δm , and whose velocity is changed from 0 to Δv over a time interval Δt . (For convenience suppose the acceleration to be uniform, though this is not necessary.)

We could write

$$(1) \quad F = m \frac{dv}{dt} \quad \text{as} \quad F = (\Delta m) \left(\frac{\Delta v}{\Delta t} \right)$$

$$(2) \quad F = v \frac{dm}{dt} \quad \text{as} \quad F = (\Delta v) \left(\frac{\Delta m}{\Delta t} \right)$$

If (2) we effectively apply $F = ma$ to an *element* of mass Δm .

► (C) Variable mass and variable velocity

Refer back to the diagram of the rocket, fig. 5.3. The force exerted on the rocket (the thrust) has a size $m(dv/dt)$, and is to the right. The force exerted on the exhaust gases has a size $u(dm/dt)$, and is to the left. Since the total force exerted on the system is zero

$$m \frac{dv}{dt} = -u \frac{dm}{dt}$$

or

$$dv = -u \left(\frac{dm}{m} \right)$$

This is the same result as that found from the law of conservation of linear momentum (p. 49).

5.5 COLLISIONS

A **collision** is a process in which the time interval during which bodies interact is small relative to the time for which we can observe them: we can then make a clear distinction between *before* and *after*. If such a collision takes a small time interval, large impulsive forces are brought into play. We can usually ignore the relatively small change of momentum brought about by an external force compared to that imposed on the bodies that collide by the impulsive (internal) forces.

We use the law of conservation of momentum in the form:

The total momentum of a system just before collision is equal to that just after collision if the time of collision is sufficiently small.

Experimental verification of this statement in one dimension is conveniently carried out by studying the collisions of suitable vehicles on an air track. The velocities of the vehicles, of measured mass, can be deduced from stroboscopic photographs or from electronic timing using light beams and photodiodes.

Classification of collisions

Energy is always conserved (when all forms of energy are considered), but when mechanical energy is converted to internal energy it is not always possible to apply this principle usefully.

(a) **Elastic collisions** are those in which kinetic energy is conserved. Truly elastic collisions can only occur in practice on an atomic scale. (Even then they are not *always* elastic.)

(b) **Inelastic collisions** are those in which kinetic energy is not conserved: it may be converted to internal energy (as usually happens), or perhaps elastic potential energy of deformation. On a macroscopic scale this is the most common type of collision.

A **completely inelastic** collision is one in which two bodies stick together after impact (as a bullet being embedded in a target). The loss of kinetic energy is large but not complete.

(c) **Explosive collisions** are those in which there is an *increase* of kinetic energy. This could occur if potential energy were released by the impact, as when the collision of one trolley with another releases a compressed spring. (Then the action–reaction forces of separation would exceed those of approach.) These collisions are sometimes called **superelastic**—see p. 198.

Note carefully:

(1) momentum *is* always conserved, but

(2) kinetic energy is *not* always conserved, as it is usually converted to some other form of energy.

The coefficient of restitution

By definition

$$e = \frac{\text{relative speed of separation}}{\text{relative speed of approach}}$$

for two bodies that collide. The speeds are measured along their common normal. For a given pair of bodies e is roughly constant. This is sometimes called *Newton's experimental law of impact*.

e gives us an alternative way of classifying collisions:

e	Type
>1	explosive or superelastic
1	elastic
<1	inelastic
0	completely inelastic

5.6 THE BALLISTIC BALANCE

This device is discussed here to illustrate the conservation laws of momentum and mechanical energy. Refer to fig. 5.4.

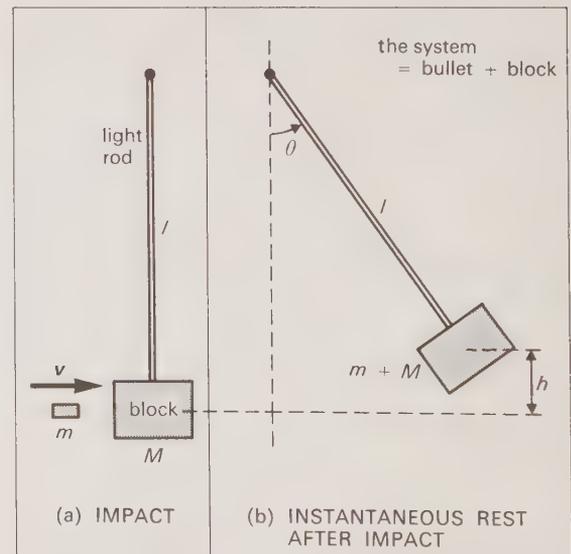


Fig. 5.4. The ballistic balance.

(A) The impact

During this process

(a) mechanical energy is *not* conserved (because of the non-conservative internal frictional force), and internal energy is produced, but

(b) linear momentum *is* conserved in the horizontal direction, along which there is no external force.

Put V = the block's velocity immediately after impact.

Using (b), and taking the horizontal direction to the right as positive, we have

$$mv = (M + m)V$$

(B) The swing after impact

During this process

(a) mechanical energy *is* conserved: the conservative gravitational force causes the conversion of k.e. to p.e., but

(b) momentum is *not* conserved, as an external resultant force (the pull of the Earth) acts on the bullet-block system for a significant time interval.

Using (a)

$$\frac{1}{2}(m + M)V^2 = (m + M)gh$$

$$V = \sqrt{(2gh)}$$

Uses of the balance

(a) As described, the balance can be used to measure the speed of the bullet.

(b) With modification the balance can be used to *compare masses*. m could be a second pendulum which is made to interact with the block M .

Instead of $\frac{m_1}{m_2} = \frac{a_2}{a_1}$ (p. 37)

we use $\frac{m}{M + m} = \frac{V}{v}$

in which both v and V are measured. This can be done directly (photographically) or indirectly (by measuring θ).

Note in fig. 5.4(b), that $h = l(1 - \cos \theta)$.

This experiment illustrates the principle used by Chadwick (1932) to establish the existence of the **neutron**. He bombarded a block of beryllium with α -particles from a polonium source and this led to the emission of neutrons. These neutrons then collided with paraffin wax and knocked out hydrogen nuclei (protons) at high speeds, which were detected in an ionization chamber. In a separate experiment he replaced the paraffin wax with paracyanogen so that nitrogen nuclei were

knocked out. The maximum speeds of the hydrogen and nitrogen nuclei were measured from ionization chamber data. This led to the deduction of the neutron mass by assuming that the neutron collisions were elastic and that momentum was conserved.

5.7 CONSERVATION OF MOMENTUM IN TWO AND THREE DIMENSIONS

Many collisions between two bodies are *oblique*, rather than head-on, and in these cases the two initial velocity vectors define a plane which is *two-dimensional*. After the collision, the plane defined by the velocities of the two bodies may be different from the initial plane and so the collision becomes *three-dimensional*. In all cases, momentum is conserved but two- and three-dimensional situations are more complicated because vectorial additions are involved.

Two-dimensional collisions can be studied in the laboratory using stroboscopic photographs of magnetic pucks supported by air or by CO_2 . In fig. 5.5, puck B is at rest and puck A, of equal mass, is approaching it obliquely.

Conservation of momentum gives

$$mu_A = mv_A + mv_B$$

$$\therefore u_A = v_A + v_B$$

Conservation of k.e. gives

$$\frac{1}{2}mu_A^2 = \frac{1}{2}mv_A^2 + \frac{1}{2}mv_B^2$$

$$\therefore u_A^2 = v_A^2 + v_B^2$$

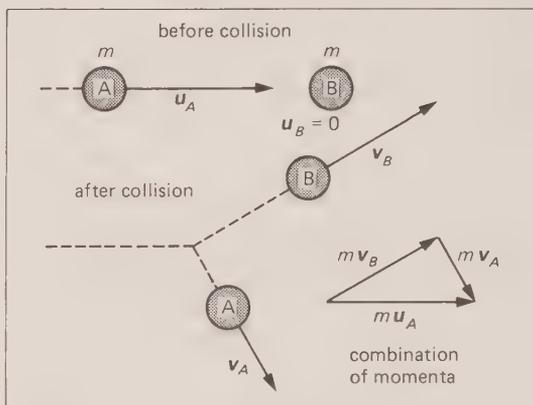


Fig. 5.5. An oblique collision between two pucks of equal mass.

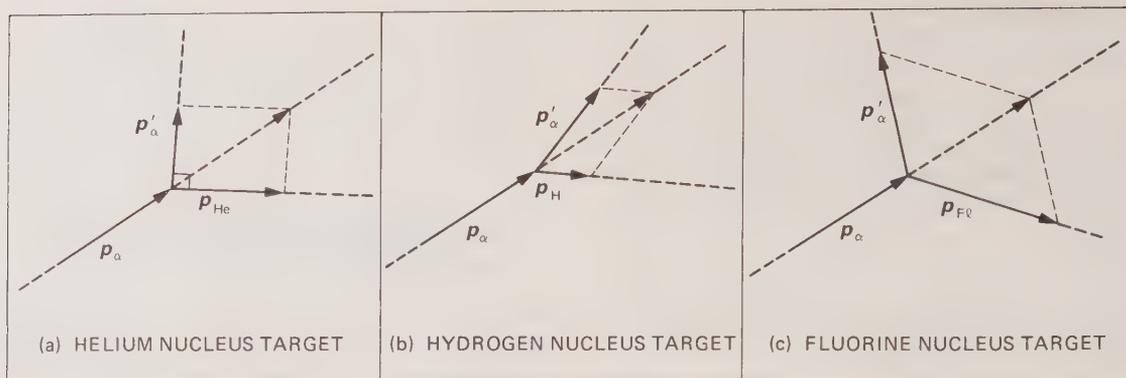


Fig. 5.6. A helium nucleus (α -particle) colliding with different target nuclei.

By considering *Pythagoras's theorem*, it is clear that this special case requires the angle between v_A and v_B to be 90° , whatever the directions of v_A and v_B .

If $m_A > m_B$, the angle would be less than 90° .

If $m_A < m_B$, the angle would be greater than 90° .

Using these ideas, important deductions can be made from cloud chamber photographs of trajec-

tories of particles undergoing collisions. Fig. 5.6 shows some possible results, with a helium nucleus (α -particle) as the incident particle in each case. Result (a) indicates that α -particles have the same mass as helium nuclei. Stereoscopic photographs are usually required for full analysis because all three tracks are rarely confined to the plane of the photograph. New particles have been identified using this technique.

6 ROTATIONAL DYNAMICS

6.1 INTRODUCTION	55
6.2 ROTATIONAL KINEMATICS	55
6.3 TORQUE AND ROTATIONAL MOTION	56
6.4 CALCULATION OF MOMENT OF INERTIA	57
6.5 MOMENTS OF INERTIA THEOREMS	59
6.6 ENERGY OF ROTATIONAL MOTION	59

▶ 6.7 ANGULAR MOMENTUM AND ITS CONSERVATION	61
6.8 THE COMPOUND PENDULUM	62
6.9 OSCILLATORY DETERMINATIONS OF MOMENT OF INERTIA	63
6.10 COMPARISON OF LINEAR AND ROTATIONAL DYNAMICS	64

6.1 INTRODUCTION

A **rigid body** is one that suffers a negligible deformation when subjected to external forces. Such a body undergoes motion which is made up of *translation* and *rotation*. Motion is purely

(a) **translational** if every particle of a body has the same instantaneous velocity,

(b) **rotational** if every particle moves in a circle about the same straight line (the axis of rotation).

(a) To describe translational motion we use the equation $F = Ma$, in which F is the resultant force applied to a *body*, M is the mass of the body, and a is the acceleration of the *centre of mass* (p. 68).

(b) For the rotational motion we relate quantities that are analogous to F , M and a .

6.2 ROTATIONAL KINEMATICS

Angular velocity ω is defined on p. 31. **Average** and **instantaneous angular accelerations** α are defined by the equations

$$\alpha_{av} = \frac{\Delta\omega}{\Delta t}$$

$$\alpha = \lim_{\delta t \rightarrow 0} \frac{\delta\omega}{\delta t} = \frac{d\omega}{dt}$$

$$\alpha_{av} \left[\frac{\text{rad}}{\text{s}^2} \right] = \frac{\Delta\omega [\text{rad/s}]}{\Delta t [\text{s}]}$$

Angular acceleration α is a *pseudovector* quantity (see p. 19), measured in rad s^{-2} . It has dimensions $[\text{T}^{-2}]$.

Provided the angular acceleration is constant, the equations of uniformly accelerated motion (p. 29) follow:

$\omega = \omega_0 + \alpha t$ $\theta = \frac{\omega_0 + \omega}{2} \cdot t$ $\theta = \omega_0 t + \frac{1}{2} \alpha t^2$ $\theta = \omega t - \frac{1}{2} \alpha t^2$ $\omega^2 = \omega_0^2 + 2\alpha\theta$	<i>Uniformly accelerated angular motion</i>
---	---

θ represents the angular *displacement* (not the total angle). These equations can be obtained by integration of the defining equations $\omega = d\theta/dt$, and $\alpha = d\omega/dt$. (Note that if a point describes several revolutions, then θ is of the form $[2n\pi \text{ rad} + \phi]$.)

Linear and rotational kinematics related

Suppose we have a particle distance r from the axis of rotation of a rigid body, and that there is no translational motion of the body.

(a) *Displacements* are related by

$s = r\theta$	(p. 22)
---------------	---------

Linear and rotational kinematics compared

Linear motion		Rotational motion	
quantity	dimensions	quantity	dimensions
$s[m]$	[L]	$\theta[\text{rad}]$	none*
$v = \frac{ds}{dt}$	$[LT^{-1}]$	$\omega = \frac{d\theta}{dt}$	$[T^{-1}]$
$u = (v)_{t=0}$	$[LT^{-1}]$	$\omega_0 = (\omega)_{t=0}$	$[T^{-1}]$
$a = \frac{dv}{dt}$	$[LT^{-2}]$	$\alpha = \frac{d\omega}{dt}$	$[T^{-2}]$

* Although it has been argued that the assignment of no dimensions to angle can lead to inconsistencies.

(b) Speeds

$$\frac{ds}{dt} = \theta \frac{dr}{dt} + r \frac{d\theta}{dt}$$

$$v = r\omega$$

($dr/dt = 0$ since r is constant for a particle of a rigid body.)

v is the instantaneous linear tangential speed.

(c) Accelerations

The total acceleration of the particle is made up of two components.

(i) Tangential linear acceleration a_t

$$\frac{dv}{dt} = r \frac{d\omega}{dt}$$

$$a_t = r\alpha$$

This will be zero when $\alpha = 0$.

(ii) Radial (centripetal) linear acceleration a_r

$$a_r = \frac{v^2}{r} = \omega^2 r$$

(p. 33)

This will be zero only when $\omega = 0$.

The total acceleration must be found by vector addition: it has a magnitude

$$a = r\sqrt{(\omega^4 + \alpha^2)}$$

An important distinction follows from these equations: all points on a rigid body undergoing rotation about a fixed axis necessarily have the same angular displacement, speed and acceleration. Their linear values, however, depend on r .

6.3 TORQUE AND ROTATIONAL MOTION

The moment of a force, or its torque, T , is defined on p. 65 by $T = Fr$.

In fig. 6.1, F is the tangential component of the resultant force that acts on the particle. (The radial component cannot produce angular acceleration.) Using Newton's Second Law

$$F = ma_t$$

$$= mr\alpha$$

Thus torque

$$T = (mr\alpha)r$$

$$= mr^2\alpha$$

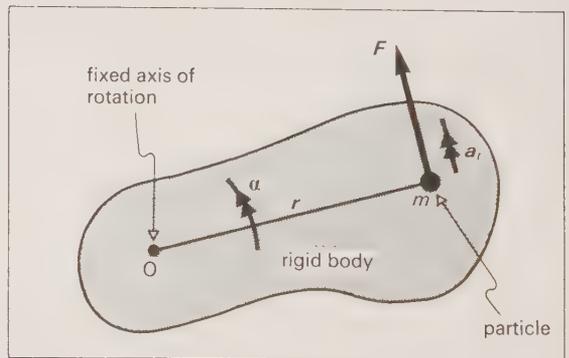


Fig. 6.1. Newton's second law applied to a particle of a rigid body.

Summing for the whole body

$$\Sigma T = \Sigma mr^2\alpha$$

or $\Sigma T_{\text{ext}} = \alpha \Sigma mr^2$

since

(a) α has the same value for all such particles m on the body, and

(b) the sum of *internal torques* is zero, in the same way that it was shown on p. 49 that the sum of the *internal forces* is zero.

Moment of inertia (m.i.)

The **moment of inertia** I of a body about a particular axis is defined by the equation

$$I = \Sigma mr^2$$

where r is the distance of a particle of mass m from the axis of rotation.

$$I[\text{kg m}^2] = \Sigma m[\text{kg}] r^2[\text{m}^2]$$

Notes

(a) The moment of inertia (m.i.) of a body is a *scalar* quantity.

(b) The dimensions of I are $[\text{ML}^2]$.

(c) I depends on

(i) the mass of a body,

(ii) the way the mass is distributed,

(iii) the *axis of rotation* (and is therefore not a constant for a rigid body).

Newton's Second Law for rotation about a fixed axis can now be written in the form

$$T = I\alpha$$

where T represents the *resultant torque* acting on the body.

Example. The simple pendulum (fig. 6.2)

Choose the anticlockwise sense of rotation as positive. For the bob

$$I = \Sigma mr^2 = ml^2$$

$$T = -(mg \sin \theta)l$$

(since P , a radial force, makes zero contribution),

$$\alpha = \ddot{\theta}$$

so

$$T = I\alpha$$

becomes

$$-mg \sin \theta l = ml^2\ddot{\theta}$$

Rearranging,

$$\ddot{\theta} = -\left(\frac{g}{l}\right) \sin \theta$$

$$\approx -\left(\frac{g}{l}\right) \theta \quad \text{for small } \theta$$

We conclude (p. 83) that the motion is simple harmonic for small amplitudes, and of period $2\pi\sqrt{l/g}$.

A graph of T^2 against l has a gradient $4\pi^2/g$, from which a reasonably reliable value for g can be determined as long as the amplitude is small, the support is rigid and the air resistance is negligible. The T values must be calculated from large numbers of oscillations and the l values measured from the centre of the bob.

The analysis of the torsional pendulum follows the same pattern (p. 148).

6.4 CALCULATION OF MOMENT OF INERTIA

Suppose a body of mass M has a moment of inertia I . The **radius of gyration** k is defined by the equation

$$I = Mk^2$$

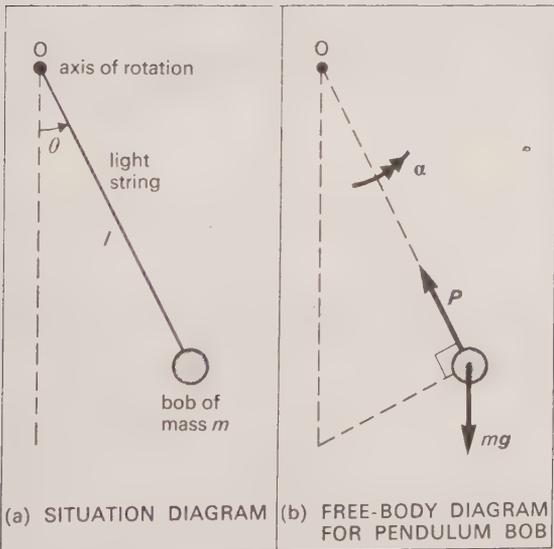


Fig. 6.2. Application of $T = I\alpha$ to the simple pendulum.

A point mass M placed a distance k from the axis of rotation would have the same value of I . (This is *not* the location of the centre of mass, which may coincide with the axis of rotation.)

► **Calculation of I by calculus**

$I = \Sigma mr^2$ enables us to calculate I for a body composed of discrete point masses.

For a continuous distribution of matter the summation is replaced by the integration

$$I = \int r^2 dm$$

where the integration is taken over the whole body.

Examples

Refer to fig. 6.3. In each case, C represents the centre of mass of the body.

(a) *A thin ring about a perpendicular axis through its centre.* Let the ring in fig. 6.3(a) consist of a number of discrete point masses m , giving a total mass M . The distance from the axis is the same for each point mass.

$$I = \Sigma mr^2 = R^2 \Sigma m$$

$$\therefore I = MR^2$$

This expression would also apply to a hollow cylinder about a similar axis.

(b) *A circular disc about a perpendicular axis through its centre.* The disc in fig. 6.3(b) can be considered as a set of concentric rings of thickness δr . If M is the total mass of the disc, the mass of the ring δm is given by

$$\delta m = \frac{2\pi r \delta r}{\pi R^2} \cdot M = \frac{2M}{R^2} \cdot r \delta r$$

$$\therefore I = \frac{2M}{R^2} \int_0^R r^3 dr$$

$$\therefore I = \frac{1}{2} MR^2$$

This expression would also apply to a solid cylinder about a similar axis.

(c) *A thin rod about a perpendicular axis through its centre.* The rod in fig. 6.3(c) can be considered as a series of elements of length δr . If M is the total mass of the rod, the mass of an element δm is given by

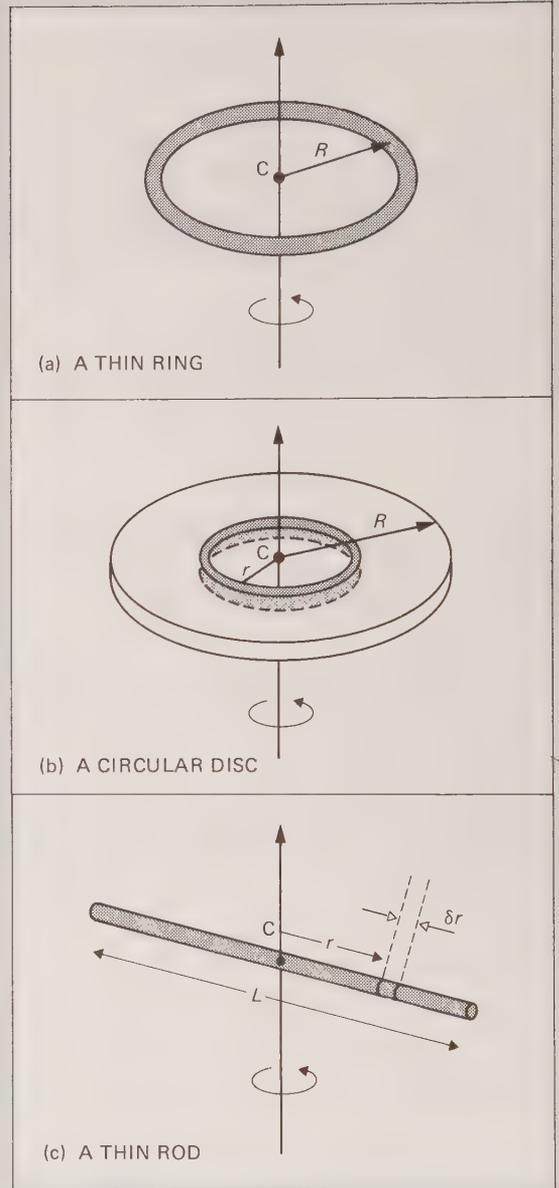


Fig. 6.3. Rotation of some simple objects.

$$\delta m = M \cdot \frac{\delta r}{L}$$

$$\therefore I = \frac{M}{L} \int_{-L/2}^{+L/2} r^2 \cdot dr$$

$$\therefore I = \frac{1}{12} ML^2$$

This expression would also apply to a flat, rectangular plate or board about an axis perpendicular to L , passing through C and lying in the plane of the lamina.

For some simple bodies it is more convenient to use **Routh's Rules**:

Suppose a body of mass M has three mutually perpendicular axes of symmetry. Let $2a$ and $2b$ be lengths of two of these axes. Then the m.i. I_0 about the third axis (which will be through the centre of mass) is given by

$$I_0 = M \left(\frac{a^2 + b^2}{n} \right)$$

where $n = 3$ for rectangular bodies,
 $n = 4$ for elliptical bodies,
 $n = 5$ for ellipsoidal bodies.

The rules have no theoretical foundation.

Examples

(a) Rectangular lamina, length l and breadth h

$$I_0 = M \left(\frac{l^2 + h^2}{12} \right)$$

for an axis perpendicular to the plane of the lamina.

(b) Sphere, radius r

$$I_0 = \frac{2}{5} Mr^2$$

The rules calculate I_0 , the m.i. about an axis through the centre of mass for a homogeneous body.

6.5 MOMENTS OF INERTIA THEOREMS

The calculation of moments of inertia about unusual axes can often be simplified by using the following theorems in conjunction with a few well-known results.

(A) The theorem of parallel axes

This can be applied to any body of mass M . Let

I_0 be the m.i. about a particular axis through the centre of mass, and

I be the m.i. about some other parallel axis distance d from the first.

Then the theorem of parallel axes states

$$I = I_0 + Md^2$$

It is clear that the m.i. of a body about a given axis is always greater than its m.i. about a parallel axis through C .

Example. A solid cylinder of mass M and radius R rotates about an axis in its surface parallel to its axis through C .

m.i. about this axis,

$$I = \frac{1}{2} MR^2 + MR^2 = \frac{3}{2} MR^2$$

(B) The theorem of perpendicular axes

This can be applied only to flat objects that can be regarded as having negligible thickness. If I_x and I_y are the m.i.s of a flat object about perpendicular axes in its plane, and I_z is its m.i. about an axis perpendicular to its plane, then it follows from *Pythagoras's theorem* that

$$I_z = I_x + I_y$$

Example. A circular disc of mass M and radius R rotates about an axis which lies along a diameter.

m.i. about this axis = $I_x = I_y$

$$\therefore I_z = 2I_x = \frac{1}{2} MR^2$$

$$\therefore I_x = \frac{1}{4} MR^2$$

Some situations may require the application of *both* theorems in turn, for example the m.i. of a circular disc rotating about an axis in the plane of the disc and tangential to its edge.

If a number of objects are combined so that they have the same axis of rotation, the m.i. of the combination is the sum of the separate m.i.s about that axis. This can lead to the experimental determination of an unknown m.i. in terms of a known m.i.

6.6 ENERGY OF ROTATIONAL MOTION

(A) Work done by a constant torque

Refer to fig. 6.4 overleaf, in which F is the tangential component of the resultant force acting at P .

Suppose F is applied while the body undergoes an angular displacement θ . Then F does work W , where

$$W = F (\text{distance along arc})$$

$$= Fr\theta$$

$$= T\theta \quad \text{since } T = Fr$$

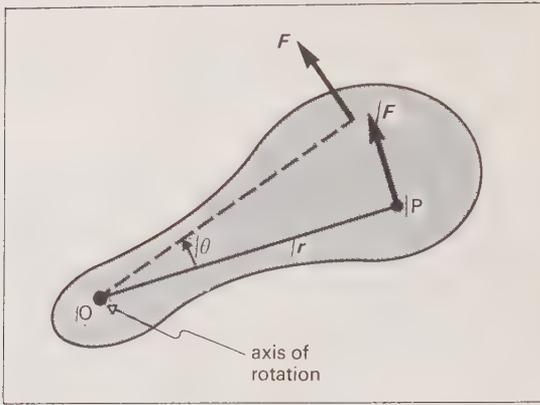


Fig. 6.4. Calculation of work done by a torque.

$$W = T\theta$$

$$W[J] = T[\text{N m}]\theta[\text{rad}]$$

When T and θ are in the same sense, W is positive. The force transfers energy to the body, whose (rotational) k.e. increases. (When the direction of T and the axis of rotation are perpendicular, $W = 0$.)

► (B) *Work done by a variable torque*

The equation $W = T\theta$ is more correctly written $\Delta W = T\Delta\theta$. Then if the magnitude of T varies with θ , the work δW done over a very small displacement $\delta\theta$ is given by

$$\delta W = T\delta\theta$$

The work done over a finite displacement can be calculated from

$$W = \int_{\theta_1}^{\theta_2} T d\theta$$

An example is given on p. 342.

Rotational kinetic energy

Suppose that the angular velocity of a rigid body is increased from ω_0 to ω when a constant resultant torque T acts through a displacement θ . Since α is constant, we can apply

$$\omega^2 = \omega_0^2 + 2\alpha\theta \quad (\text{p. 55})$$

in which $T = I\alpha$
 Using $W = T\theta$
 we have $W = I\alpha\theta$

$$= I\left(\frac{\omega^2}{2} - \frac{\omega_0^2}{2}\right)$$

$$W = \frac{1}{2}I\omega^2 - \frac{1}{2}I\omega_0^2$$

This represents the **work-energy theorem** for rotational motion. In words

$$\left(\begin{array}{c} \text{The work done by} \\ \text{the torques acting} \\ \text{on the body} \end{array}\right) = \left(\begin{array}{c} \text{The change of the} \\ \text{rotational kinetic} \\ \text{energy of the body} \end{array}\right)$$

The **rotational k.e.** of a body can be calculated from

$$E_k = \frac{1}{2}I\omega^2$$

$$E_k[J] = \frac{1}{2}I[\text{kg m}^2]\omega^2\left[\frac{\text{rad}^2}{\text{s}^2}\right]$$

$\frac{1}{2}I\omega^2$ is a convenient way of writing $\Sigma \frac{1}{2}mv^2$, since

$$\begin{aligned} \Sigma \frac{1}{2}mv^2 &= \frac{1}{2}\Sigma mr^2\omega^2 \\ &= \frac{1}{2}\omega^2 \Sigma mr^2 \\ &= \frac{1}{2}I\omega^2 \end{aligned}$$

It is more convenient to use ω than v , because all points on the rigid body have the *same value of* ω (for a given axis of rotation).

A flywheel is designed so that most of its mass is concentrated at the rim. Thus the flywheel has a high moment of inertia and, for a given angular speed, high kinetic energy. This finds application in piston engines, steam rollers, friction-drive toy cars, etc.

Total kinetic energy of a rigid body

The total kinetic energy is made up by

$$\begin{aligned} &(\text{translational k.e.}) + (\text{rotational k.e.}) \\ &= \frac{1}{2}Mv^2 + \frac{1}{2}I_0\omega^2 \end{aligned}$$

in which M is the total mass,
 v is the velocity of the centre of mass,
 I_0 is the m.i. about an axis through the centre of mass, and
 ω is the angular velocity about that axis.

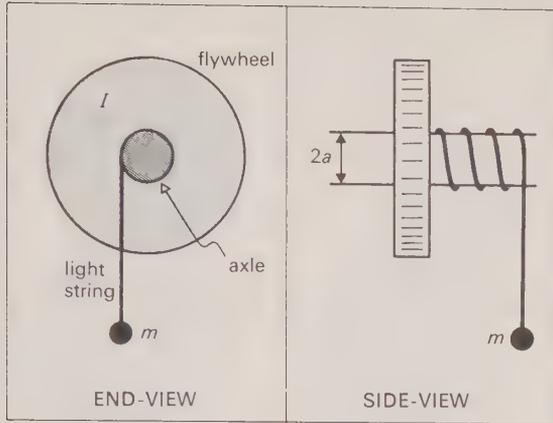


Fig. 6.5. Measurement of the moment of inertia of a flywheel.

Measuring moment of inertia

Refer to fig. 6.5.

Suppose the mass m falls a vertical distance h from rest, and acquires a speed v .

$$\text{Loss of gravitational p.e.} = mgh$$

$$\text{Gain of k.e.} = \frac{1}{2}I\omega^2 + \frac{1}{2}mv^2$$

If only conservative forces act, these energies are equal:

$$mgh = \frac{1}{2}I\omega^2 + \frac{1}{2}mv^2$$

The values of m , g and h are known, ω and v can be found from a knowledge of t , the time of fall, h , the distance fallen, and a , the radius of the axle:

$$v = 2\left(\frac{h}{t}\right)$$

$$\omega = \frac{v}{a}$$

We can correct for the frictional torque by a separate experiment in which we measure the energy dissipated during each revolution.

► 6.7 ANGULAR MOMENTUM AND ITS CONSERVATION

Angular momentum

(A) Angular momentum of a particle

Refer to fig. 6.1 (p. 56) in which the particle m has tangential speed v . Its linear momentum is mv : its angular momentum is the *moment* of the linear momentum about the axis of rotation.

$$\begin{aligned} \text{Angular momentum } L &= (mv)r \\ &= (m\omega r)r \\ &= (mr^2)\omega \end{aligned}$$

(B) Angular momentum of a body

The **angular momentum** L of a body about an axis of rotation is defined by

$$\text{angular momentum} = \Sigma(mr^2)\omega$$

$$L = I\omega$$

where I is the moment of inertia about that axis.

Notes

(a) Angular momentum is a *pseudovector* quantity, its sense and direction for a symmetrical body being that of ω (p. 31).

(b) It may be measured in $(\text{kg m}^2) \times (1/\text{s})$
 $= \text{kg m}^2 \text{ s}^{-1} = \text{N m s}$

(c) The total angular momentum of a body (such as the Earth) may be made up by

(i) its *spin angular momentum* about an axis through its centre of mass, and

(ii) its *orbital angular momentum* about some external axis.

This is an important idea in atomic physics.

The law of conservation of angular momentum

In section 5.1 (p. 48) we discuss *linear* impulse and momentum: the same arguments can be developed for *angular* impulse and momentum. They give these results:

(a) The **angular impulse** of a variable torque is equal to $\int T dt$.

(b) The **angular impulse-momentum theorem** states

$$\int T dt = I\omega - I\omega_0$$

In words

$$\left(\begin{array}{c} \text{angular impulse} \\ \text{applied to a} \\ \text{body} \end{array} \right) = \left(\begin{array}{c} \text{change of angular} \\ \text{momentum experienced} \\ \text{by body} \end{array} \right)$$

An example is given on p. 420.

(c) **The law of conservation of angular momentum** states that if no resultant external torque acts on a system, then the total angular momentum of that (isolated) system remains constant.

This is a fundamental principle of physics.

Examples of the law's application

(a) A skater can reduce his moment of inertia by drawing in his arms. A reduction of I from I_1 to I_2 causes ω to increase from ω_1 to ω_2 where

$$I_1\omega_1 = I_2\omega_2 \quad (\text{since } T = 0)$$

Since the energy of rotation $= \frac{1}{2}I\omega^2$, his kinetic energy has increased: he does work in pulling in his arms.

(b) An orbiting planet experiences a centripetal gravitational attractive force from the Sun. As this force exerts no torque on the planet, it has constant angular momentum. This means that, as the planet moves closer to the Sun thereby decreasing its m.i. about the Sun, ω must increase.

(c) In the Earth–Moon system, tidal forces are reducing the Earth's spin angular momentum: the consequent increase in orbital angular momentum occurs as the Earth–Moon separation increases. The Earth will slow down and the Moon's orbital radius will increase until the Earth's period of rotation is equal to the Moon's orbital period.

(d) A gyroscope mounted so that it experiences no resultant external torque points in a fixed direction, and can be used as a compass. The gyroscopic compass has the advantage of pointing towards true North as it is unaffected by any local magnetic field variations.

6.8 THE COMPOUND PENDULUM

Fig. 6.6(a) shows a rigid body of arbitrary shape free to make oscillations about a horizontal axis through S.

Put $I_C = Mk^2$

then, using the *theorem of parallel axes* (p. 59)

$$I_S = Mk^2 + Mh_1^2$$

Using $T_S = I_S\alpha$

and choosing the anticlockwise sense to be positive we have

$$-Mgh_1 \sin \theta = M(k^2 + h_1^2)\ddot{\theta}$$

or, for small amplitudes of oscillation,

$$\ddot{\theta} \approx \frac{-(gh_1)}{(k^2 + h_1^2)}\theta$$

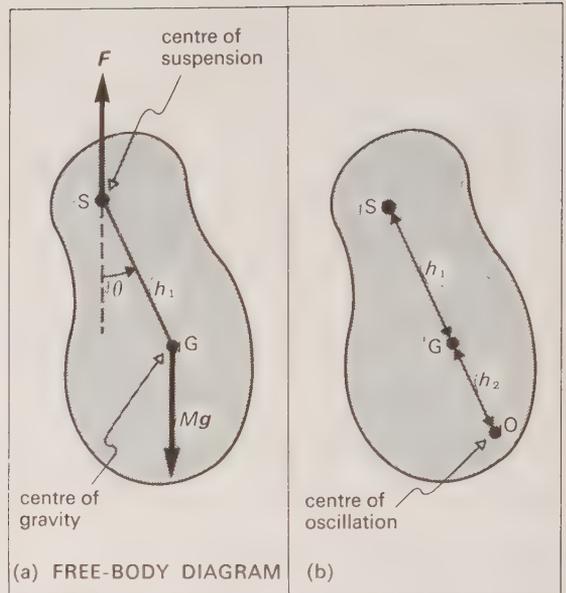


Fig. 6.6. The compound pendulum.

We conclude (p. 83) that the motion is simple harmonic for small amplitudes, and of period

$$2\pi \sqrt{\frac{k^2 + h_1^2}{h_1g}}$$

Minimum time period

If suspension from O (fig. 6.6(b)) gives the same time period as from S, then O and S are *centres of oscillation and suspension*, respectively, in fig. 6.6(a).

$$\frac{k^2 + h_2^2}{h_2} = \frac{k^2 + h_1^2}{h_1}$$

$$k^2(h_1 - h_2) = h_1h_2(h_1 - h_2)$$

We deduce either

- (a) $h_1 = h_2$, or (b) $k^2 = h_1h_2$, or
- (c) $k = h_1 = h_2$ (special case).

In general four axes through a compound pendulum (such as a metre rule) will give equal time periods. When $k = h_1 = h_2$, we have two coincident pairs, and the period is a minimum: the smallest value of

$$\begin{aligned} \frac{k^2 + h_1^2}{h_1} &= \frac{(k - h_1)^2 + 2kh_1}{h_1} \\ &= \frac{(k - h_1)^2}{h_1} + 2k \end{aligned}$$

occurs when $k = h_1$, and is $2k$.

The minimum period is $2\pi\sqrt{2k/g}$.

The simple equivalent pendulum

We have shown above that $k^2/h_1 = h_2$.

$$\therefore \frac{k^2 + h_1^2}{h_1} = \frac{k^2}{h_1} + h_1 = (h_1 + h_2)$$

Then time period = $2\pi\sqrt{[(h_1 + h_2)/g]} = 2\pi\sqrt{l/g}$

where $l = (h_1 + h_2)$ is the length of the simple equivalent pendulum: it is the distance between the centres of oscillation and suspension.

The metre rule as a compound pendulum

Suppose the time period T is measured for an oscillating metre rule for different values of h measured from the centre of gravity. Then

$$T^2 = 4\pi^2 \left(\frac{k^2 + h^2}{hg} \right)$$

Fig. 6.7 shows T^2 plotted against h .

Notes

- (a) T_{\min} occurs when $h_1 = h_2 = k$.
- (b) $T_{\max} = \infty$ occurs when $h = 0$.
- (c) The horizontal line (for a given T) may cut the graph at four points A, B, C and D.

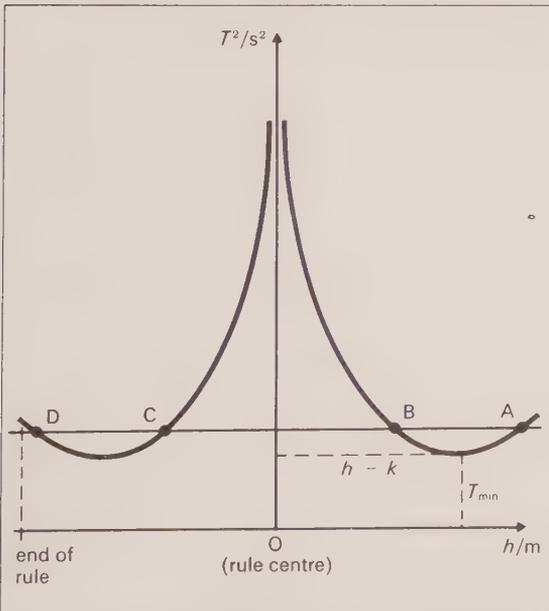


Fig. 6.7. Measurements with a metre rule as compound pendulum.

Then $AC = BD = (h_1 + h_2)$ and is the length of the simple equivalent pendulum.

(d) Rearrangement of the above equation gives

$$T^2h = \frac{4\pi^2}{g} \cdot h^2 + \frac{4\pi^2}{g} \cdot k^2$$

so that, if T^2h is plotted against h^2 , the straight-line graph would have gradient $4\pi^2/g$ and intercept on the h^2 -axis of $-k^2$. This leads to the determination of values for g and k . The reversible compound pendulum was used by Kater (1818) for an early accurate measurement of g , which avoided the plotting of graphs.

6.9 OSCILLATORY DETERMINATIONS OF MOMENT OF INERTIA

(A) The bifilar suspension

As shown in fig. 6.8, a body of mass M is suspended by two equal, vertical strings arranged symmetrically about C. If the body is displaced through a small angle in a horizontal plane about a vertical axis through C, it oscillates with s.h.m. when released. For specified values of l and d ,

$$\text{the time period, } T_0 \propto \sqrt{\frac{I_0}{M}}$$

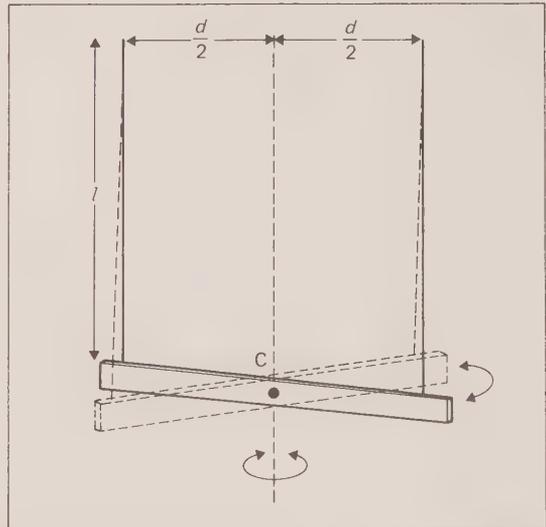


Fig. 6.8. The bifilar suspension.

where I_0 is the m.i. about the axis being considered. If two different bodies are suspended in turn by the same strings,

$$\frac{I_1}{I_2} = \frac{M_1 T_1^2}{M_2 T_2^2}$$

which enables an unknown m.i. to be calculated.

(B) The torsional pendulum

As shown in fig. 6.9, the torsional pendulum is suspended by wire or thread of torsion constant c . If the body is given an angular displacement about an axis passing through C , it will oscillate with s.h.m. when released. If I_0 is the m.i. about the axis being considered, it can be calculated from the expression for the time period,

$$T_0 = 2\pi \sqrt{\frac{I_0}{c}}$$

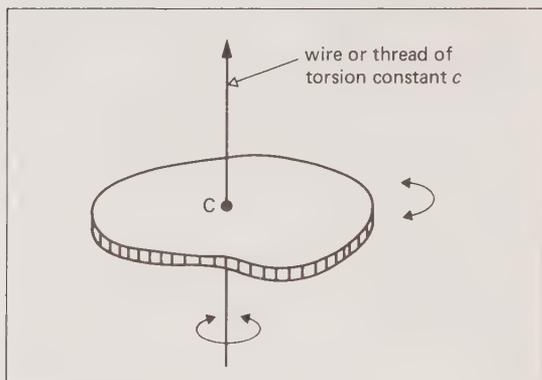


Fig. 6.9. The torsional pendulum.

Once I_0 is known, the m.i. I of another body, such as a ring, can be calculated by placing it on the pendulum with the same axis of oscillation, timing the period T of the combination and using the expression

$$\frac{T_0^2}{T^2} = \frac{I_0}{I_0 + I}$$

6.10 COMPARISON OF LINEAR AND ROTATIONAL DYNAMICS

Analogous quantities

For kinematic analogues, see p. 56.

Linear motion		Rotational motion	
dimensions	quantity	quantity	dimensions
[M]	m	$I = \Sigma mr^2$	[ML ²]
[MLT ⁻²]	F	$T = Fr$	[ML ² T ⁻²]
[ML ² T ⁻²]	$E_k = \frac{1}{2}mv^2$	$E_k = \frac{1}{2}I\omega^2$	[ML ² T ⁻²]
[MLT ⁻¹]	$p = mv$	$L = I\omega$	[ML ² T ⁻¹]

Note carefully the respective dimensions.

Analogous equations

Linear motion	Rotational motion
$F = ma$	$T = I\alpha$
$W = Fs$	$W = T\theta$
$W = \frac{1}{2}mv^2 - \frac{1}{2}mu^2$	$W = \frac{1}{2}I\omega^2 - \frac{1}{2}I\omega_0^2$
power $P = Fv$	$P = T\omega$
$Ft = mv - mu$	$Tt = I\omega - I\omega_0$

7 EQUILIBRIUM

7.1 EQUILIBRIUM OF A PARTICLE	65
7.2 EQUILIBRIUM OF A BODY	65
7.3 EQUILIBRIUM CONDITIONS	66
7.4 COUPLES	67

7.5 SYSTEMS OF FORCES	67
7.6 CENTRE OF MASS	68
7.7 CENTRE OF GRAVITY	69
7.8 STABILITY OF EQUILIBRIUM	69

7.1 EQUILIBRIUM OF A PARTICLE

A particle is in **equilibrium** if its acceleration is zero. The equilibrium is **static** if, in addition, the particle's velocity is also zero, so that it remains continuously at rest relative to the observer.

Since $F = ma$, in which F and a are vectors, if $a = 0$, then $F = 0$. This gives the **condition for the equilibrium** of a particle:

the vector sum of all external forces acting on the particle must be zero.

Coplanar forces, which we meet most frequently, are those whose lines of action are confined to one plane. Suppose this plane is the x - y plane: then

(a) To prove that a particle is in equilibrium, we must show that the sums of the resolved parts of the forces on the particle in *any two directions* are each zero. This is called a **sufficient condition**.

(b) Given that a particle is in equilibrium, we know that the sums of the resolved parts of the forces on the particle in *all directions* are each zero. This is called a **necessary condition**.

Procedure for solving problems

(Refer also to p. 39.)

- Draw a situation diagram.
- Draw a free-body diagram for the particle in equilibrium.
- Resolve *twice* in *any two convenient directions*:

$$\text{e.g. put } \Sigma F_x = 0, \quad \Sigma F_y = 0.$$

This will result in two equations, which can be used to find *two* unknowns.

Notes

(a) Since a particle is a point mass, forces acting on it must pass through a single point.

(b) If the particle moves with constant velocity, the same procedure can be followed.

(c) When the forces are not coplanar then we obtain *three* (rather than two) independent equations.

7.2 EQUILIBRIUM OF A BODY

The moment of a force, or torque T

The moment of a force is a measure of its ability to rotate a body about a given axis. Refer to fig. 7.1.

The **moment** or **torque** T of the force F about the axis through O is defined by the equation

$$T = Fr$$

$$T[\text{N m}] = F[\text{N}] r[\text{m}]$$

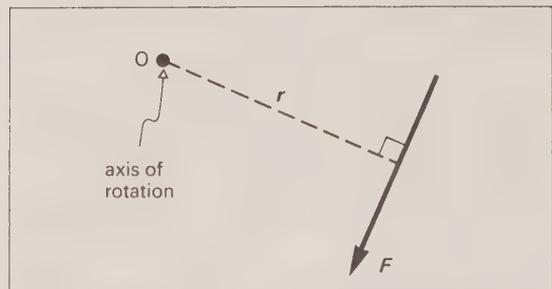


Fig. 7.1. Definition of torque.

Notes

(a) Torque has dimensions $[\text{ML}^2\text{T}^{-2}]$ (the same as energy, but this does *not* imply any direct relationship).

(b) Torque is a *pseudovector* quantity: the direction of the torque in fig. 7.1 is perpendicular to the paper, and its sense is into the paper (so as to obey a right-hand corkscrew rule).

(c) 1 N m is a unit which here has no special name. (Be careful not to confuse the unit for *torque* with that for *work*—the joule.)

(d) The length r is sometimes called the **moment arm**.

(e) For coplanar forces we speak loosely of taking *moments about a point* (rather than an axis), since the axis is represented as a point in a two-dimensional diagram.

Equilibrium of a body

(A) Translational

The centre of mass of the body must be non-accelerated. The conditions of section 7.1 apply.

(B) Rotational

A body is in rotational equilibrium if its angular acceleration α about every axis is zero.

Since $T = I\alpha$, in which T and α are vectors, if $\alpha = 0$, then $T = 0$.

The condition is that the vector sum of all the external torques acting on the body must be zero.

Suppose we confine ourselves to forces in the x - y plane: then

(a) To prove that a body is in equilibrium, we must show that the sum of the torques acting on the body about *any one axis* is zero. This is a *sufficient* condition.

(b) Given that a body is in equilibrium, we know that the sums of the torques acting on the body about *all axes* are each zero. This is a *necessary* condition.

In (a) and (b), the axes will be drawn perpendicular to the x - y plane, and therefore along the z -direction. When the forces are not coplanar, then we have *three* independent directions about which we can take moments.

The necessary condition (b) is called the **principle of moments**.

If a body is in equilibrium, the algebraic sums of the torques acting on the body about all axes are each zero.

7.3 EQUILIBRIUM CONDITIONS

Summary of necessary conditions

(A) Forces in three dimensions

$$(a) \text{ Translation} \quad \mathbf{a} = 0$$

$$\therefore \Sigma \mathbf{F} = 0$$

which implies $\Sigma F_x = \Sigma F_y = \Sigma F_z = 0$

$$(b) \text{ Rotation} \quad \mathbf{\alpha} = 0$$

$$\therefore \Sigma \mathbf{T} = 0$$

which implies $\Sigma T_x = \Sigma T_y = \Sigma T_z = 0$

(B) Forces confined to x - y plane

$$(a) \text{ Translation} \quad \Sigma \mathbf{F} = 0$$

which implies $\Sigma F_x = \Sigma F_y = 0$

$$(b) \text{ Rotation} \quad \Sigma \mathbf{T} = 0$$

which implies $\Sigma T_z = 0$

Thus in (A) we have six independent conditions, and in (B) we have three. When solving problems involving forces which are coplanar, we can only establish three independent equations, and find three unknowns: a fourth equation (found by resolving in a new direction, or taking moments about a new axis) contains no new information.

Theorems about equilibrium

(a) If three forces maintain a rigid body in equilibrium, their lines of action are concurrent. (Suppose three forces P , Q and R act on a rigid body. If R does not pass through the intersection of P and Q , there will be a resultant torque which gives the body an angular acceleration about an axis through that point.)

(b) (i) The Triangle of Forces Theorem

If three forces acting on a particle can be represented in size and direction by the sides of a triangle taken in order, they will maintain the particle in equilibrium.

(ii) Its converse

If three forces maintain a body in equilibrium, they can be represented in size and direction by the sides of a triangle taken in order.

(i) and (ii) are consequences of the parallelogram law of addition (p. 18). The converse is the more useful for solving problems.

(c) The Polygon of Forces Theorem

When more than three forces act on a particle, as in fig. 7.2(a), the force vectors must form a closed polygon ($\Sigma \mathbf{F} = 0$) if that particle is in equilibrium.

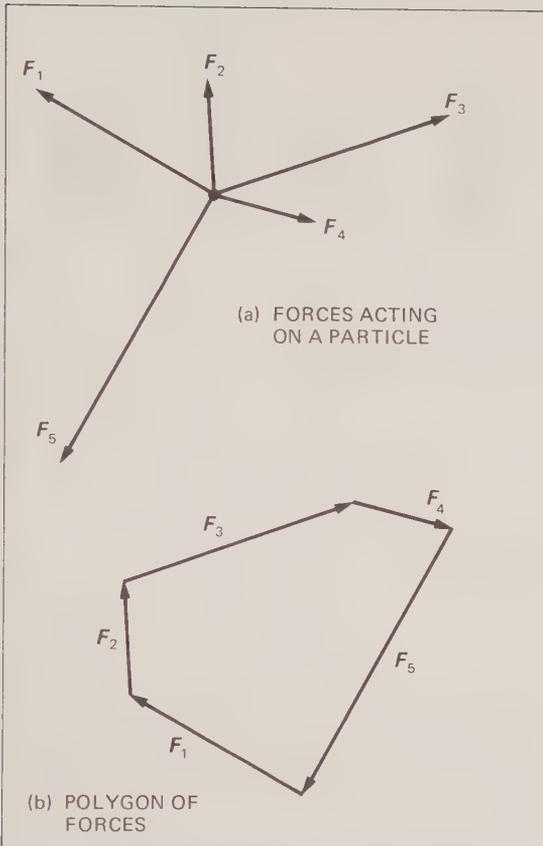


Fig. 7.2. The polygon of forces.

Note that the arrows in fig. 7.2(b) must all have the same sense round the polygon. In general, the polygon of forces will not be restricted to a plane unless all the forces involved are coplanar. A polygon of forces can be considered as a series of triangles of forces.

These theorems can be used to calculate the forces in frameworks of structures such as bridges, cranes and buildings. Beams and their trusses contain components that are in a state of tension (**ties**) and others that are in a state of compression (**struts**). The design engineer will select suitable materials, e.g. steel for tension and concrete for compression, for the various parts of the structure.

7.4 COUPLES

The resultant of two coplanar forces is usually found by vector addition using the parallelogram

law, but this method cannot be used for *parallel* forces.

(a) Like parallel forces P and Q have a resultant along the same direction of magnitude $P + Q$. Its line of action is such that it produces the same torque as P and Q , and is between their lines of action.

(b) Antiparallel forces P and Q ($P > Q$ say) have a resultant magnitude $(P - Q)$ whose line of action lies outside those of P and Q .

(c) Equal antiparallel forces cannot be replaced, for all effects, by a single force, and constitute an example of a **couple**.

A **couple** is a system of forces which produces a turning effect only (i.e. zero translational effect).

A couple often consists of a pair of equal antiparallel forces acting along a straight line, as in

(i) the system of forces applied to a car steering wheel, or

(ii) the forces experienced by two sides of a rectangular coil which carries a current while in a magnetic field (p. 406).

We cannot always identify two *distinct* forces, as in

(i) the forces applied by one car clutch-plate on the other, or

(ii) the forces exerted by a torsion fibre on the moving coil of a galvanometer.

The torque of a couple

The couple shown in fig. 7.3 (overleaf) illustrates

(a) the moment of a couple is the *same about any axis* drawn perpendicular to the plane it defines (since O is an arbitrary point), and

(b) the torque of a couple is calculated from

$$T = Fd$$

where F is the magnitude of either force, and d is the moment arm, the perpendicular distance between the forces' lines of action.

7.5 SYSTEMS OF FORCES

A system of concurrent, coplanar forces can be replaced by a single force F , its resultant. The resultant torque T about a specified axis is equal to the vector sum of the torques of the separate,

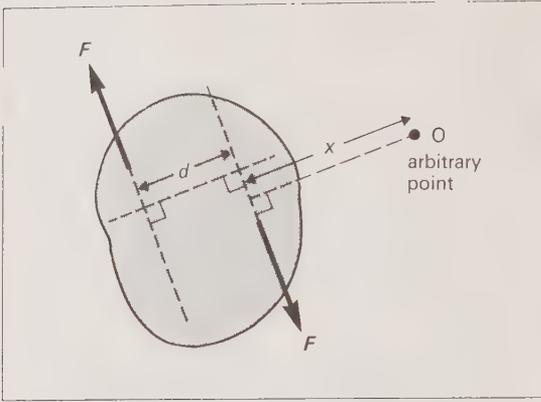


Fig. 7.3. The torque T of the forces is clockwise and of size $F(x + d) - Fx = Fd$.

concurrent forces. F determines the translational motion of a rigid body and T determines its rotational motion. When the applied coplanar forces are not concurrent, they can always be reduced to a single force and a couple. This couple simulates the net turning effect of the whole system about the specified axis. The resultant force must not, therefore, contribute any further torque and so must be constructed to pass through the point where the chosen axis meets the plane containing the coplanar forces. In this coplanar case, each force in the system has a moment about an axis (the torque axis) at right angles to the plane.

In the general case of a system of non-coplanar forces, the orientations of these torque axes are generally different for the different forces. Nevertheless, it can still be shown that such a general force system can always be reduced to a single force applied at a specified point together with a couple. This couple is a vector quantity which is the vector sum of the moments of all the individual forces in the system about the chosen point.

7.6 CENTRE OF MASS

Suppose a force is applied to the rod of fig. 7.4 at the point C , which accelerates the rod and both masses.

It is found by experiment that the rod *does not* have angular acceleration if

$$m_1 d_1 = m_2 d_2$$

C is the **centre of mass** of the two point masses. Suppose that relative to an arbitrary origin, m_1, m_2 and C have coordinates $(x_1, y_1), (x_2, y_2)$ and (\bar{x}, \bar{y}) respectively. Then

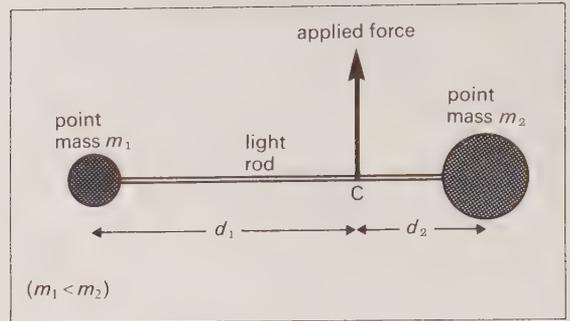


Fig. 7.4. The concept of centre of mass.

$$\bar{x} = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2}, \quad \bar{y} = \frac{m_1 y_1 + m_2 y_2}{m_1 + m_2}$$

We define the **centre of mass** \bar{x} of a number of particles by the equation

$$\bar{x} = \frac{\sum (m_i x_i)}{\sum m_i}$$

where the particle m_1 has coordinate x_1 relative to the origin.

► For a continuous distribution of matter (i.e. for a body not composed of discrete point masses) we would calculate the location of C from

$$x = \frac{\int x \, dm}{\int dm}$$

Location of the centre of mass

When a body possesses one or more symmetry elements, the location of C is simplified. C must coincide with any centre of symmetry, it must lie on any rotational axis of symmetry, and any planes of symmetry must pass through C . Some common examples include:

- (a) sphere, circular plate, cylinder – at the geometrical centre;
- (b) triangular plate – at the point of intersection of the three medians;
- (c) pyramid, cone – at $\frac{1}{4}$ of the height, measured from the base along the axis;
- (d) regular polygon – at the geometrical centre.

7.7 CENTRE OF GRAVITY

The weights of the individual particles of which a body is composed form a system of forces which are effectively parallel. Their single resultant constitutes the **weight G** of the body.

By the **centre of gravity** of a body we mean that single point through which the line of action of the weight passes however the body is oriented.

It follows that the resultant gravitational torque on the body about an axis through that point will always be zero.

Suppose the individual particles of a body have weight w_1 etc. Then the centre of gravity is located at \bar{x} , where

$$G\bar{x} = \Sigma(w_1x_1)$$

$$\bar{x} = \frac{\Sigma(w_1x_1)}{\Sigma w_1}$$

► For a continuous distribution of matter we again write

$$\bar{x} = \frac{\int x dw}{\int dw}$$

Since $w_1 = m_1g$, it follows that where g can be assumed constant, the centre of mass and centre of gravity coincide. This is true on the surface of the Earth.

Location of the centre of gravity

(a) Use obvious physical homogeneity and geometrical *symmetry*.

(b) By calculation, using

(i) for simple bodies, the principle of moments, or

(ii) in more complex situations

$$\bar{x} = \frac{\int x dw}{\int dw}$$

(which incorporates the principle of moments).

(c) By *experiment*. Suspend the body freely from two different points on it. In each case the centre of gravity must lie vertically below the point of suspension. Its location is the intersection of the lines on which it must lie.

7.8 STABILITY OF EQUILIBRIUM

Equilibrium or not?

Distinguish carefully between a body in *unstable* equilibrium, and a body which is *not* in equilibrium (accelerated). For a body to be in equilibrium when resting on a surface, the vertical line passing through its centre of gravity must pass within the (implied) boundary of contact with the surface.

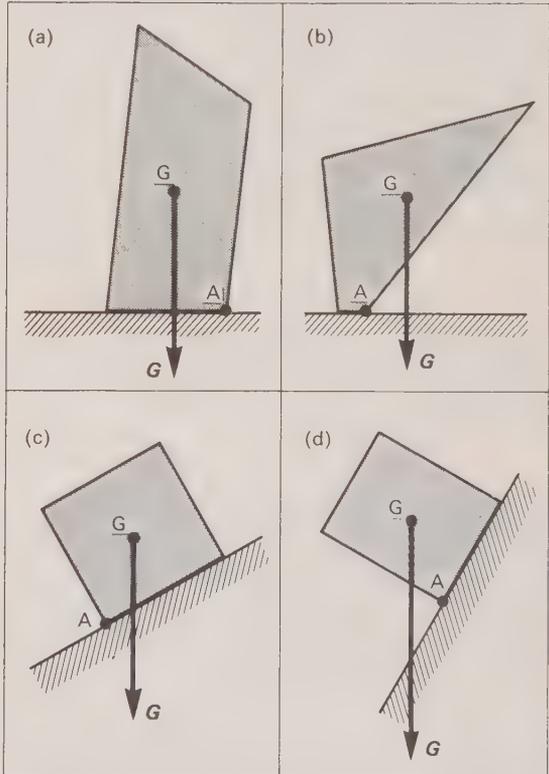


Fig. 7.5. Equilibrium, and line of action of weight.

In fig. 7.5, (a) and (c) are possible equilibrium positions, but (b) and (d) are not.

For each example, consider the torque acting on the body about A, the corner on which pivoting is possible. In (a) and (c) the gravitational torque is balanced by the torque of the contact force: the body is *in equilibrium*. In (b) and (d) the gravitational torque is unbalanced: the body is *not in equilibrium*.

Stability of equilibrium

The table overleaf shows how to distinguish **stable**, **unstable** and **neutral** equilibrium situations.

Equilibrium situations

Situation	Displacement	What happens to c.g.	What happens to p.e.	Subsequent behaviour	Name of original equilibrium
marble in saucer	displace from centre	raised	increased	marble returns to original position	stable
marble on upturned saucer	displace from centre	lowered	reduced	marble rolls off saucer	unstable
marble on flat table	displace horizontally	stays at same level	no change	marble remains in displaced location	neutral

The potential energy has a stationary value at a position of equilibrium. It is a *maximum* for unstable equilibrium, and a *minimum* for stable equilibrium.

► Analysis of equilibrium stability

The magnitudes of the changes in potential energy in gravitational and electrostatic interactions are given by $\Delta E_p = mgh$, and $\Delta W = QEy$, respectively. If F represents the force of interaction, and x the displacement vector, we write

$$\delta E_p = -F \cdot \delta x$$

The negative sign is necessary because potential increases in a direction opposite from that of field. The result

$$F = -\frac{dE_p}{dx}$$

is of general application, and is used in this book on p. 133 to analyse the interaction between a pair of molecules.

(a) For a body to be in *equilibrium* $F = 0$, so $dE_p/dx = 0$. The E_p-x curve must show a *stationary value*.

(b) For the equilibrium to be *stable* a small displacement must lead to an increase in E_p . The

E_p-x curve must be at a *minimum* value. d^2E_p/dx^2 must be positive.

(c) For the equilibrium to be *unstable* the curve must show a *maximum* value. d^2E_p/dx^2 must be negative.

(d) When $d^2E_p/dx^2 = 0$ the equilibrium is apparently neutral, but it is wise to investigate the situation more fully.

These points are illustrated in fig. 7.6. The graph can be visualized simply in terms of a ball bearing rolling along a curved track in a vertical plane. Then E_p is proportional to height above an arbitrary level.

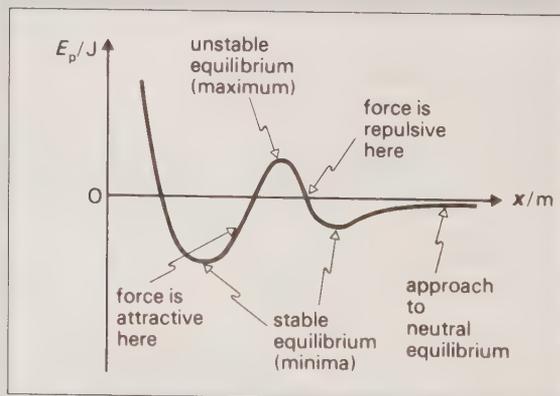


Fig. 7.6. A schematic graph to illustrate stability of equilibrium.

8 FLUIDS AT REST

8.1 DENSITY AND PRESSURE	71
8.2 FACTS ABOUT FLUID PRESSURE	72

8.3 ARCHIMEDES' PRINCIPLE AND FLOTATION	72
8.4 MEASUREMENT OF DENSITY AND PRESSURE	73

This chapter summarizes concisely what is taught as **hydrostatics** in GCSE courses.

8.1 DENSITY AND PRESSURE

Density

Suppose the sample of a particular substance has a mass m and volume V .

The **average density** ρ of the substance is defined by the equation

$$\rho = \frac{m}{V}$$

$$\rho \left[\frac{\text{kg}}{\text{m}^3} \right] = \frac{m[\text{kg}]}{V[\text{m}^3]}$$

The table indicates the extent to which densities can vary.

Material	$\rho/\text{kg m}^{-3}$
interstellar space	10^{-20}
best laboratory vacuum	10^{-16}
air	1
water	10^3
dense metal	10^4
atomic nucleus	10^{17}

The **relative density** of a substance is defined by

$$\text{relative density} = \frac{\text{density of substance}}{\text{density of water}}$$

For example, the density of mercury is $1.36 \times 10^4 \text{ kg m}^{-3}$, and its relative density is 13.6.

Pressure

Suppose an area ΔA totally immersed in a fluid experiences a normal force ΔF .

The **average pressure** p_{av} over the area is defined by the equation

$$p_{\text{av}} = \frac{\Delta F}{\Delta A}$$

$$p_{\text{av}} \left[\frac{\text{N}}{\text{m}^2} \right] = \frac{\Delta F[\text{N}]}{\Delta A[\text{m}^2]}$$

The **pressure** p at a point in the fluid is defined by the equation

$$p = \lim_{\delta A \rightarrow 0} \frac{\delta F}{\delta A}$$

where the area δA contains the point.

Notes

(a) The unit N m^{-2} is, for convenience, called a **pascal** (Pa). Thus

$$1 \text{ Pa} \equiv 1 \text{ N m}^{-2}$$

(b) Pressure is a *scalar* quantity. ΔF is a *vector*, and ΔA is a *pseudovector*, the direction of ΔA being that of the normal to the surface. The *force* that we associate with a pressure has a direction, but the *pressure* does not.

(c) It is found experimentally that if the orientation of ΔA is altered, then the magnitude of ΔF (and hence the value of p) remains fixed, even though ΔF remains normal to ΔA .

The origin of pressure forces

(a) *The pressure caused by a gas* is discussed fully on p. 199.

(b) *In a liquid* molecules have mainly vibrational energy (kinetic and intermolecular potential energy), and a small proportion of translational energy. When a liquid molecule strikes a surface, it will suffer a much smaller momentum change than would an analogous gas molecule because its speed is (on average) much smaller at the instant of impact. Nevertheless there may be 10^3 times as many molecules in unit volume, and the pressures in the gas and liquid may have the same order of magnitude.

(c) *In a solid* the mechanism is similar, but differs in that the solid's molecules have only vibratory (and no translatory) kinetic energy. When vibration causes overlap between the electron clouds of adjacent molecules, they will repel each other (p. 127).

8.2 FACTS ABOUT FLUID PRESSURE

(a) *Pressure increase with depth*
Imagine two points separated by a vertical distance h in a fluid of constant density ρ . The pressure difference Δp between the points is given by

$$\Delta p = h\rho g$$

where h is measured as positive downward.

$$\Delta p \left[\frac{\text{N}}{\text{m}^2} \right] = h[\text{m}] \rho \left[\frac{\text{kg}}{\text{m}^3} \right] g \left[\frac{\text{m}}{\text{s}^2} \right]$$

(b) It follows from (a) that *the pressures at two points at the same horizontal level in the same liquid are the same*, provided that the points are connected by that liquid only, and that the liquid is static. (If this were not so, fluid would move from the place of higher pressure to that of lower until the condition was satisfied.)

(c) The change of pressure Δp depends on the vertical distance h , ρ and g only: it is therefore independent of the cross-sectional area and shape of any containing vessel.

(d) *For a gas* (whose density is very low) Δp is small unless h is large. *The gas pressure is the same at all points in volumes of reasonable size.*

► For volumes of large size (such as the Earth's atmosphere) the gas density ρ will not be constant because the lower layers are compressed. One cannot apply $\Delta p = h\rho g$ under these conditions: one must use the more rigorous form

$$\frac{dp}{dh} = -\rho g$$

in which h is measured vertically upward, and the negative sign is inserted because p decreases as h increases. ρ depends upon p .

(e) **Pascal's Principle** states that pressure applied to an enclosed liquid is transmitted without change to every part of the liquid, whatever the shape of the liquid.

If the pressure at a liquid surface is p_0 (e.g. atmospheric pressure), then the pressure at a point in the liquid can be found from

$$p = p_0 + \Delta p$$

where Δp is caused by changes of vertical height in the liquid.

8.3 ARCHIMEDES' PRINCIPLE AND FLOTATION

Archimedes's Principle states that when a body is immersed in a fluid (liquid or gas) it experiences an upthrust equal in magnitude to the weight of fluid displaced.

We can justify the principle by experimental verification, a pressure argument, or a force argument.

Force argument

Consider the equilibrium of that part of the fluid which has the same volume and shape as the part of the body to be immersed. It experiences a force, through its centre of gravity, of the same magnitude as its weight, but oppositely directed. When that part of the fluid is replaced by the body to be immersed, then the body will experience the same pressure forces. This proof, although not analytical, is rigorous.

Notes

(a) The upthrust still acts when the body rests on some support (such as the bottom of a vessel) provided the space between the body and the support is not evacuated.

(b) *Newton's Third Law* requires that a body which experiences a buoyant upthrust should exert an equal but downward thrust on the fluid responsible.

Flotation

The Principle of Flotation states that a floating body displaces its own weight of fluid.

It is a consequence of the condition for equilibrium and *Archimedes' Principle*. It is applied in situations with

- (a) *variable weight and constant upthrust* (such as the submarine), and
- (b) *constant weight and variable upthrust*, such as
 - (i) the balloon, and
 - (ii) the **hydrometer**. When it floats freely, the hydrometer experiences an upthrust equal in magnitude to its weight (a constant): it therefore reaches equilibrium with a particular volume under the surface depending on the density of the liquid in which it is immersed.

8.4 MEASUREMENT OF DENSITY AND PRESSURE

(A) Density and relative density

- (a) Direct measurement of m and V .
- (b) Measurement of V by displacement methods.
- (c) Use of the *relative density bottle* for liquids, or finely divided solids.
- (d) Use of *Archimedes' Principle*:
 - (i) *Solid*. If the sample floats use a sinker. If it dissolves in water use some other non-solvent liquid.
 - (ii) *Liquid*. Use a solid (such as a glass stopper) of constant volume.

(B) Pressure

Absolute pressure is the actual pressure at a point in a fluid.

Gauge pressure is that recorded by a pressure gauge, and is frequently the difference between absolute pressure and atmospheric pressure.

The equation $\Delta p = h\rho g$ indicates why it is convenient to refer to pressures by **heads of liquid**.

A unit commonly used for gas pressures is the **atmosphere** (atm), which is *defined* to be 101325 Pa. It is equivalent (but not *exactly* equal) to that exerted by 760 mm of mercury (mmHg) of specified density under **standard gravity** ($g_0 = 9.80665 \text{ N kg}^{-1}$). Thus

$$\begin{aligned} 1 \text{ atmosphere} &= 760 \text{ mmHg} \\ &= (0.760 \text{ m}) \times (1.36 \times 10^4 \text{ kg m}^{-3}) \\ &\quad \times (9.81 \text{ N kg}^{-1}) \\ &= 1.013 \times 10^5 \text{ N m}^{-2} \text{ (Pa)}. \end{aligned}$$

[Note that the conventional millimetre of mercury exerts a pressure of 133 Pa. Neither the atm nor the mmHg is an SI unit.]

Atmospheric pressure can be measured by

- (a) a **simple barometer** (normally using mercury),
- (b) a **Fortin** barometer, which incorporates refinements such as a *vernier scale*,
- (c) an **aneroid** barometer, which consists of an evacuated flexible metal chamber.

The simple barometer is an example of a closed-tube **manometer**.

Other pressure-measuring devices include

- (a) the **U-tube** manometer,
- (b) the **Bourdon** gauge,
- (c) the **McLeod** gauge (a device for measuring very low pressures).

9 THE BERNOULLI EQUATION

9.1 DESCRIPTION OF STREAMLINE FLOW	74	9.4 APPLICATIONS	76
9.2 THE EQUATION OF CONTINUITY	74	9.5 FIELDS OF FLOW	78
9.3 THE BERNOULLI EQUATION	75		

9.1 DESCRIPTION OF STREAMLINE FLOW

In general, fluid dynamics is too complex to analyse, but we can proceed if we make the following assumptions.

(a) *The fluid is non-viscous* (see p. 161). No mechanical energy is converted to internal (thermal) energy.

(b) *The fluid is incompressible*. This means that its density remains constant.

(c) *The fluid angular momentum plays no part*. Flow patterns for which this is true are said to be **irrotational**.

(d) *The flow is steady*. Flow is said to be *laminar*, *orderly* or **streamline** if successive particles while passing a given point have the same velocity. (Flow is **turbulent** or *disorderly* if the velocity of particles at the given point depends on the instant of observation.)

Although *irrotational steady* flow of a *non-viscous incompressible* fluid does not occur in practice, the analysis which follows is a reasonable description of some real physical situations. (See also the equation that describes an adiabatic change, p. 209.)

For *steady flow*, we define terms as follows.

(i) A **streamline** is the path taken by a fluid particle. The instantaneous velocity of a fluid particle at a point lies along the tangent to the streamline at that point. Like electric field lines (p. 340), streamlines can never cross.

(ii) A **tube of flow** is an imaginary boundary defined by streamlines drawn so as to enclose a tubular region of fluid. No fluid crosses the *side* boundary of such a tube (see fig. 9.1).

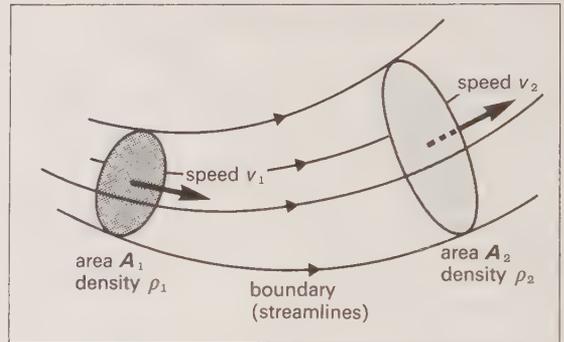


Fig. 9.1. A tube of flow.

9.2 THE EQUATION OF CONTINUITY

This equation follows from an application of the law of conservation of *mass* to ideal fluid flow.

In time Δt , a mass m_1 enters the tube of flow (fig. 9.1) across the cross-section A_1 , where $m_1 = \rho_1 A_1 v_1 \Delta t$. A mass m_2 leaves at A_2 in the same time, where $m_2 = \rho_2 A_2 v_2 \Delta t$.

Equating $m_1 = m_2$

$$\rho_1 A_1 v_1 = \rho_2 A_2 v_2 \quad \text{The equation of continuity}$$

Since we are to consider incompressible fluids only, $\rho_1 = \rho_2$, so

$$A_1 v_1 = A_2 v_2$$

Av is called the *flow rate*, or **volume flux**. The fact that it is constant demonstrates that v is large where A is small, and vice versa.

9.3 THE BERNOULLI EQUATION

This equation follows in a similar way from the law of conservation of *energy*.

(A) Change of speed (fig. 9.2)

Consider a time interval Δt .

(a) The net work done *on* the fluid during this time by the pressure forces due to the fluid to the left of point A and to the right of point B is given by

$$\begin{aligned} W &= \left(\text{work done by forces at A} \right) + \left(\text{work done by forces at B} \right) \\ &= (p_1 A_1)(v_1 \Delta t) - (p_2 A_2)(v_2 \Delta t) \\ &= (p_1 - p_2)(A_1 v \Delta t) \end{aligned}$$

since $A_1 v_1 = A_2 v_2$. Note that the forces at B do *negative* work on the fluid, because they oppose the fluid's motion.

(b) The gain of kinetic energy is that of a mass

$$(A_1 v_1 \Delta t) \rho \quad [\text{or } (A_2 v_2 \Delta t) \rho]$$

whose speed changes from v_1 to v_2 .

$$\text{Gain of k.e.} = \frac{1}{2}(A_1 v_1 \Delta t \rho)(v_2^2 - v_1^2)$$

We now apply the *work-energy theorem*

$$W = \text{change of k.e.}$$

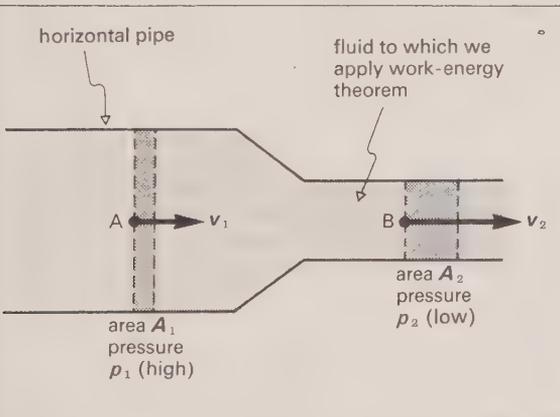


Fig. 9.2. Change of pressure that accompanies change of speed.

$$\begin{aligned} (p_1 - p_2)(A_1 v_1 \Delta t) &= \frac{1}{2}(A_1 v_1 \Delta t \rho)(v_2^2 - v_1^2) \\ (p_1 - p_2) &= \frac{1}{2} \rho (v_2^2 - v_1^2) \\ p_1 + \frac{1}{2} \rho v_1^2 &= p_2 + \frac{1}{2} \rho v_2^2 \end{aligned} \quad (1)$$

(B) Change of vertical height (fig. 9.3)

Consider a time interval Δt . Because there is no change of kinetic energy, the net work done *on* the system is *zero*. The (positive) work done by the pressure forces has the same magnitude as the (negative) work done by the gravitational forces:

$$(\text{work done by pressure forces}) = (\text{gain of p.e.})$$

$$\begin{aligned} (p_1 - p_2)(Av \Delta t) &= (Av \Delta t \rho)g(h_2 - h_1) \\ (p_1 - p_2) &= \rho g(h_2 - h_1) \\ p_1 + h_1 \rho g &= p_2 + h_2 \rho g \end{aligned} \quad (2)$$

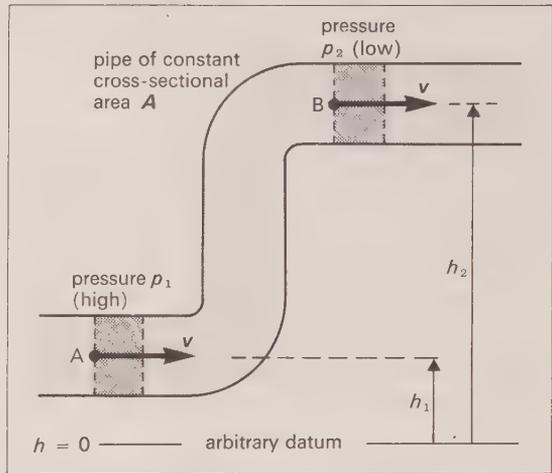


Fig. 9.3. Change of pressure that accompanies change of vertical height.

(C) Change of speed and height

If (A) and (B) happen simultaneously, so that both kinetic energy and gravitational potential energy change, then (1) and (2) are combined as

$$p_1 + \frac{1}{2} \rho v_1^2 + h_1 \rho g = p_2 + \frac{1}{2} \rho v_2^2 + h_2 \rho g$$

More generally, since A and B are arbitrary points, we have

$$p + \frac{1}{2} \rho v^2 + h \rho g = \text{constant}$$

which is the **Bernoulli equation**, and which applies to all points along a streamline for steady, non-viscous incompressible flow. When the flow is irrotational, the constant has the same value for all streamlines.

Notes

(a) Each of the terms has the unit N m^{-2} (Pa) and dimensions $[\text{ML}^{-1}\text{T}^{-2}]$ of pressure, and can be regarded as energy per unit volume.

(b) Suppose $v_1 = v_2 = 0$, so that the fluid is at rest. Then

$$(p_1 - p_2) = -(h_1 - h_2)\rho g \text{ or } \Delta p = h\rho g$$

This is the familiar result of p. 72.

(c) It is sometimes helpful to write the *Bernoulli* equation in the form

$$p = C - \frac{1}{2}\rho v^2 - h\rho g \text{ where } C = \text{constant.}$$

When $v = 0$, the *static pressure* can be written as

$$p_{v=0} = C - h\rho g$$

As h is taken to be positive in an *upward* direction, it is clear that the static pressure increases with increasing depth. C is the value of the pressure when $v = 0$ and $h = 0$. The term $-\frac{1}{2}\rho v^2$ is known as the *dynamic component* of the pressure.

9.4 APPLICATIONS

In all real situations, the conditions for the *Bernoulli* equation to be exact are violated. (We can seldom escape turbulence; nearly every fluid is viscous, and gases are easily compressed.) Nevertheless the qualitative argument holds: where streamlines are drawn together, as at a constriction, the speed of the fluid increases (equation of continuity). If this happens at a horizontal level, then the pressure of the fluid falls (*Bernoulli*). This general result has many applications.

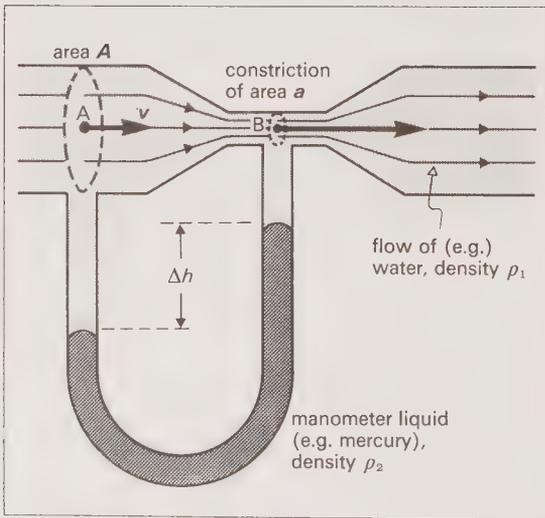


Fig. 9.4 The Venturi meter.

(A) The Venturi meter

This device (fig. 9.4) measures the rate of flow of fluid (its *volume flux*) through a pipe.

The pressure difference between A and B is recorded by the attached manometer. By measuring A , a , ρ_1 , ρ_2 and Δh , one can find

(a) the fluid speed v at A (by using the *Bernoulli* equation), and hence

(b) the volume flux = Av (by using the equation of continuity).

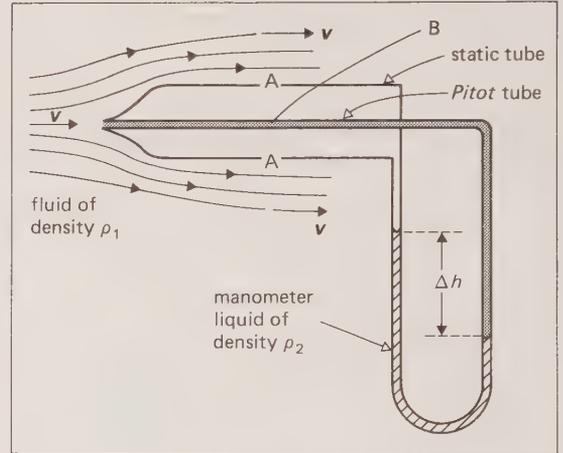


Fig. 9.5. The Pitot static tube.

(B) The Pitot static tube

This device (fig. 9.5) measures the *speed* of a fluid, e.g. air flowing past an aircraft. The fluid cannot flow through the tube and its velocity is reduced to zero at B. The pressure at B in the *Pitot* tube is the full *ram* (or *stagnation*) pressure, p_B , whereas the pressure at A in the static tube is the *free-stream* pressure, p_A . Application of the *Bernoulli* equation to points A and B gives

$$p_B = p_A + \frac{1}{2}\rho_1 v^2 \tag{1}$$

The difference in height, Δh , of the liquid in the manometer arms gives a second equation linking the pressures in the two tubes,

$$p_A = p_B - \Delta h \rho_2 g \tag{2}$$

Equations (1) and (2) give

$$\frac{1}{2}\rho_1 v^2 = \Delta h \rho_2 g$$

so that
$$v = \sqrt{\frac{2\Delta h \rho_2 g}{\rho_1}}$$

which gives the speed of the fluid.

(C) Dynamic lift

This force is exerted on a body which is moving relative to a fluid in such a way that the streamlines are close on one side and further apart on the other—the pattern is not symmetrical.

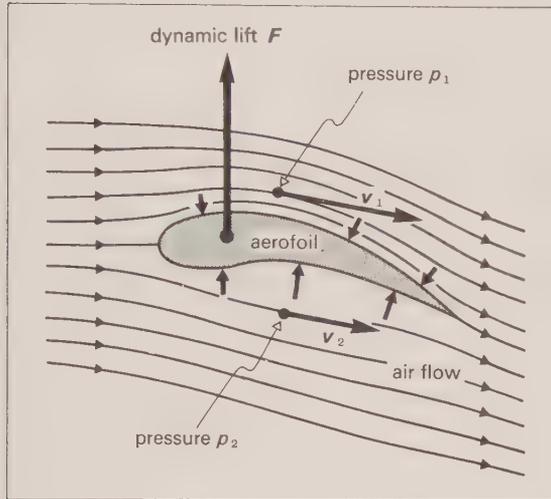


Fig. 9.6. An aerofoil experiences dynamic lift.

The aerofoil

The aerofoil of fig. 9.6 is shaped so that air flows faster over the top than underneath.

Since $v_1 > v_2$, $p_1 < p_2$ (Bernoulli).

The fact that the pressure below an aeroplane wing exceeds that above accounts for most of the force which supports the plane. (Note that fig. 9.6 is drawn for an observer travelling with the aerofoil.) When the angle of attack is too great the flow over the upper surface becomes turbulent, which reduces the pressure difference, and leads to stalling.

The design and orientation of curved surfaces in an aeroplane determines its aerodynamic stability. *Pitching* (about a lateral axis), *yawing* (about a vertical axis) and *rolling* (about a longitudinal axis) all result from rotational torques which must be counteracted. In normal horizontal flight, a tail-plane which is curved symmetrically about a horizontal plane produces no resultant lift. A pitching movement would lead to the tail-plane having an angle of attack which results in a restoring torque. Similarly, a tail-fin which is curved symmetrically about a vertical plane produces no resultant sideways force. A yawing movement would introduce an angle of attack and a corresponding restoring torque. The wings are set at a dihedral angle so

that a rolling movement generates different resultant lifting forces on the wings, thereby producing a correcting torque. The effects of these curved surfaces can be controlled by *elevators* on the tail-plane (lateral axis), a *rudder* on the tail-fin (vertical axis) and *ailerons* on the wings (longitudinal axis).

The spinning ball

This time the dynamic force on the ball (fig. 9.7) is caused, not by its shape, but because it spins.

The resulting lateral forces causes the ball to have a curved trajectory, such as that of a sliced golf ball.

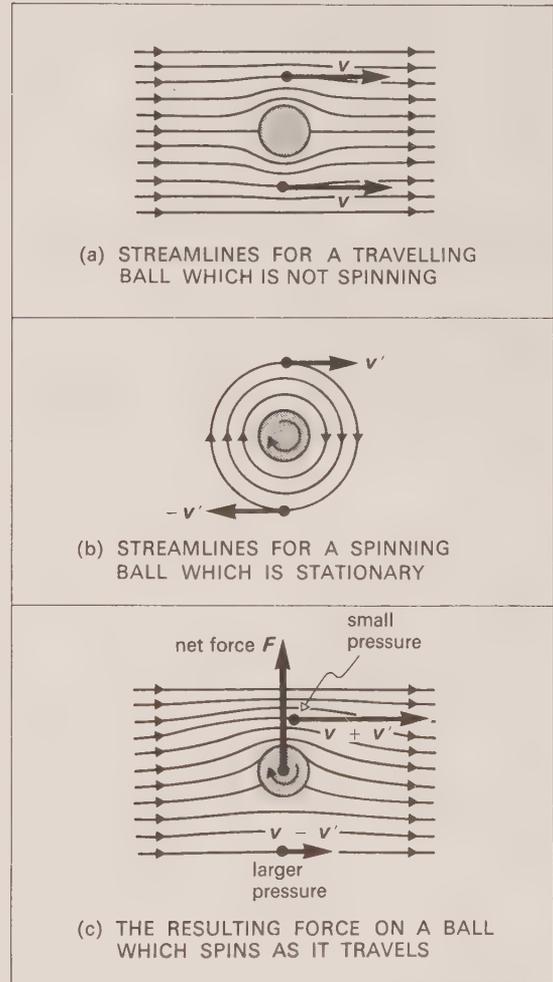


Fig. 9.7. The lateral force exerted on a moving spinning ball. The streamlines are drawn for an observer travelling with the ball.

Note that a *viscous* medium is needed: in fig. 9.7(b) air is dragged round by the ball.

(C) Torricelli's theorem

Assume that the fluid of fig. 9.8 is non-viscous, and apply the *Bernoulli* equation to the points A and B on the streamline shown.

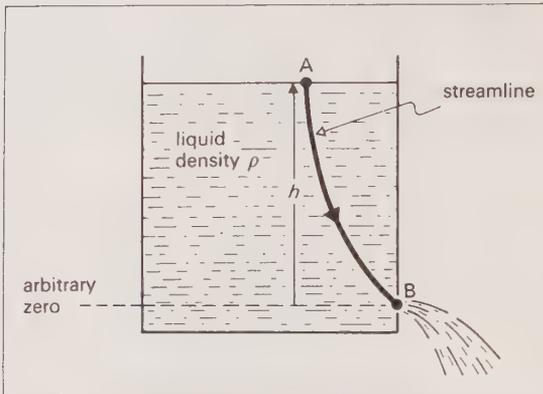


Fig. 9.8. Liquid escapes from a hole punched in a container.

At A $p = \text{atmospheric}, v = 0.$

At B $p = \text{atmospheric}, v$ is to be found.

Then $p + \frac{1}{2}\rho v^2 + h\rho g = \text{constant}$ becomes

$$(A) \quad p + 0 + h\rho g = p + \frac{1}{2}\rho v^2 + 0 \quad (B)$$

$$v = \sqrt{2gh}$$

The horizontal speed of efflux equals the vertical speed acquired by free fall from a height h.

The *Bernoulli* equation also finds application in the action of filter-pumps, bunsen burners, carburettors, boomerangs, atomizers, sails, propellers, etc.

9.5 FIELDS OF FLOW

A moving fluid can be described by a **field of flow**. Each point in the fluid has a flow velocity vector v associated with it, and this flow velocity can be related to a velocity potential. This can be compared with gravitational field strength and gravitational potential (chapter 20). Analogies can also be made between fluid fields and electromagnetic fields, as illustrated in the following examples:

(a) a *homogeneous* field of flow, with parallel streamlines, corresponds to the electric field of a parallel-plate capacitor or to the magnetic field inside a solenoid,

(b) a *source* or *sink* field of flow, with radial streamlines, corresponds to the electric field of a positive or negative charge,

(c) a *dipole* field of flow, consisting of a separated source and sink, corresponds to the electric field of two opposite separated charges,

(d) a *vortex* field of flow, with circular streamlines, represents the magnetic field near a straight, current-carrying wire.

SUMMARY OF MECHANICS

	Defining equations	Laws	Important derivations
v a ω	$v = \frac{ds}{dt}$ $a = \frac{dv}{dt}$ $\omega = \frac{d\theta}{dt}$		$v = u + at$ $s = \frac{u + v}{2} \cdot t$ $T = \frac{2\pi}{\omega}$ $v = \omega r$ $a = v^2/r = \omega^2 r = v\omega$
p F mass ratio	$p = mv$ (i) defined by <i>Newton I</i> (ii) measured from <i>Newton II</i> $\frac{m_0}{m} = \frac{a}{a_0}$	NEWTON I If $F = 0, \Delta v = 0$ II $F \propto \frac{d}{dt}(mv)$ III $F_{AB} = -F_{BA}$	$F = ma$ $G = mg$
W E_k P	$W = Fs \cos \theta$ $E_k = \frac{1}{2}mv^2$ $P = \frac{dW}{dt}$	Law of conservation of energy	$W = Fs$ $= \frac{1}{2}mv^2 - \frac{1}{2}mu^2$
	impulse = $F \Delta t$	Law of conservation of momentum: if $F = 0, \Delta p = 0$ for an isolated system.	$F \Delta t = mv - mu$
α T I	$\alpha = \frac{d\omega}{dt}$ $T = Fr$ $I = \Sigma mr^2$		$a_t = r\alpha$ $a_r = \frac{v^2}{r} = \omega^2 r = v\omega$ $T = I\alpha$ (from <i>Newton II</i>) $I = I_0 + Mh^2$ $\Delta W = T \Delta \theta$ $E_{k,rot} = \frac{1}{2}I\omega^2$

Summary of mechanics—cont.

	Defining equations	Laws	Important derivations
L	$L = I\omega$	Law of conservation of angular momentum: if $T = 0$, $\Delta L = 0$ for an isolated system.	$W = \frac{1}{2}I\omega^2 - \frac{1}{2}I\omega_0^2$ $T\Delta t = I\omega - I\omega_0$
\bar{x} \bar{x}	$\bar{x} = \frac{\int x dm}{\int dm}$ $\bar{x} = \frac{\int x dw}{\int dw}$		Equilibrium conditions $\Sigma F = 0$ $\Sigma T = 0$
ρ p	$\rho = \frac{m}{V}$ $p = \lim_{\delta A \rightarrow 0} \left(\frac{\delta F}{\delta A} \right)$		$\Delta p = h\rho g$
			Equation of continuity $\rho_1 A_1 v_1 = \rho_2 A_2 v_2$ $A_1 v_1 = A_2 v_2$ Bernoulli $p + \frac{1}{2}\rho v^2 + h\rho g = \text{constant}$

III OSCILLATIONS AND WAVE MOTION

10	SIMPLE HARMONIC OSCILLATION	82
11	DAMPED AND FORCED OSCILLATION	91
12	WAVE MOTION	96
13	HUYGENS' CONSTRUCTION	106
14	THE PRINCIPLE OF SUPERPOSITION	109
15	DOPPLER EFFECT	119

10 SIMPLE HARMONIC OSCILLATION

10.1	INTRODUCTION	82
10.2	THE KINEMATICS OF S.H.M.	83
10.3	THE DYNAMICS OF S.H.M.	85
10.4	EXAMPLES OF EXACT S.H.M.	86

10.5	EXAMPLES OF APPROXIMATE S.H.M.	87
10.6	ENERGY IN S.H.M.	88
▶ 10.7	ENERGY APPROACH TO S.H.M.	89
▶ 10.8	SUPERPOSITION OF TWO S.H.M.	90

10.1 INTRODUCTION

A **periodic** or **cyclic** motion is one during which a body continually retraces its path at equal time intervals.

Examples

- (a) Planetary motion,
- (b) the motion of an air molecule disturbed by a sound wave (this being superimposed on its random motion).

Oscillations

Examples

Nature of oscillating system	p.e. stored as	k.e. possessed by moving
mass on helical spring	elastic energy of spring	mass
cantilever	elastic energy of bent rod	rod
simple pendulum	gravitational p.e. of bob	bob
vertical rod floating in liquid of zero viscosity	gravitational p.e. of rod or liquid	rod

Before a system can be set into oscillation, it must satisfy two conditions:

- (1) it must be able to store *potential energy* (so a mechanical system would need some springiness or elasticity), and
- (2) it must have some inertia (or mass) which enables it to possess *kinetic energy*.

These two concepts (*elasticity* and *inertia*) are very useful for developing an intuitive understanding of any oscillation.

An oscillation represents the continual interchange of potential and kinetic energy.

Simple harmonic oscillations

Dimensional argument

Problem: for what type of oscillation is the time period independent of the amplitude?

Put T = time period,

F = restoring force at displacement x ,

m = mass that oscillates.

Suppose

$$T = k(F)^a(x)^b(m)^c$$

Equating dimensions for M , L and T (as on p. 20), we find

$$a = -\frac{1}{2}$$

$$b = \frac{1}{2}$$

$$c = \frac{1}{2}$$

so

$$T = k \left(\frac{mx}{F} \right)^{1/2}$$

The motion is **isochronous** (has T independent of amplitude) provided x/F is constant, since m is constant.

Then $x \propto F$

Definition of s.h.m.

(a) **Linear.** If the acceleration \ddot{x} of a point is always directed toward, and is proportional to its distance from, a fixed point, then the motion is simple harmonic (fig. 10.1).

$$\ddot{x} = -(\text{const})x$$

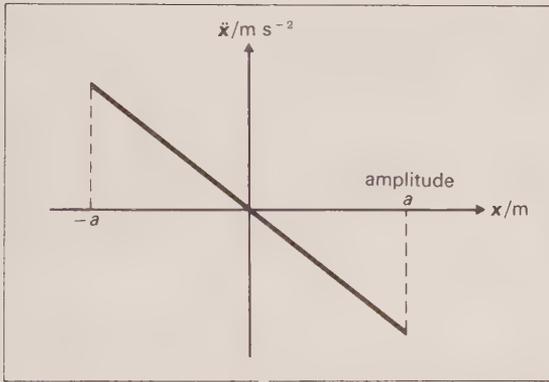


Fig. 10.1 Acceleration-displacement graph for simple harmonic motion.

(b) **Angular.** Suppose θ represents the angular displacement. Then

$$\ddot{\theta} = -(\text{const})\theta$$

The importance of s.h.m.

- (a) It occurs very frequently (see below).
- (b) T does *not* depend on amplitude.
- (c) It can be described by relatively simple mathematics.
- (d) Any *periodic* motion which is not simple harmonic can be analysed into its simple harmonic components: it can then be regarded as the superposition of the separate components, which *can* be analysed.

s.h.m. in practice

s.h.m. will be shown by any system subjected to a force which obeys *Hooke's Law*, since if

$$F = -kx \quad (\text{p. 139})$$

then $\ddot{x} = -\omega^2 x$

where ω^2 is some positive constant, and ω is called the **pulsatance**.

In practice *Hooke's Law* is nearly obeyed for most *small* distortions (see p. 139): hence many particles and systems will show at least approximate s.h.m. for *small* oscillations about a point of stable equilibrium.

The value of T will increase if

- (a) the inertia factor increases, or
- (b) the elasticity factor decreases.

Note that in some systems (such as the simple pendulum) the two may be related, since mass and gravitational force are proportional.

10.2 THE KINEMATICS OF S.H.M.

Definitions

Refer to fig. 10.2 in which P describes s.h.m.

Displacement x is the vector \vec{OP} : it takes negative values when P lies between O and B.

Amplitude a is the magnitude of the maximum displacement from the central position, i.e.

$$a = OA = \frac{AB}{2}$$

(It is always positive.)

Period T is the time taken to describe a complete oscillation (cycle), such as the path OAOBO.

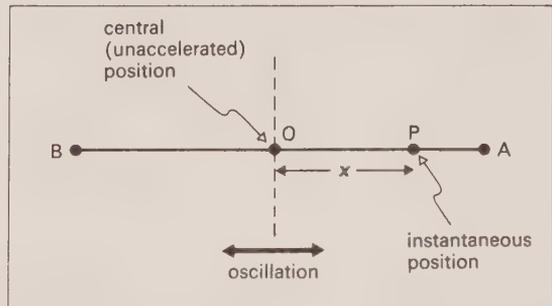


Fig. 10.2. Description of linear s.h.m.

Frequency f is the number of oscillations performed in each second. Thus

$$f = \frac{1}{T}$$

$$f[\text{Hz}] = \frac{1}{T[\text{s}]}$$

The unit *cycle per second* is called the **hertz** (Hz).

The auxiliary circle

Suppose (fig. 10.3) that N describes a circle of radius a at steady angular speed ω . We will show that P, under these conditions, describes s.h.m. according to our earlier definition.

Acceleration of N = $\omega^2 a$ toward O.

Resolved part of this acceleration parallel to AB

$$= \omega^2 a \cos \theta \text{ toward O.}$$

Acceleration of P = $\omega^2 a \cos \theta$ toward O.

$$\ddot{x} = -\omega^2 (a \cos \theta)$$

$$\therefore \ddot{x} = -\omega^2 x$$

where \ddot{x} is measured in the same direction as x .
 Since ω^2 is a positive constant, P describes s.h.m.

The *time period* T of P is the time period for one revolution of N.

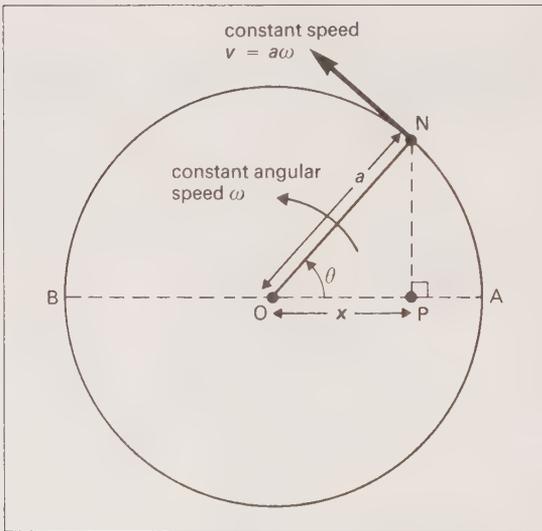


Fig. 10.3. The auxiliary circle.

$$T[\text{s}] = \frac{\text{angle through which N turns [rad]}}{\text{angular speed } \left[\frac{\text{rad}}{\text{s}} \right]}$$

$$T = \frac{2\pi}{\omega}$$

which gives $\omega = 2\pi f$.

Variation of x, \dot{x} and \ddot{x} with time t

(A) Displacement x

From fig. 10.3, $x = a \cos \theta$
 $= a \cos \omega t \quad (\theta = \omega t)$

(B) Velocity \dot{x}

Velocity of P in direction OA is that of the resolved part of N's velocity in that direction.

Thus $\dot{x} = (a\omega) \sin \theta$ along AO
 $= -a\omega \sin \theta$ along OA
 $= -a\omega \sin \omega t$
 $= \pm a\omega \sqrt{(1 - \cos^2 \omega t)} \quad (\text{p. 22})$
 $= \pm a\omega \sqrt{\left[1 - \left(\frac{x}{a} \right)^2 \right]}$
 $= \pm \omega \sqrt{(a^2 - x^2)}$

Special cases:

When $x = 0, \quad \dot{x} = \pm a\omega$
 (its largest value)

When $x = \pm a, \quad \dot{x} = 0$.

(C) Acceleration \ddot{x}

As was shown above

$$\ddot{x} = -a\omega^2 \cos \theta$$

$$= -a\omega^2 \cos \omega t$$

Special cases:

When $x = 0, \quad \ddot{x} = 0$
 When $x = a, \quad \ddot{x} = -\omega^2 a$ } its largest value
 When $x = -a, \quad \ddot{x} = +\omega^2 a$ } value

These results can also be found using calculus.
 Since $x = a \cos \omega t$

$$\frac{dx}{dt} = \dot{x} = -a\omega \sin \omega t$$

$$\frac{d^2x}{dt^2} = \ddot{x} = -a\omega^2 \cos \omega t$$

These results assume $\theta = 0$ (and so $x = a$) when $t = 0$. If we had chosen $\theta = 3\pi/2$ rad (and so $x = 0$) when $t = 0$, we would have found

$$x = a \sin \omega t$$

$$\dot{x} = a\omega \cos \omega t$$

$$\ddot{x} = -a\omega^2 \sin \omega t$$

Summary

Quantity	Variation with x	Variation with t
x	—	$x = a \cos \omega t$
\dot{x}	$\dot{x} = \pm \omega \sqrt{a^2 - x^2}$	$\dot{x} = -a\omega \sin \omega t$
\ddot{x}	$\ddot{x} = -\omega^2 x$	$\ddot{x} = -a\omega^2 \cos \omega t$

It is instructive to examine the graphs (fig. 10.4) which illustrate these equations. Note particularly the phase relationships.

See also figs. 10.2 and 10.3.

Note. The equations of uniformly accelerated motion (p. 29) can *never* be applied to s.h.m., since the acceleration changes with time.

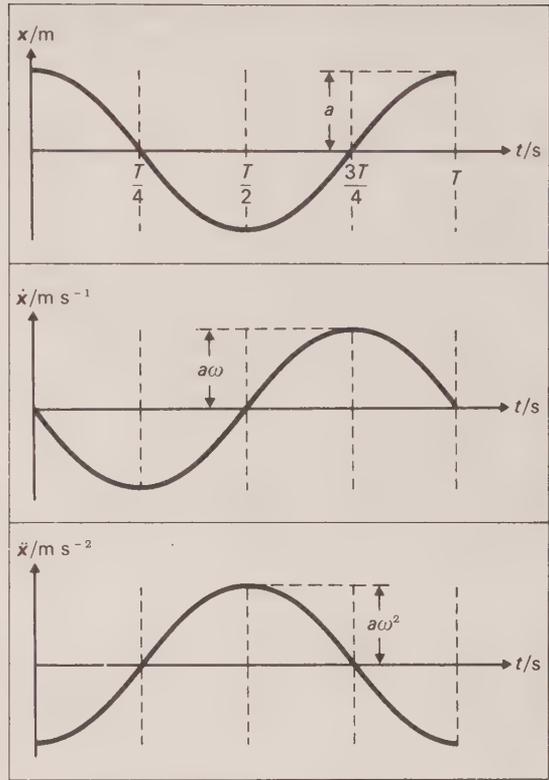


Fig. 10.4. Variation with t of x , \dot{x} and \ddot{x} .

Use of the auxiliary circle

Any linear s.h.m. can be regarded as having associated with it a point moving with constant speed in a circle whose diameter is the path of the object. Such a circle greatly simplifies the calculation of the time taken by the object to cover a given distance, without resort to algebra. If we relate \ddot{x} and x by the equation

$$\ddot{x} = -\omega^2 x,$$

then

(a) For the *object performing s.h.m.*, ω^2 is simply a positive constant, of dimensions $[T^{-2}]$, from which T can quickly be found.

$$T = \frac{2\pi}{\omega}$$

(b) For the *point in the auxiliary circle*, ω is an angular speed, of dimensions $[T^{-1}]$, and unit rad s^{-1} .

Beware of confusing ω and $\dot{\theta}$ for an angular s.h.m. described by

$$\ddot{\theta} = -\omega^2 \theta$$

Here ω does *not* represent an angular speed, but is merely a constant of the motion. It is here called the **pulsatance**.

10.3 THE DYNAMICS OF S.H.M.

To discover whether a body executes s.h.m.

(A) Linear motion

(1) Draw the body an *arbitrary* displacement x from the mean position O .

(2) Mark in *all* the forces that act on the body, and evaluate the restoring force ΣF .

(3) Mark the acceleration \ddot{x} in the chosen x -direction (that in which x increases).

(4) Apply the equation of motion

$$\Sigma F = m\ddot{x}$$

in a specified direction.

(5) The motion is s.h.m. if the equation of motion can be written

$$\ddot{x} = -(\text{positive constant}) x$$

The positive constant is usually written as ω^2 for convenience, since time period $T = 2\pi/\omega$.

Examples are given below.

(B) Angular motion

Repeat the procedure for an arbitrary angular displacement θ .

Apply torque $T = I\alpha$.

An equation of the form

$$\ddot{\theta} = -(\text{positive constant}) \theta$$

means that the body executes angular s.h.m.

Examples are given on pp. 57 and 409.

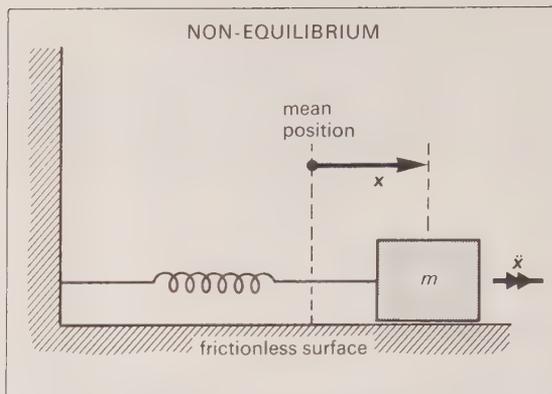


Fig. 10.5. A mass attached to a helical spring executes s.h.m.

10.4 EXAMPLES OF EXACT S.H.M.

(A) Mass on a helical spring

Suppose that in fig. 10.5 the spring exerts a force F on the mass, where

$$F = -kx$$

k is the **spring constant**, measured in N m^{-1} . The negative sign means that the force is to the left.

Apply the equation of motion to the right. Then $F = ma$ becomes

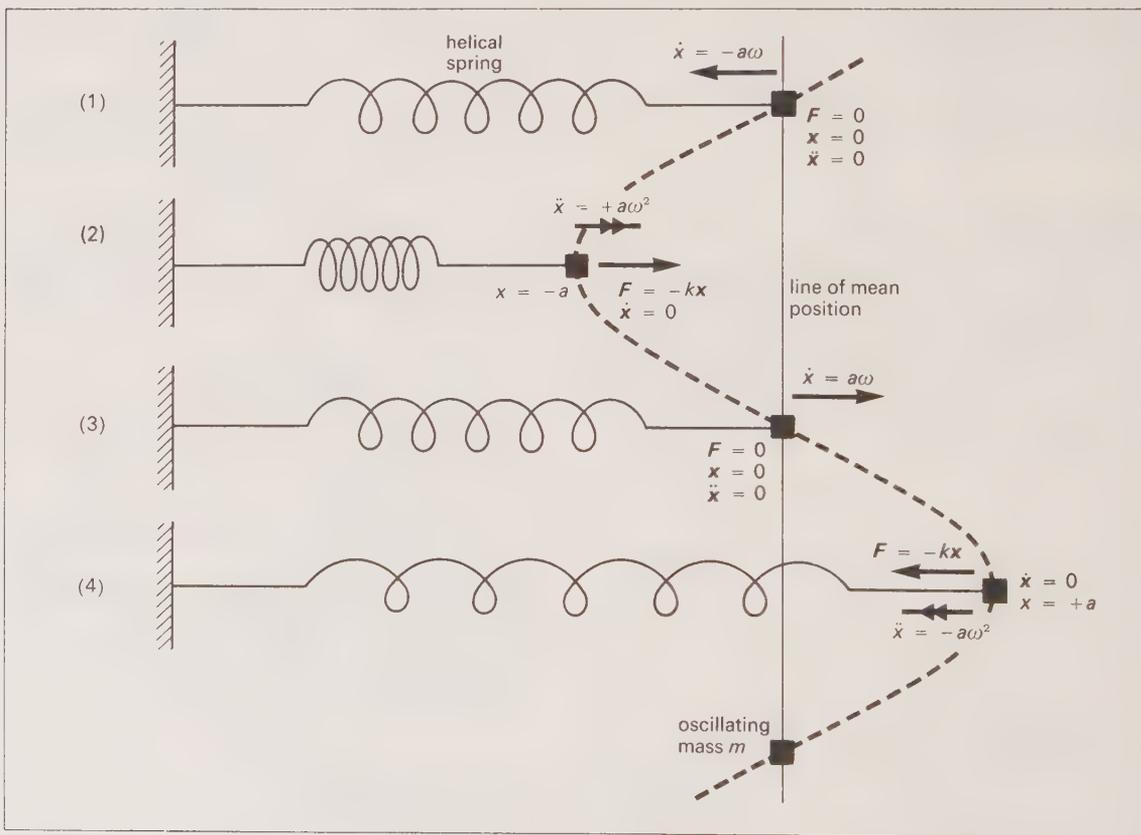


Fig. 10.6. The oscillation of the mass on the spring.

$$-kx = m\ddot{x}$$

$$\ddot{x} = -\left(\frac{k}{m}\right)x$$

This is s.h.m., so long as Hooke's Law is obeyed, of period

$$T = 2\pi\sqrt{\frac{m}{k}}$$

The motion is illustrated by fig. 10.6.

Notes

(a) T is independent of

(1) amplitude

(2) g (the weight mg of the mass m is not relevant).

$$(b) T = 2\pi\sqrt{\frac{\text{mass}}{\text{force at unit displacement}}}$$

(B) Cylindrical hollow tube floating in a liquid

In fig. 10.7 the test-tube is not in equilibrium, because it has been displaced x from the mean position.

The Archimedean upthrust exceeds the tube's weight by $(Ax)\rho g$.

Applying $F = ma$ in the downward direction, we have

$$-(A\rho g)x = m\ddot{x}$$

$$\ddot{x} = -\left(\frac{A\rho g}{m}\right)x$$

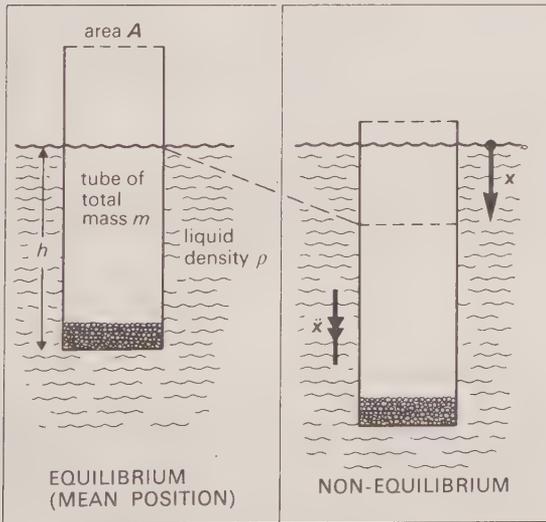


Fig. 10.7. A floating tube executes s.h.m.

We deduce that the motion is s.h.m. (in the absence of friction).

For the equilibrium situation

$$(Ah)\rho g = mg$$

$$\ddot{x} = -\left(\frac{g}{h}\right)x$$

so

$$T = 2\pi\sqrt{\frac{h}{g}}$$

h is a measure of the system's inertia, and g of its springiness. In the absence of a gravity field there would be no method of storing p.e., and so no oscillation could occur.

(In practice the liquid's viscosity produces two effects:

(i) it causes heavy damping (p. 91), and

(ii) it causes a significant mass of liquid to be set into oscillation. This liquid has taken its k.e. from the rod.)

A graph of T^2 against h would have gradient $4\pi^2/g$, from which g can be determined.

(C) Liquid oscillating in a U-tube

The motion is s.h.m. because the restoring force has a magnitude proportional to the length of displaced liquid, and so to the displacement. The force always acts so as to restore the liquid to the mean position. Viscous effects cause the motion to be damped (p. 91), but in their absence

$$T = 2\pi\sqrt{\frac{l}{2g}}$$

where l is the length of the liquid column.

A graph of T^2 against l would have gradient $2\pi^2/g$ which would again lead to a value for g .

(D) The torsional pendulum

On p. 148 it is shown that the time period T of a torsional pendulum consisting of a body of moment of inertia I suspended by a system of torsion constant c is given by

$$T = 2\pi\sqrt{I/c}$$

10.5 EXAMPLES OF APPROXIMATE S.H.M.

(A) The simple pendulum

(An angular treatment is given on p. 57.)

Refer to fig. 10.8 overleaf. The bob can move only along the tangent. Apply $F = ma$ to the right. Then

$$-mg \sin \theta = m\ddot{x}$$

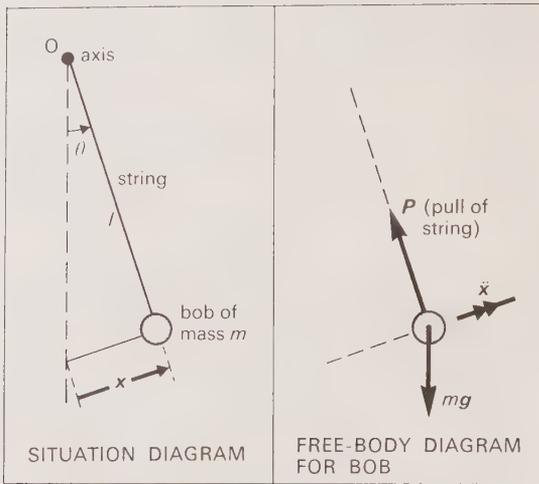


Fig. 10.8. The simple pendulum treated as a linear motion

where $mg \sin \theta$ is the magnitude of the resolved part of mg along the tangent.

When $\theta < \frac{1}{6}$ rad (10°) or so,

$$\sin \theta \approx \theta \approx \frac{x}{l}$$

if θ is measured in radians.

Then
$$-m\left(\frac{g}{l}\right)x \approx m\ddot{x}$$

$$\ddot{x} \approx -\left(\frac{g}{l}\right)x$$

So the motion is *approximate* s.h.m., and

$$T = 2\pi\sqrt{\frac{l}{g}}$$

T is independent of amplitude, *provided a is small*.

The angular treatment on p. 57 demonstrates that ml^2 is a measure of the system's inertia (its m.i.), while $mg l \theta$ is a measure of the springiness (the restoring torque).

Because inertial mass \propto gravitational mass, m cancels: T is independent of m . (Expressed differently, different masses always give the same T , demonstrating that all masses experience the same acceleration g at a given point.)

(B) Heavy piston inside frictionless cylinder

In fig. 10.9 suppose that atmospheric pressure can be ignored, and that all dissipative effects (p. 91) are negligible.

If the displacement x is small, then the pressure

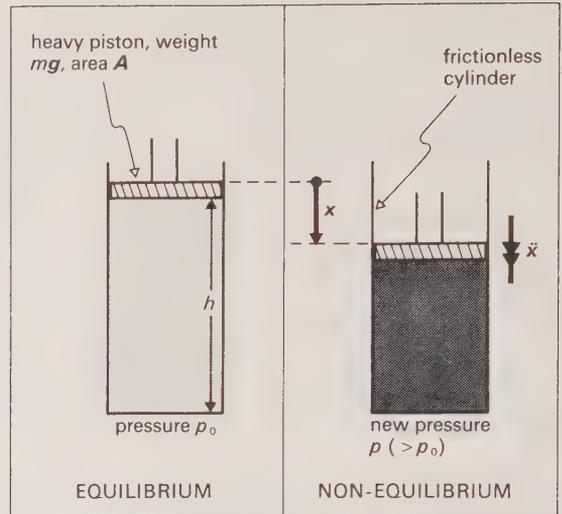


Fig. 10.9. Approximate s.h.m. shown by a heavy piston.

changes will be so small that they are approximately proportional in size, but opposite in sign, to the volume changes, and hence to the displacements x . (Expressed differently, we can approximate the *Boyle's Law* hyperbola to a straight line over small pressure ranges: the line's slope is negative.) The force resulting from the pressure change is then proportional to x , but oppositely directed: s.h.m. occurs.

It can be shown that

$$T = 2\pi\sqrt{\frac{h}{g}}$$

for small displacements.

The inertial factor here is m , the mass of the piston, and the springiness factors the pressure of the gas (since a compressed gas can store p.e.) and the gravitational p.e. of the piston.

(C) The compound pendulum

On p. 62 it is shown that the time period T of a compound pendulum is given by

$$T = 2\pi\sqrt{\frac{k^2 + h^2}{hg}}$$

10.6 ENERGY IN S.H.M.

When an oscillation is **undamped**, the system does no work against resistive forces, and so its total energy remains constant in time. Fig. 10.10 demonstrates that the p.e. and k.e. oscillate with double the frequency of the motion. (See also a.c., on p. 442.)

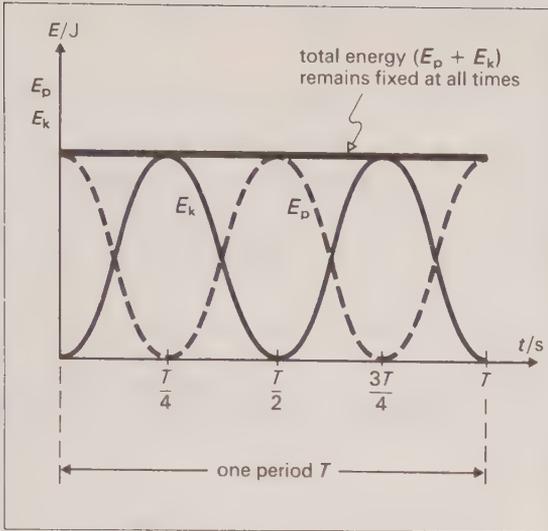


Fig. 10.10. Time variation of the energy of an s.h.m.

(a) Potential energy

$$\begin{aligned}
 \text{at extension } x \quad E_p &= \int_0^x F dx \\
 &= \int_0^x kx dx \\
 &= \frac{1}{2}kx^2 \quad \left(\omega^2 = \frac{k}{m} \right) \\
 &= \frac{1}{2}m\omega^2x^2
 \end{aligned}$$

(b) Kinetic energy

$$\text{at extension } x \quad E_k = \frac{1}{2}mv^2 = \frac{1}{2}m\omega^2(a^2 - x^2)$$

(c) Total energy

$$\begin{aligned}
 \text{Total energy} &= E_p + E_k \\
 &= \frac{1}{2}m\omega^2x^2 + \frac{1}{2}m\omega^2(a^2 - x^2) \\
 &= \frac{1}{2}m\omega^2a^2
 \end{aligned}$$

Since the total energy is independent of the variable x , it remains constant. Note that the energy of an oscillation is proportional to

- (i) mass,
- (ii) (frequency)²,
- (iii) (amplitude)².

► 10.7 ENERGY APPROACH TO S.H.M.

Suppose that a system capable of oscillation can possess energy in the forms

$$E_p = \frac{1}{2}ax^2$$

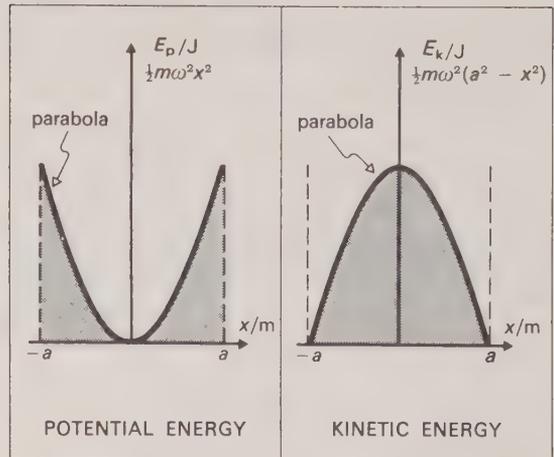


Fig. 10.11. Variation with position of the energies of an s.h.m. As in fig. 10.10, the total energy remains constant.

and

$$E_k = \frac{1}{2}bx^2$$

where a and b are positive constants for the system, and x is a variable.

If the system loses no energy (is undamped) the total energy

$$E = (E_p + E_k)$$

is constant, so $E = \frac{1}{2}ax^2 + \frac{1}{2}bx^2$

Differentiating w.r.t. time

$$ax\dot{x} + b\dot{x}x = 0$$

$$\therefore \ddot{x} = -\left(\frac{a}{b}\right)x$$

(if $\dot{x} \neq 0$).

The system is capable of s.h.m. of time period

$$T = 2\pi\sqrt{\frac{b}{a}}$$

Examples

Oscillation	E_p	E_k	T
mass on a helical spring	$\frac{1}{2}kx^2$	$\frac{1}{2}m\dot{x}^2$	$2\pi\sqrt{\frac{m}{k}}$
torsional pendulum	$\frac{1}{2}c\theta^2$	$\frac{1}{2}I\dot{\theta}^2$	$2\pi\sqrt{\frac{I}{c}}$
charge in LC circuit	$\frac{1}{2}\left(\frac{1}{C}\right)Q^2$	$\frac{1}{2}L\dot{Q}^2$	$2\pi\sqrt{LC}$

► 10.8 SUPERPOSITION OF TWO S.H.M.

In this section we discuss the behaviour of a particle which is subjected to two simultaneous simple harmonic motions in the same plane.

(A) Same line, same frequency

The resultant motion is an s.h.m. of the same frequency. The particle's resultant displacement is found by using the *principle of superposition* (p. 109), and a special case of this behaviour is discussed under *stationary waves* (p. 110).

(B) Same line, different frequencies

The resultant motion is not *simple* harmonic because the amplitude oscillates with time—it is said to be **modulated**. The frequency of oscillation of the amplitude equals the difference of the frequency of the superposed motions. The behaviour is analogous to the sound emitted by two tuning forks of close but different frequencies—the variation with time of the sound intensity is called **beating** (p. 327). Fig. 10.12 shows what happens when two motions have different amplitudes a_1 and a_2 ($a_1 > a_2$).

See p. 115 for a situation where $a_1 = a_2$.

(C) Perpendicular directions

(a) *Same frequency*. In general the path of the particle is an ellipse. When the amplitudes are equal, and the phase differences are $0, \pi/2$ rad and π rad, we have a straight line, circle and straight line respectively (p. 103).

(b) *Different frequencies*. The path of the particle is a continuous curve whose general form is controlled by the phase difference and frequency ratio of the superposed motions. Fig. 10.13 shows the **Lissajous figures** for a phase difference of $\pi/2$ rad between two motions whose frequency ratio f_y/f_x takes some simple values.

If the two simple harmonic motions have different amplitudes, the *Lissajous* figures have different shapes in detail, but their general form is not altered.

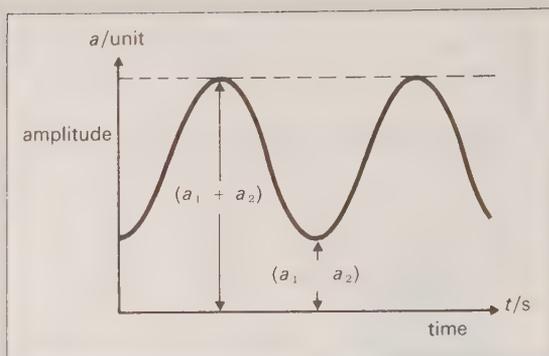


Fig. 10.12. Amplitude fluctuation (beats).

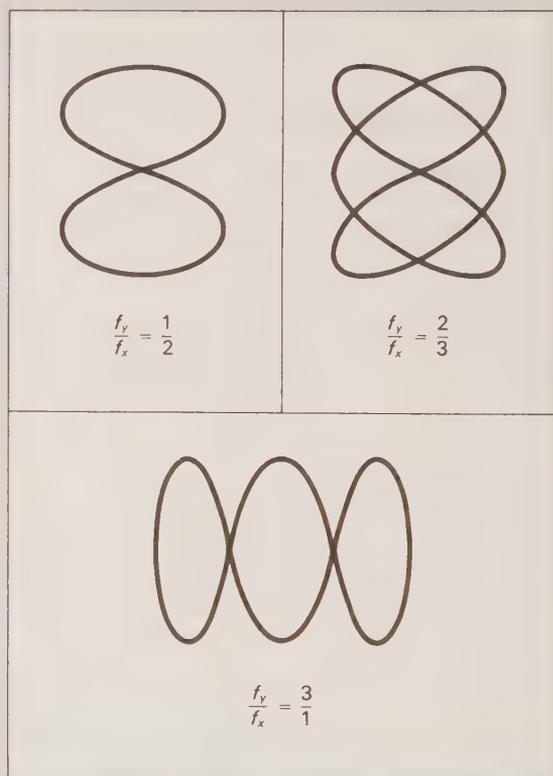


Fig. 10.13. Some simple Lissajous figures.

11 DAMPED AND FORCED OSCILLATION

11.1	DAMPED OSCILLATIONS	91
11.2	AMPLITUDE OF DAMPED OSCILLATIONS	92
11.3	FORCED VIBRATION	93

11.4	RESONANCE	94
11.5	RESONANCE AND ENGINEERING STRUCTURES	94
▶ 11.6	QUALITY FACTOR OR Q-FACTOR	95

11.1 DAMPED OSCILLATIONS

When a system executes *true* s.h.m., then

- (a) its period is independent of its amplitude,
- (b) its total energy remains constant in time.

In practice many bodies execute *approximate* s.h.m. because

- (a) F is *not* proportional to $-x$, and so the oscillations are not isochronous, and
- (b) the energy of the system decreases in time (fig. 11.1).

A real oscillating system is opposed by dissipative forces, such as air viscosity. The system then does positive work: the energy to do this work is taken from the energy of oscillation, and usually

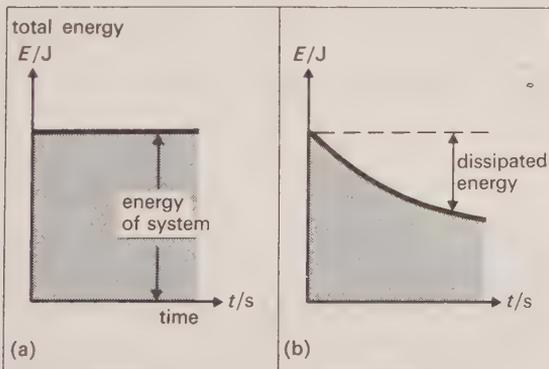


Fig. 11.1. Energy comparisons for (a) ideal and (b) approximate (damped) s.h.m.

appear as thermal energy. **Damping** is the process whereby energy is taken from the oscillating system.

Examples of damping

(A) Natural

(a) The internal forces of a spring can be dissipative, and prevent the spring being perfectly elastic.

(b) Fluids exert viscous forces that prevent relative movement of fluid layers, and hence oppose movement through the fluid. The internal energy of the fluid increases.

(B) Artificial

(a) Galvanometers utilize either electromagnetic damping (p. 430) or dashpots to prevent tiresome oscillation. (See also below.)

(b) The panels of cars are coated to reduce vibration.

(c) The shock absorber of a car provides a damping force which prevents excessive oscillation.

Examples of damped vibrations

(A) Natural

(a) A percussive musical instrument (such as a bell) gives out a note whose intensity decreases with time.

(b) The paper cone of a loudspeaker vibrates, but is heavily damped so as to lose energy (as sound wave energy) to the surrounding air.

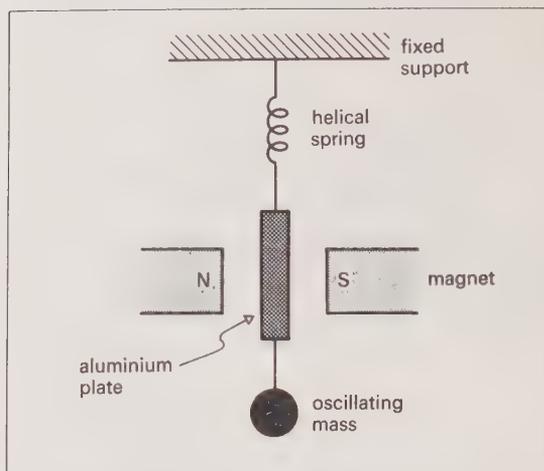


Fig. 11.2. Apparatus to demonstrate the electromagnetic damping of a mass oscillating on a helical spring.

(B) Demonstration

Refer to fig. 11.2.

Note three characteristics:

(a) the mass, which can possess k.e. because of its inertia,

(b) the spring, which can store p.e. because of its elasticity, and which provides a restoring force $F = -kx$, and

(c) the aluminium plate which experiences a resistive force

$$R \approx -(\text{constant})\dot{x}$$

because of eddy current damping (p. 430). This force causes *Joule* heating in the plate.

The resistive forces in damping are frequently roughly proportional to the speed of movement (p. 165). An assumption that they are so, greatly simplifies the mathematical analysis.

Notes

(a) When the mass is at rest, $R = 0$. R , on its own, would never stop the body from reaching the equilibrium position.

(b) The frequency of a damped motion is slightly less than that of an undamped motion.

11.2 AMPLITUDE OF DAMPED OSCILLATIONS

Fig. 11.3 shows displacement–time graphs for three categories of damping.

(a) **Slight damping** results in definite oscillation, but the amplitude of oscillation decays exponentially with time. The **decrement**

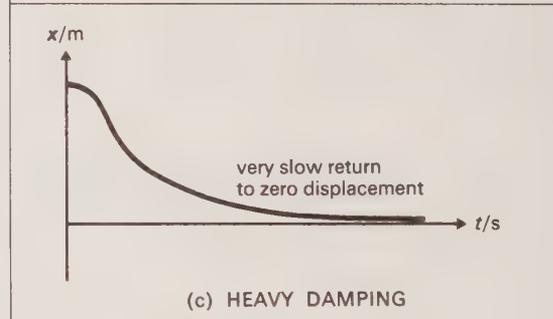
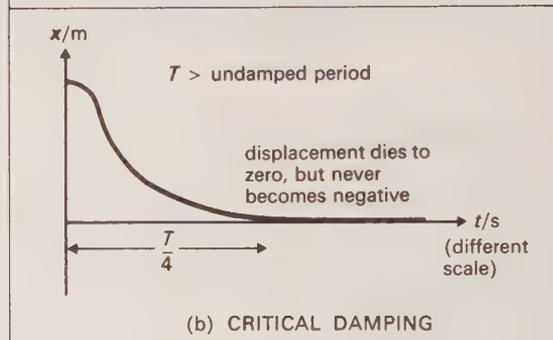
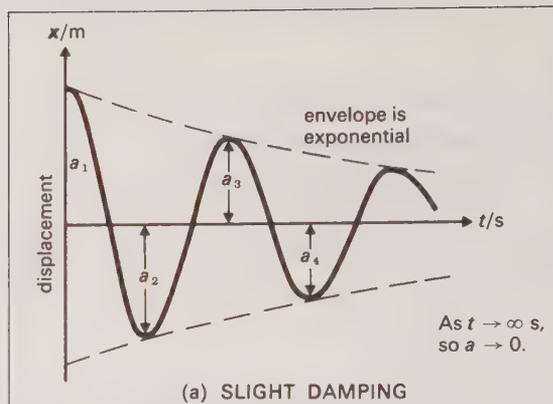


Fig. 11.3. Three kinds of damped oscillation.

$$\delta = \frac{a_1}{a_2} = \frac{a_2}{a_3} = \frac{a_3}{a_4} \dots \text{etc.}$$

For exponential decay δ is a constant. The **logarithmic decrement** $\ln \delta$ is used as a criterion of the amount of damping, and for making corrections.

(b) **Critical damping** results in no real oscillation as such: the time taken for the displacement to become effectively zero is a *minimum*.

(c) **Heavy damping** is a damping in which the resistive forces exceed those of critical damping. The system returns very slowly to the equilibrium position.

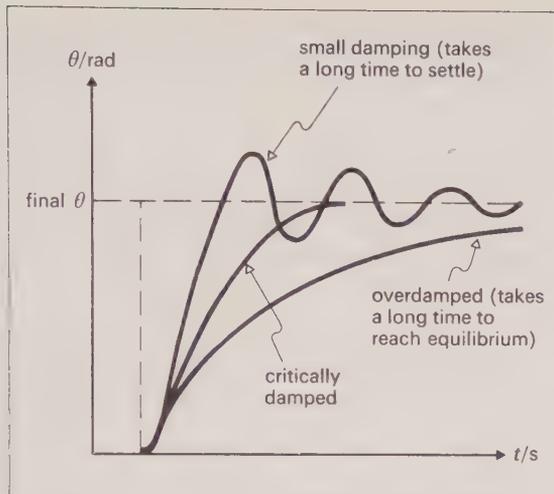


Fig. 11.4 θ - t curves for a moving-coil galvanometer being used to measure an electric current which becomes steady.

The moving-coil galvanometer

A galvanometer is

- (a) **dead-beat** when it is critically damped,
- (b) used **ballistically** when the damping is as small as possible. (See p. 420.)

Fig. 11.4 should be studied carefully.

11.3 FORCED VIBRATION

Suppose that in fig. 11.2, the support is set into continuous oscillation by some external agent, so that a periodic driving force is imposed on the oscillating mass. We will consider what happens when conditions are *steady*. (In the early stages beats (p. 114) will occur between the forced and natural vibration, giving rise to **transient** oscillations which we will ignore. The frequency of the natural free vibration is determined by the inertia and elasticity factors of the system.)

One observes the following:

(A) Energy

Under steady conditions the amplitude of vibration is fixed for a fixed driving frequency. We deduce that the driving force does work *on* the system at the same rate as the system loses energy by doing work against dissipative forces. The power of the driver is controlled by damping.

(B) Amplitude

The amplitude of vibration depends on

- (a) the relative values of the natural frequency

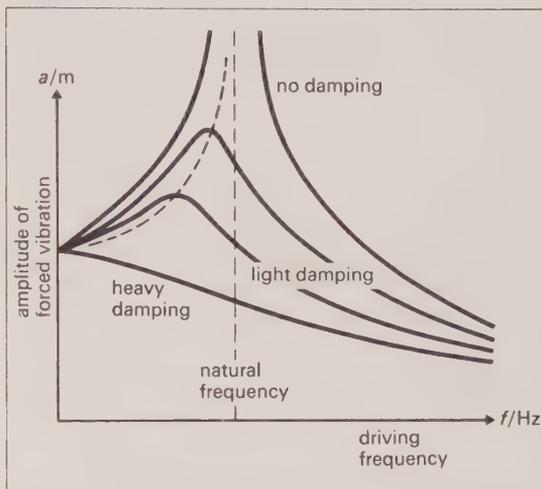


Fig. 11.5. The response of damped systems to varying driving frequencies.

of free oscillation, and the frequency of the driving force, and also

- (b) the extent to which the system is damped.

These observations are summarized by fig. 11.5.

Damping shows two effects:

- (i) It causes the maximum amplitude to be reached when the driving frequency is a little less than either the damped or undamped natural frequencies.
- (ii) It reduces the response of the forced system over a range of frequencies. (It cuts down the sharp peak of fig. 11.5.)

(C) Phase

Whatever its natural frequency, the forced vibration takes on the frequency of the driving force, but it shows the *phase lag* illustrated by fig. 11.6 (overleaf).

Barton's pendulums

The apparatus of fig. 11.7 (overleaf) can be used to demonstrate these observations.

When X oscillates perpendicular to the plane of the paper, the steady oscillations of pendulums A to E demonstrate the effects shown in figs. 11.5 and 11.6. While A and B may be almost $\frac{1}{2}$ period behind X, C will be $\frac{1}{4}$ period behind, and D and E will be nearly in phase.

- C will show the largest amplitude.

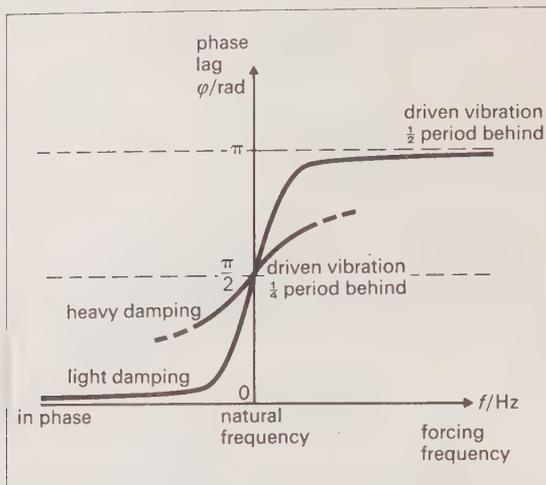


Fig. 11.6. Phase relationships for the amplitude of a forced vibration.

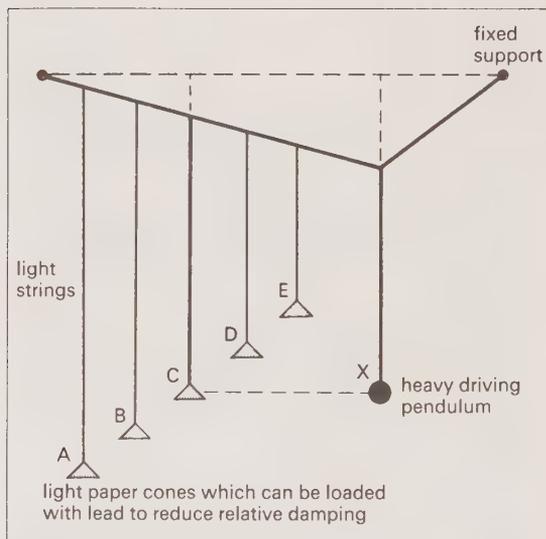


Fig. 11.7. Barton's pendulums.

11.4 RESONANCE

Fig. 11.5 illustrates that the *amplitude* of forced vibration reaches a maximum when the driving frequency is just less than the natural frequency of free oscillation. Several definitions of resonance are possible.

Energy resonance (velocity resonance) occurs when an oscillator is acted upon by a second driving oscillator whose frequency equals the natural frequency of the system.

When this happens the driving oscillator (some external agent) most easily transfers its energy to the oscillating system: *the energy of the system becomes a maximum.*

(a) For zero damping, the energy would (in theory) become infinite,

(b) for light damping, the driving frequency is close to that required for amplitude resonance.

Examples of resonance are widespread in all branches of physics.

(a) **Mechanics**

(i) The oscillation of a child's swing.

(ii) The vibration of machinery and vehicles containing rotating parts.

(b) **Sound**

(i) Resonance tube (p. 332).

(ii) Kundt's tube (p. 332).

(c) **Electricity.** Tuning a radio circuit—the natural frequency of the circuit is made equal to that of the incoming electromagnetic wave by changing its capacitance. (See p. 454.)

(d) **Light.** Maximum absorption of infra-red waves by an NaCl crystal occurs when their frequency equals that of the vibration of the Na^+ and Cl^- ions.

► (e) **Modern physics.** Reference should be made to more detailed books for **paramagnetic resonance** and the concept of **scattering cross-section**.

Resonant systems

(a) Consider a mass on a helical spring. The inertia of the system and the restoring force are both *concentrated*: there is *one value* of the driving frequency for resonance.

(b) In a system such as the resonance tube (p. 332), the inertia and the restoring force are *evenly distributed* in the system: there are *several* (related) *driving frequencies* which give resonance.

11.5 RESONANCE AND ENGINEERING STRUCTURES

High damping in buildings consisting of brick or stone results from the many joints which exert frictional forces on the wind thereby dissipating its energy. Metal or reinforced concrete structures such as bridges, chimney stacks and cables are more likely to experience fluctuating forces due to **vortex shedding**. If the frequency of vortex shedding coincides with a structure's natural frequency, the amplitude of vibration of the structure

may become too great to be supported, e.g. the *Tacoma Narrows* bridge collapse (near *Seattle*, 1940). Structures can resist these forced vibrations by having:

- (a) high stiffness — and therefore a high natural frequency,
- (b) the ability to absorb energy — they might require additional damping, e.g. resilient pads in chimneys and dumb-bell dampers on cables,
- (c) large mass — note that the tendency nowadays is towards structures of smaller mass.

Extensive wind-tunnel tests are carried out on structures such as bridges in order to determine the most effective designs. One result of such tests has been to change the shape of tall chimney stacks by fitting helical strakes which prevent regular vortex shedding.

► 11.6 QUALITY FACTOR OR *Q*-FACTOR

The *Q*-factor of a system undergoing forced oscillation is one way of indicating the sharpness of resonance. It is defined by

$$Q = 2\pi \frac{\text{(maximum energy contained in system)}}{\text{(average energy dissipated per oscillation)}}$$

Under steady conditions the energy dissipated by the system equals the work done on it by the external agent. A system of high *Q* (lightly damped) would be expected to continue vibrating for a long time as the energy taken from the system during each oscillation is small. Some typical values of this dimensionless quantity *Q* are given below:

test tube in water	10^1
typical inductor	2×10^2
simple pendulum in air	4×10^2
quartz crystal in air	3×10^4

It can be seen from fig. 11.5 that the smaller the damping (i.e. the higher the *Q*-factor) the sharper the resonance. This is particularly important in the tuned circuit of a radio receiver where a highly selective response is required.

12 WAVE MOTION

12.1	PROGRESSIVE WAVES	96
12.2	ANALYTICAL DESCRIPTION OF WAVES	97
12.3	WAVE CLASSIFICATION	98
12.4	SUMMARY OF WAVE PROPERTIES	99
12.5	REFLECTION	99

12.6	REFRACTION AND DISPERSION	100
12.7	POLARIZATION	102
▶ 12.8	THE EQUATION OF A SINUSOIDAL TRAVELLING WAVE	103
▶ 12.9	THE POWER OF A WAVE MOTION	104

Some familiarity with elementary wave motion is assumed in this chapter and in chapter 14. Most of the wave properties described can be conveniently demonstrated using (a) a ripple tank, (b) 30 mm electromagnetic wave equipment.

12.1 PROGRESSIVE WAVES

A **progressive or travelling wave** is the movement of a disturbance from a source which transfers energy and momentum from the source to places around it.

Waves are of two fundamental types:

- (a) **mechanical waves**, which require a material medium for their propagation, and
- (b) **electromagnetic waves**, which can travel through a vacuum. They are discussed in detail in chapter 37, p. 282.

Nevertheless all types of wave motion have the same basic properties, and can be treated analytically by equations of the same form. Wave phenomena occur in many branches of physics, and an understanding of the general *principles* underlying their behaviour is very important.

Wave characteristics

(a) A wave transfers *energy*: there is no net translation of matter. (Waves breaking on a beach are *not* typical of the waves that interest us.)

(b) In a mechanical wave the medium has particles which have inertia and elasticity (some means of interacting with their neighbours). These

particles execute oscillations of small amplitude about their equilibrium positions.

(c) Each particle oscillates in the same way as its neighbour, but shows a time lag if it is further from the source of energy.

Longitudinal and transverse waves

A wave is said to be

(a) **longitudinal** if the displacement of mechanical particles is *parallel* to the direction of translation of energy (as in sound waves), or

(b) **transverse** if the wave has associated with it some (vector) displacement which is *perpendicular* to the direction of translation of energy (as in electromagnetic waves).

These wave types can be demonstrated using a 'slinky' spring.

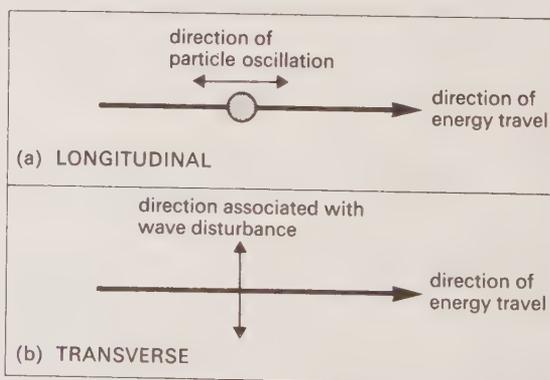


Fig. 12.1 Longitudinal and transverse waves.

Examples of progressive waves

Wave type	Nature of disturbance	Possible result of energy transfer
water waves	water molecules describe a periodic motion in a closed path	oscillation of a floating body
sound waves	vibratory motion superposed on random motion of air molecules	diaphragm of a microphone set oscillating
secondary earthquake waves	shaking motion (transverse movement) of material of Earth's crust	buildings collapse
electromagnetic waves	transverse variation of electric and magnetic fields	could stimulate retina of human eye

12.2 ANALYTICAL DESCRIPTION OF WAVES

Definitions

The **wavefront** is a surface (reducing to a line and point respectively in two and one dimension(s)) over which the disturbance has the same phase at all points.

Examples for an isotropic medium (one in which the wave speed is the same in every direction):

Nature of source	Shape of wavefront
point	point (in one dimension) circle (in two dimensions) sphere (in three dimensions)
line	line (in two dimensions) cylinder (in three dimensions)
plane	plane
any source	plane at a distant point

A **ray** is the path taken by the wave energy in travelling from the source to the receptor. A ray is usually perpendicular to a wavefront, and is then a **wave normal**.

The **speed** c of a *mechanical* wave motion is that with which an observer must move so as to see the wave pattern apparently stationary in space. (We will not distinguish between **phase** and **group velocity** in this book.) According to the theory of relativity, *electromagnetic* waves have the same speed in vacuo relative to all observers. The speed of the wave motion is then defined as that at which the wavefront passes the observer.

The **wavelength** λ is the spatial period of the wave pattern at a fixed instant, and is therefore the distance between two consecutive particles which have the same phase. (It could, for example, be the distance between adjacent crests.) The wavelength of most waves remains constant, but when waves have a small radius of curvature, the wave crests may not be evenly spaced.

The **frequency** f of the wave is the number of vibrations performed in each second by the source, and is thus the number of crests that pass a fixed point in each second. Unit: hertz (Hz) (see p. 84). (But see also *Doppler* effect, p. 119.)

Hence in time Δt a number $f\Delta t$ waves, each of length λ , pass a fixed point.

$$\therefore c = \frac{(f\Delta t)\lambda}{\Delta t}$$

$$c = f\lambda$$

$$c \left[\frac{\text{m}}{\text{s}} \right] = f \left[\frac{1}{\text{s}} \right] \lambda \left[\text{m} \right]$$

Wave diagrams

We will use the following symbols:

c for the *wave speed* (see below).

x for the *location* of a particular point in the medium.

y for the *disturbance* transmitted by the wave.

Thus for

- (a) string waves, y is the transverse displacement of a string particle,
- (b) sound waves, the longitudinal displacement of an air molecule,
- (c) electromagnetic waves, the electric field strength E (p. 282).

Two types of graph are valuable.

(A) Displacement–position curves

A plot of y against x can be called a **wave profile**: it represents what is happening *at an instant of time*.

On the graph the axes for x and y are drawn perpendicular to each other. In practice x and y are only perpendicular when we consider a transverse wave: thus fig. 12.2 can never represent *directly* the displacements of a longitudinal wave (as is done by a photograph).

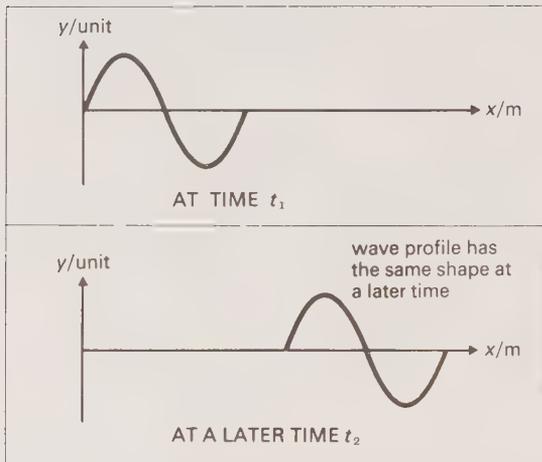


Fig. 12.2 Propagation of a wave profile.

Notes

- (a) The wave profile retains its shape when there is no **distortion**.
- (b) A **harmonic wave** is one whose wave profile is sinusoidal. (Refer to the comment on p. 103.)
- (c) Be careful not to confuse
 - (i) \dot{y} , which might represent an instantaneous particle speed, with
 - (ii) $\dot{x} = c$, which is the speed at which the wave profile propagates (the *wave speed*).

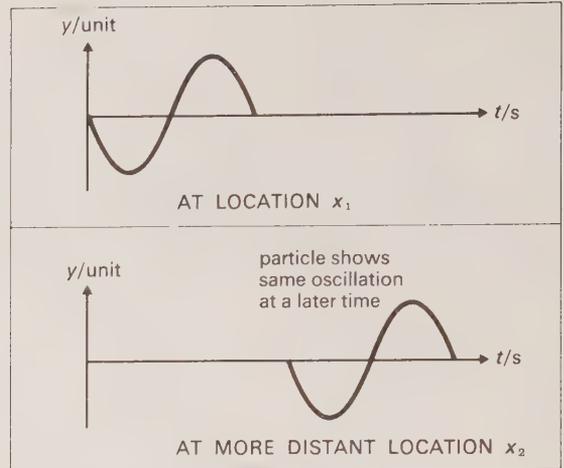


Fig. 12.3. Particles at two different locations showing the same oscillations at different times.

(B) Displacement–time curves

A plot of y against t (fig. 12.3) shows the nature of the oscillation of a particle at a *particular location*.

These ideas are developed in section 12.8 on p. 103.

12.3 WAVE CLASSIFICATION

Waves can be classified by

(A) Periodicity

(a) A *periodic wave* is generated by a periodic source which gives a **wavetrain**, a succession of disturbances.

(b) A *non-periodic wave* consists of a single wavefront (as is shown by a shock wave, p. 122).

(B) Energy movement

(a) A *progressive or travelling wave* transfers energy from one location (the source) to another.

(b) A *stationary or standing wave* shows no net energy transfer in space (but in time it shows an interchange of energy between kinetic and potential).

(C) Direction of displacement

- (a) *Longitudinal* (p. 96).
- (b) *Transverse* (p. 96).
- (c) *Torsional waves* are those propagated along a rod when one end is given a rapid twist.

(D) Nature of disturbance

Some examples are given on p. 97.

- Thus a hand-clap gives rise to a wave which is
- (1) non-periodic,
 - (2) progressive,
 - (3) longitudinal, in which
 - (4) the disturbance is a compression caused by the longitudinal displacement of air molecules.

12.4 SUMMARY OF WAVE PROPERTIES

These properties are brought together for reference.

(A) Properties controlled by (change of) medium

- (a) Wave speed (depends on inertial and elastic characteristics).
- (b) Reflection occurs at an interface where wave speed changes.
- (c) Transmission into the second medium occurs in (b), and may be accompanied by (i) refraction, and (ii) dispersion.
- (d) Polarization can occur with transverse waves only.

These are properties which can be shown by particles under appropriate conditions.

(B) Properties which can be explained by superposition (chapter 14)

- (a) Interference, beats and stationary waves appear when several waves coincide in the same region of space.
- (b) Diffraction and scattering occur when different obstacles obstruct the path of a wave, and distort the wavefront.

Apart from scattering, these properties are not shown by particles in classical physics.

It is again to be emphasized that, apart from dispersion and polarization, all waves should exhibit these properties: obviously some wave types lend themselves more easily to demonstration.

12.5 REFLECTION

(A) One dimension

Fig. 12.4 shows diagrammatically the phases of reflected and transmitted pulses for transverse waves.

These phase changes can be demonstrated when shaken pulses cross from one stretched spring to another of different physical properties.

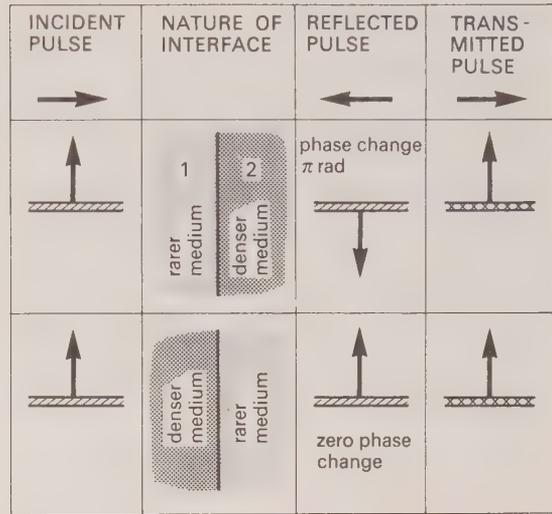


Fig. 12.4. Partial reflection and transmission of transverse waves.

- Notes. (a) Energy (not amplitude) is conserved.
 (b) There is a phase change of π rad when the wave is reflected at the denser medium: we will meet this again in physical optics (p. 292).
 (c) For an infinitely massive boundary there is no transmission and total reflection. For a string there is total reflection at a free end.

Fig. 12.5 illustrates total reflection for a longitudinal pulse on a spring.

We will use these ideas when we discuss sound reflection (p. 329).

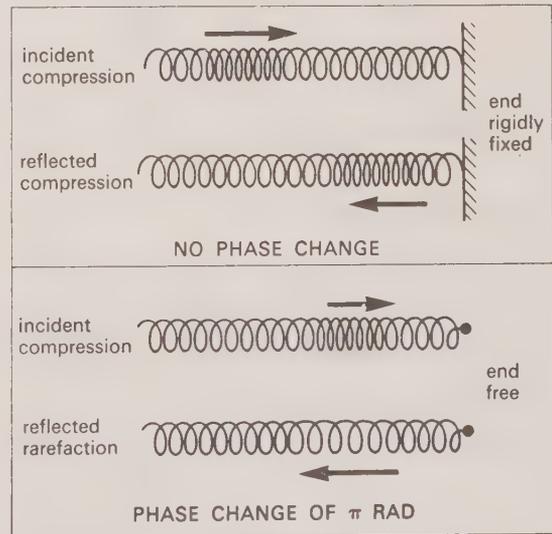


Fig. 12.5. Total reflection of a longitudinal pulse.

(B) Two dimensions

[A knowledge of the behaviour of waves in a ripple tank is assumed.]

(a) *Plane waves* readily demonstrate the law of reflection $\theta = \theta'$ (p. 252). Note that an angle is measured either

- (i) between a wavefront and a boundary, or
- (ii) between a ray and a normal.

(b) *Circular waves* illustrate the ideas of

- (i) *object* (centre of circular waves sent out by the source), and
- (ii) *image* (centre of circular waves after reflection or refraction).

Fig. 12.6 shows the formation of a virtual image (p. 256) by reflection of a circular wave at a plane surface.

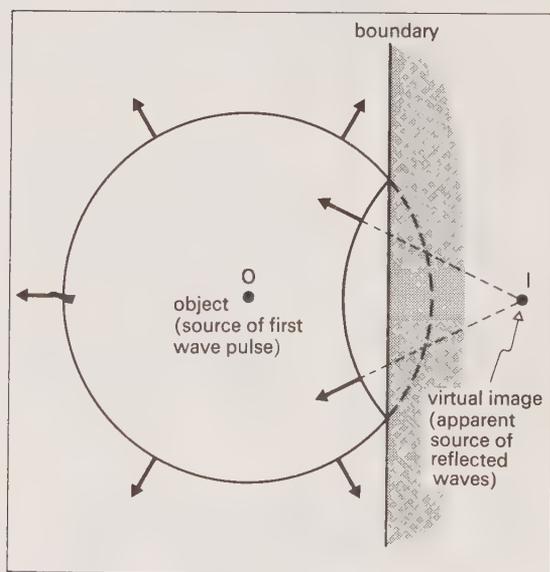


Fig. 12.6. Reflection at a plane surface.

(C) Uses of reflected waves

(a) *Radar*—pulses of high-frequency radio waves are transmitted towards distant objects and their reflections are received between the transmitted pulses. The direction and range of the object, e.g. land, aircraft or missile, are given by the direction of the receiving aerial and the time taken (measured using a c.r.o.) for the reflected part of the pulse to return. A **plan position indicator** (PPI) is a c.r.o. tube which produces an echo map of an area scanned by rotating transmitting and receiving aerials. The *Doppler* effect (p. 119) can be used to measure the speed of moving reflectors, e.g. a car in a police speed trap.

(b) *Sonar*—pulses of ultrasonic sound waves ($f > 20$ kHz) are transmitted towards underwater objects (e.g. sea bed, shoals of fish), and their reflected pulses are detected. The depth of the object can be calculated from the time taken for the reflected part of the pulse to return.

(c) *Ultrasound in medicine*—beams of ultrasonic sound waves are directed into the body and valuable information is given by their reflections. **Ultrasonic scanning** is used in echoencephalography, obstetrics and cardiography. This technique is extremely useful and safe, and the equipment is relatively simple.

(d) *Ionospheric waves*—a proportion of the radio waves from a transmitter travels through the air and is reflected by the ionosphere, layers of ionized gas 50–100 km from the Earth's surface. In this way, ionospheric or *sky waves* can be used for long-range communication. A further proportion (the *ground waves*) is diffracted over the Earth's surface and fading can occur when the ground waves interfere destructively with the sky waves. At high frequencies there is no ionospheric reflection and so v.h.f. radio and u.h.f. television transmissions for world-wide communication require reflection (and amplification) by artificial satellites (p. 176). Low-flying aircraft can cause fading of television reception if they obstruct the path of the transmitted waves towards the reflecting satellite.

12.6 REFRACTION AND DISPERSION**Refraction**

Fig. 12.7 shows in diagram form what happens when a wave impinges on an interface between two media.

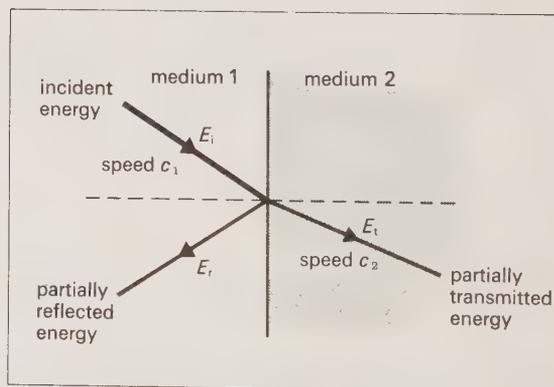


Fig. 12.7. Partial reflection and transmission.

When speed $c_1 \neq c_2$, some energy is nearly always reflected back into the original medium. At the same time the direction of travel of the transmitted energy deviates from that of the incident energy. This process (change of wavefront direction) is called **refraction**.

Since $c = f\lambda$, and f is fixed by the source, a decrease in c causes a decrease in λ .

Snell's law

The **relative refractive index**, ${}_1n_2$, for waves passing from medium 1 to 2 is given by

$${}_1n_2 = \frac{c_1}{c_2} \quad (\text{See p. 253})$$

Refer to fig. 12.8.

$${}_1n_2 = \frac{c_1}{c_2} = \frac{f\lambda_1}{f\lambda_2} = \frac{\lambda_1/AB}{\lambda_2/AB} = \frac{\sin \theta_1}{\sin \theta_2}$$

Since c_1 and c_2 are constants,

$(\sin \theta_1)/(\sin \theta_2)$ always has the same value for waves of a given frequency passing from medium 1 into medium 2.

This is **Snell's law**.

We have demonstrated that this constant value is the relative refractive index (defined as c_1/c_2).

Image formation

The general idea is shown in fig. 12.9, in which diffraction effects are ignored. An image is formed only if the shaded region is correctly shaped.

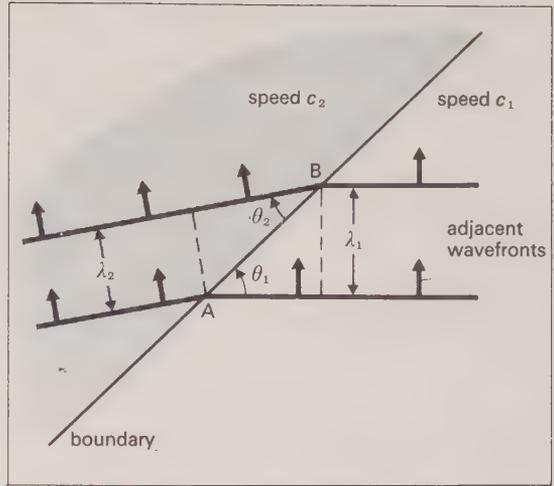


Fig. 12.8. 'Photograph' of a pair of adjacent wavefronts crossing a boundary.

The circular wavefront diverging from O is re-shaped by refraction to one which converges on I. The converging agent would be concave in shape if the medium of which it was made speeded up the wave motion. Examples of refractive media are given in the table overleaf.

Dispersion

Dispersion is that phenomenon which occurs when the speed of propagation of a wave motion depends on the wave frequency (and thus wave-

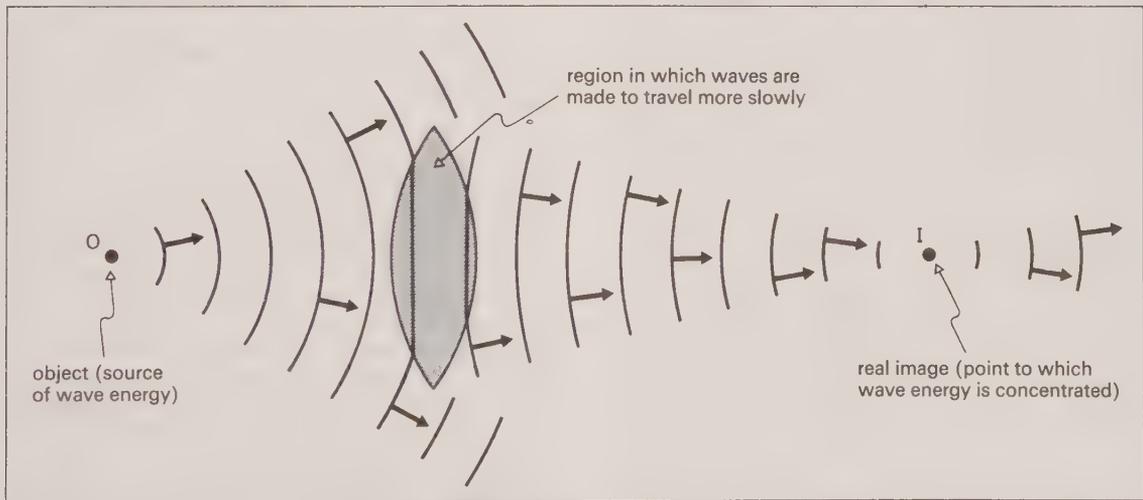


Fig. 12.9. Formation of a real image by refraction (ignoring diffraction).

Examples of media that cause refraction

Wave type	Medium in which wave travels more quickly ('rarer')	Medium in which wave travels more slowly ('denser')
ripple waves on water	deep water	shallow water
visible light	air	glass
sound	hydrogen	carbon dioxide
transverse waves on spring	spring of small mass per unit length	spring of greater mass per unit length
earthquake	dense rocks below surface	less dense surface rocks

length). The medium is called a **dispersive medium**.

Waves of different frequency may become separated

(a) by travelling different distances through the same medium in the same time, and/or

(b) by being deviated through different angles when refracted at an interface between two media.

Fig. 35.1 on p. 269 illustrates (b) for light waves.

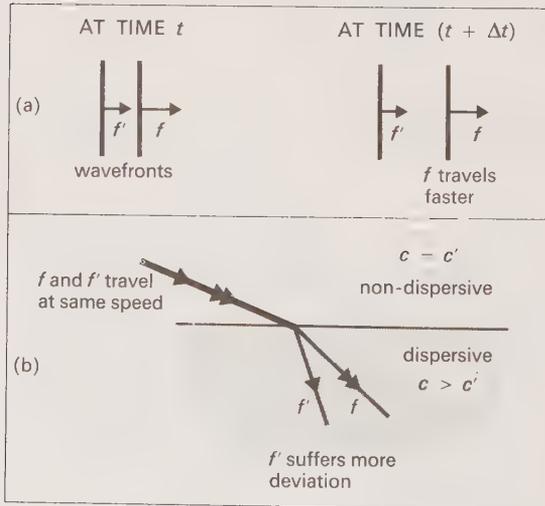


Fig. 12.10. Dispersion of two frequencies f and f' , (a) in one medium, (b) at an interface.

Examples of dispersion

Wave type	Dispersive medium
light	glass: higher frequencies (such as blue) travel at lower speeds
surface water waves	at large λ , $c = \sqrt{(g\lambda/2\pi)}$
surface water waves	at small λ , $c = \sqrt{(2\pi\gamma/\lambda\rho)}$ (ripples)
electro-magnetic	ionosphere

Air is *not* a dispersive medium for sound waves at audible frequencies.

► Dispersion and distortion

A genuine harmonic wave (which must in theory extend to infinity) has a unique frequency, and so remains undistorted when traversing different media. Any other waveform can be made by the superposition of harmonic waves of *different frequencies*: since these frequencies travel at different speeds in dispersive media, non-harmonic waves will then show distortion.

12.7 POLARIZATION

Refer to fig. 12.11.

(a) Suppose that the source oscillates parallel to the x -axis. Then any string particle is made to oscillate along a straight line, and the wave is **linearly (plane) polarized**. The **plane of polarization** is that containing the direction of energy travel, and the string displacement (the x - z plane). It is sometimes called the **plane of vibration**.

(b) Suppose that the source oscillates parallel to the x - and y -axes simultaneously, and that the two motions are *in phase*. Then we can superpose these two wave motions; their resultant is another linearly polarized wave.

(c) Suppose that in (b) the two vibrations *differ in phase by $\pi/2$ rad* but have the same amplitude. Now superposition gives a **circularly polarized** wave. At any given instant the string is *helical*.

(d) When the two motions of (c) have different amplitudes, the wave is **elliptically polarized**.

(e) When the motions have the same amplitude, and any phase difference other than 0,

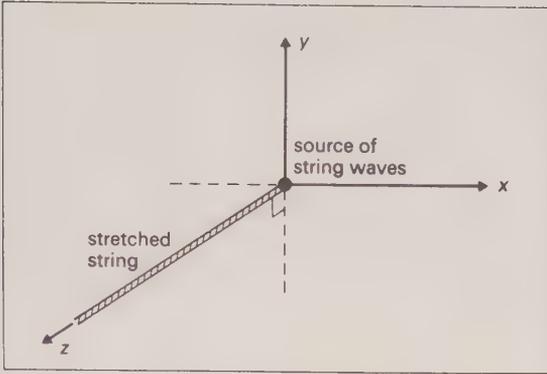


Fig. 12.11. Situation to describe polarization.

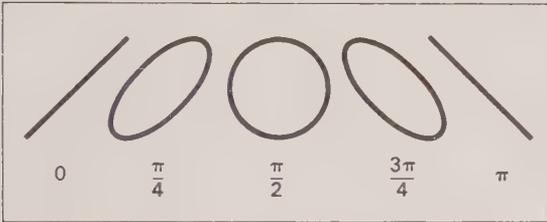


Fig. 12.12. End-on views of the string for superposed x - and y -motions of the source of the same amplitude, but which differ in phase.

$\pi/2$ rad or π rad, then the wave is again **elliptically polarized**.

The points are illustrated by fig. 12.12.

► Note that whereas a transverse wave on a string transmits *linear* momentum, a circularly polarized wave transmits *angular* momentum.

Example of polarization

Television and v.h.f. transmissions are polarized, and so receiving aerials are oriented so as to be parallel to the electric field vector of the electromagnetic wave.

In order to *match* the incoming wave, an aerial consists of a dipole made of two conducting rods each of length $\lambda/4$ where λ is the wavelength to be received. A reflector is used behind the receiving dipole and placed a distance $\lambda/4$ from it so as to boost the signal by constructive superposition. A series of parallel conducting rods in front of the dipole increases the strength of signal but this type of aerial has to be directed *accurately* towards the transmitter.

► 12.8 THE EQUATION OF A SINUSOIDAL TRAVELLING WAVE

Importance of sinusoidal waves

(a) They are generated by sources that describe s.h.m., and are therefore very common (p. 82).

(b) They propagate without distortion (other than that introduced by dissipation of energy).

(c) Any wave shape can be regarded as the superposition of sinusoidal waves of appropriate frequencies, amplitudes and phase relationships.

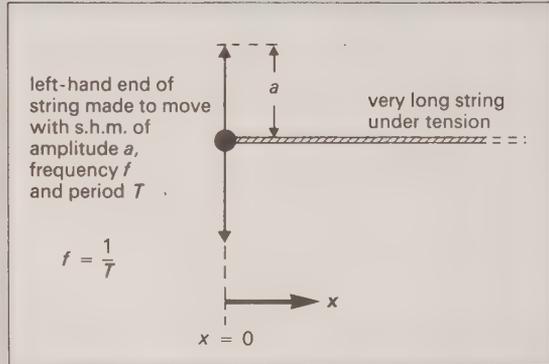


Fig. 12.13. To derive a wave equation.

Development of the equation

We write the transverse displacement as $y(x, t)$. Thus at the left end (where $x = 0$)

$$y(0, t) = a \sin \omega t$$

where $\omega = 2\pi f = \frac{2\pi}{T}$

At some location x ($x > 0$), the displacement y

- (a) varies with t in the same way, but
- (b) shows a **phase lag** δ .

For this location we write

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

Now if the wave speed c is constant

$$\delta \propto x$$

$$\delta = kx$$

where k is the circular wave number (see overleaf).

The equation of a one-dimensional wave travelling to the right is

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

It is important to note that the x and t terms have *opposite signs*.

For a wave travelling *to the left*, the displacement at x would show a **phase lead** of δ . We would write

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

in which the x and t terms have the *same sign*.

Useful terms and relationships

The **wavelength** λ is the distance between points having the same phase (same displacement, velocity and acceleration).

If x changes by λ , the phase changes by 2π rad.

$$\therefore k\lambda = 2\pi$$

$$k = \frac{2\pi}{\lambda}$$

The **circular wave number** k is the number of wavelengths in a length 2π units.

$$k[\text{m}^{-1}] = \frac{2\pi}{\lambda[\text{m}]}$$

Since $\omega = 2\pi/T$, we can write the equation in the alternative form

$$y = a \sin 2\pi \left(\frac{t}{T} - \frac{x}{\lambda} \right)$$

When t changes by a period T , the phase is unchanged so the location has changed by one wavelength λ . The wave profile advances by a distance λ in time T .

$$c = \frac{\lambda}{T} = f\lambda$$

It follows that
$$c = (2\pi f) \left(\frac{\lambda}{2\pi} \right)$$

$$= \frac{\omega}{k}$$

The **phase constant** ϕ is introduced for situations where

$$y \neq 0 \quad \text{when } x = 0, t = 0$$

Thus

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

The ideas of this section are developed in chapter 14, p. 109.

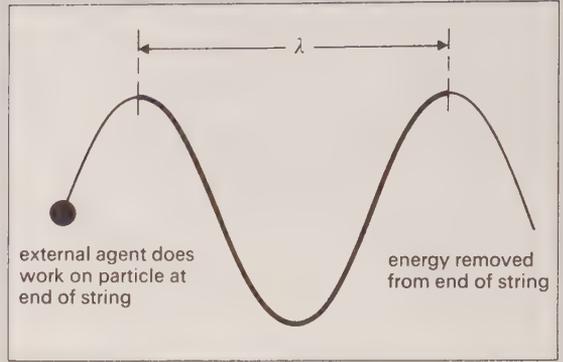


Fig. 12.14. Calculation of wave power.

► 12.9 THE POWER OF A WAVE MOTION

In this section we consider string waves to deduce a result which is of more general application. Refer to fig. 12.14.

Momentum and energy are transferred from particle to particle as the state of motion is propagated. The energy of each oscillating particle of mass m is $\frac{1}{2}m\omega^2a^2$, being partly k.e., and partly p.e.

If the string has mass/unit length ρ , a length λ has energy E where

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

The energy of length λ is taken steadily from the string in time T , so the **power of the wave** P is given by

$$P = \frac{E}{T}$$

$$= \frac{\frac{1}{2}\rho\lambda\omega^2a^2}{T} \quad \left(c = \frac{\lambda}{T} \right)$$

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

Note. The power of a wave motion is proportional to

- (a) (frequency)²
- (b) (amplitude)²
- (c) wave speed c .

Suppose a wave motion causes energy E to cross an area A at right angles to the wave velocity in time t . The **wave intensity** I is defined by

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

$$I \left[\frac{\text{W}}{\text{m}^2} \right] = \frac{P[\text{W}]}{A[\text{m}^2]}$$

Care must be taken to distinguish variations of *intensity* and variations of *amplitude* with distance r from the source (summarized in the table alongside).

The student should note the close similarity between these results and those on p. 344 (*Gauss's Law* in electrostatics).

$I \propto 1/r^2$ is simply the well-known **inverse square law**. Note the conditions to which it can be applied.

Shape of wavefront	Dependence of a on r	Dependence of I on r
plane	independent	independent
cylindrical	$a \propto \frac{1}{\sqrt{r}}$	$I \propto \frac{1}{r}$
spherical	$a \propto \frac{1}{r}$	$I \propto \frac{1}{r^2}$

13 HUYGENS' CONSTRUCTION

13.1 STATEMENT OF HUYGENS' HYPOTHESES

106

13.2 APPLICATION TO REFLECTION

107

13.3 APPLICATION TO REFRACTION

107

13.1 STATEMENT OF HUYGENS' HYPOTHESES

The hypotheses

Each point on an existing wavefront can be considered as a source of secondary wavelets.

(1) The wave amplitude at any point ahead can be obtained by superposition of these wavelets.

(2) From a given position of an uninterrupted wavefront, a later position can be determined as the envelope of the secondary wavelets.

Validity of the hypotheses

Huygens' construction is not a fundamental physical principle, but a useful mathematical device for the convenient solution of some problems in optics. It applies (in diffraction)

(a) to obstacles large relative to the wavelength, and

(b) at large distances for small angles of diffraction.

It takes no account of the physical interaction of a wave with an obstacle.

Uses of the construction

It can be applied to all types of wave motion, but is particularly useful for diffraction problems in physical optics (p. 303). In geometrical optics it is frequently clumsier than the use of the ray concept.

Explanation of propagation of wave motion

The following procedure has been adopted in fig. 13.1:

(1) Several arbitrary points on the existing wavefront have been chosen as centres of the secondary (spherical) wavelets.

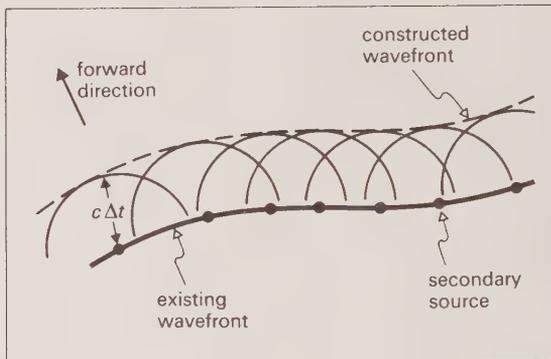


Fig. 13.1. Propagation of an arbitrary wavefront described by Huygens' construction.

(2) The radius of the secondary wavelets has been taken to be $c\Delta t$.

(3) The envelope of these wavelets has been taken to be the new wavefront after the time interval Δt .

[We assume (so that our result agrees with observation) the intensity of the spherical wavelets to vary continuously from a maximum in the forward direction to zero in the backward. Fresnel showed this assumption to be justified.]

Malus's Theorem

The discussions of the next two sections are based on

(a) Huygens' construction, together with

(b) **Malus's Theorem**. This states:

The time taken for one point on a wavefront to travel to its corresponding point (at a later time) is the same for all such pairs of points.

13.2 APPLICATION TO REFLECTION

Two situations serve as examples.

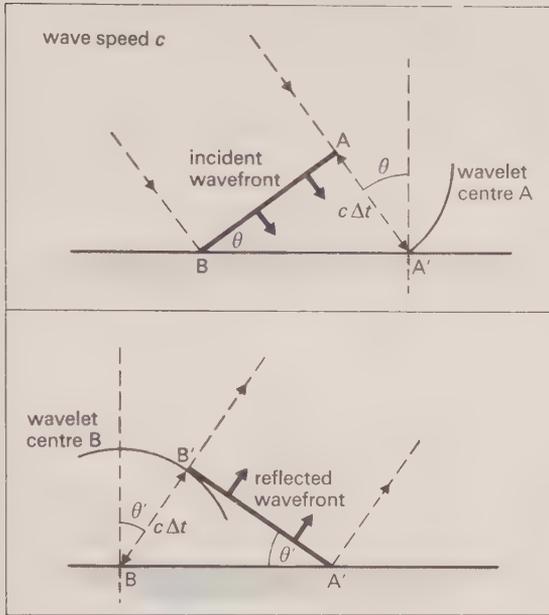


Fig. 13.2. The reflection of a plane wave at a plane surface explained by Huygens' construction.

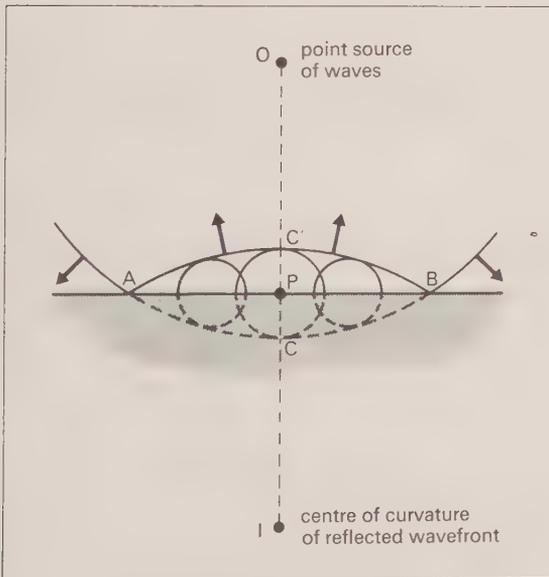


Fig. 13.3. Formation of an image described by Huygens' construction.

The law of reflection

Refer to fig. 13.2.

Suppose the wavefront at A takes time Δt to reach A'. Then $AA' = c \Delta t$.

The radius of secondary wavelet centre B is then also $c \Delta t$: the tangent to this wavelet from A' gives B', the point to which the wavefront previously at B has advanced in time Δt .

It follows that $\theta = \theta'$.

Image formation

Fig. 13.3 shows the reflection of a spherical wave at a plane surface.

The secondary wavelets have been drawn as they originate along the reflecting boundary: it follows that they have different radii.

$$\begin{aligned} \text{But} \quad & CP = C'P \\ \therefore & OP = IP \end{aligned}$$

a familiar result of elementary optics.

13.3 APPLICATION TO REFRACTION

Huygens' construction can be used to derive results such as the lens-maker's formula (p. 260), but ray optics is usually much neater. These examples illustrate some important ideas.

The law of refraction

Refer to fig. 13.4 (overleaf), which has been drawn by the method used for fig. 13.2.

Dispersion

If white light is used in a dispersive medium, the radius BB' ($= c_2 \Delta t$) will depend on the frequency of the waves. Then θ_2 will be different for different colours: this is the phenomenon of **dispersion**.

The *wave theory* can explain observed phenomena on the assumption that light travels *more slowly* in the medium in which it makes a smaller angle with the normal. *Newton's corpuscular theory* required that the light should travel *more rapidly* in such a medium if the corpuscular mass did not change as it crossed the boundary. Experiment (p. 289) verifies the wave assumption, and so for a time the wave explanation for refraction was preferred.*

* Nowadays we can combine the two since the theory of relativity suggests that the *mass* of the photon will change at the interface. *Momentum* is conserved parallel to the boundary, and in the denser medium it increases normal to the boundary. We must be careful when drawing conclusions about changes to the photon *velocity*.

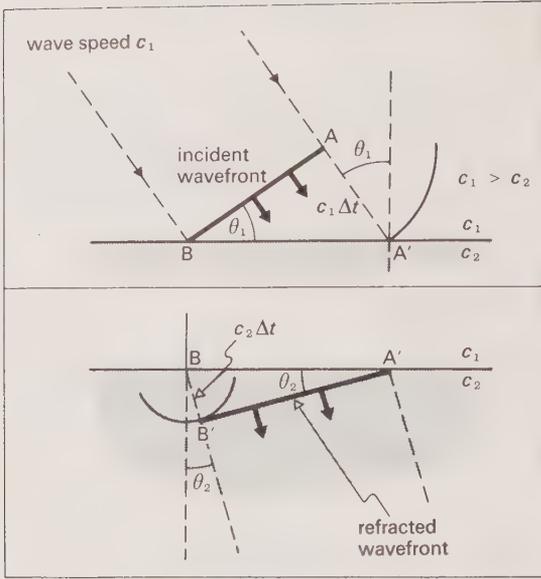


Fig. 13.4. The refraction of a plane wave at a plane surface explained by Huygens' construction.

Total internal reflection

Refer to fig. 13.5.

If the radius of the secondary wavelet centre A exceeds AC, no refracted wavefront can be drawn in medium 2 (since C would be inside the circle). **Grazing emergence** occurs when

$$\left(\frac{\text{time to travel}}{AC \text{ in medium 2}} \right) = \left(\frac{\text{time to travel}}{BC \text{ in medium 1}} \right)$$

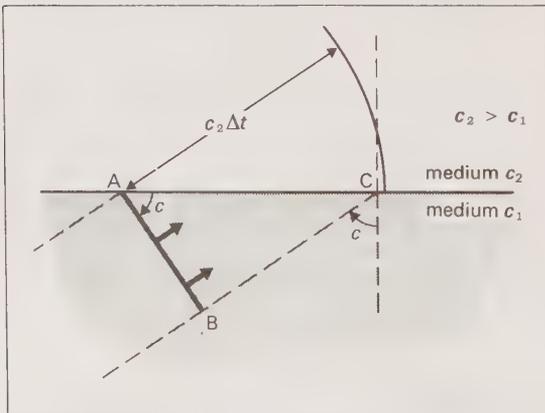


Fig. 13.5. Total internal reflection described by Huygens' construction.

$$\frac{AC}{c_2} = \frac{BC}{c_1}$$

$$\frac{BC}{AC} = \frac{c_1}{c_2} = \frac{n_2}{n_1} = \sin c$$

c is the **critical angle** for that pair of media.

Prism at minimum deviation

In fig. 13.6 wavefronts have been drawn for the ray that passes symmetrically.

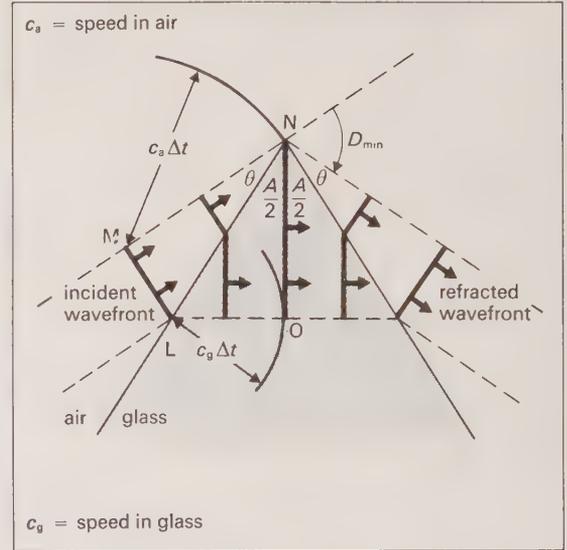


Fig. 13.6. Wavefronts drawn for a prism at minimum deviation.

The wavelets centred on M and L are drawn for the same interval Δt. Thus NO is a new wavefront which bisects the refracting angle at N.

Since

$$MN = c_a \Delta t \quad \text{and} \quad LO = c_g \Delta t$$

$$\begin{aligned} \frac{n_g}{n_a} &= \frac{c_a}{c_g} = \frac{MN}{LO} \\ &= \frac{NL \cos \theta}{NL \sin (A/2)} \end{aligned}$$

But $2\theta + A + D_{\min} = 180^\circ$

so $\frac{n_g}{n_a} = \frac{\sin [(A + D_{\min})/2]}{\sin (A/2)}$

a result which is proved for ray optics on p. 257.

14 THE PRINCIPLE OF SUPERPOSITION

14.1 THE PRINCIPLE OF SUPERPOSITION	109
14.2 STATIONARY OR STANDING WAVES	110
▶ 14.3 ANALYTICAL TREATMENT OF STATIONARY WAVES	112
14.4 INTERFERENCE	112
▶ 14.5 ANALYTICAL TREATMENT OF YOUNG'S EXPERIMENT	113

14.6 BEATS	114
▶ 14.7 ANALYTICAL TREATMENT OF BEATS	115
14.8 DIFFRACTION	116
14.9 THE IMPACT OF WAVES ON OBSTACLES	118

14.1 THE PRINCIPLE OF SUPERPOSITION

It is observed experimentally that when simple water waves of small amplitude collide, each wave has afterwards the same shape, momentum and energy as it had before the collision. We deduce that such waves do not interact with each other. (In more advanced physics it is observed that a collision between say a photon and an electron *does* affect the waves concerned: nevertheless our general conclusion still holds good.)

The principle of superposition follows:

The net disturbance at a given place and time caused by a number of waves which are traversing the same space is the vector sum of the disturbances which would have been produced by the individual waves separately.

$$y = y_1 + y_2 + y_3 \text{ etc.}$$

This should be compared with the superposition principle for forces (p. 35). Travelling particles cannot be superposed in this way: a collision between two particles alters the velocity of both.

The principle can be applied

(a) to all electromagnetic waves,

(b) to elastic waves in a deformed medium so long as the deformation is proportional to the restoring force, i.e. so long as *Hooke's Law* applies.

One importance of the principle is that it enables us to analyse complicated wave motions by

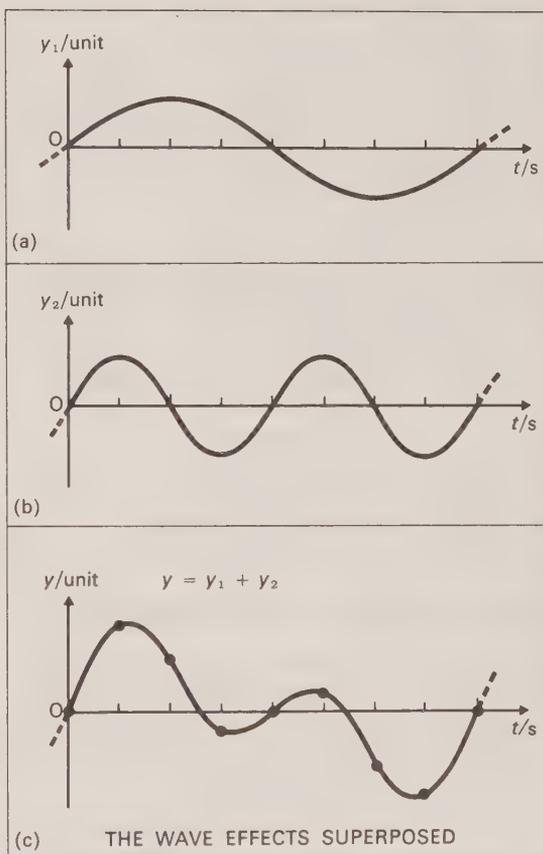


Fig. 14.1. An example of wave superposition.

representing them in terms of simple harmonic motions (p. 83).

An example of superposition

Fig. 14.1 shows the variation with time of the disturbance caused at a particular point by

- (a) a sinusoidal wave,
- (b) a second sinusoidal wave of the same amplitude but with twice the frequency, and
- (c) both waves together.

One can draw similar curves for y plotted against x . These would be for a fixed time instant (i.e. would be 'photographs').

Application of the principle

Wave superposition results in the phenomena classified in this table:

Phenomenon	Waves that are superposed
stationary waves	two wavetrains of same amplitude and frequency travelling at the same speed in opposite directions (a special case of interference)
interference	coherent waves from identical sources
beats	two wavetrains of close frequency travelling in the same direction at the same speed
diffraction	secondary wavelets originating from coherent sources on the same wavefront

The word *interference* is sometimes used where in this book we use *superposition*.

14.2 STATIONARY OR STANDING WAVES

Stationary or standing waves result from superposition of two wavetrains which have

- (a) the same amplitude and frequency (and therefore wavelength), but
- (b) equal but opposite velocities.

Two such wavetrains (supposed sinusoidal for convenience) are shown in fig. 14.2. The formation of stationary waves is thus a special case of interference (p. 112) which will be observed along the line joining the two sources in a *Young's* experiment.

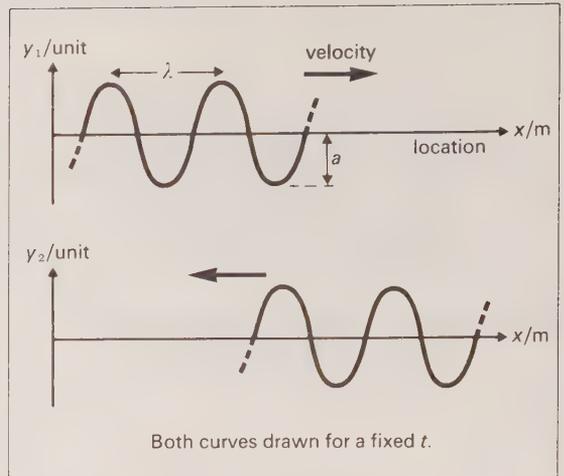


Fig. 14.2. Two wavetrains which give stationary waves when superposed.

It is convenient to describe the result of superposition in terms of what would be observed using transverse waves on a stretched string (fig. 14.3), and the first table on p. 111 compares and contrasts this with a progressive wave motion.

The *eye* sees a *time-average*, and cannot distinguish between the phases of adjacent segments, whereas a *photograph* can.

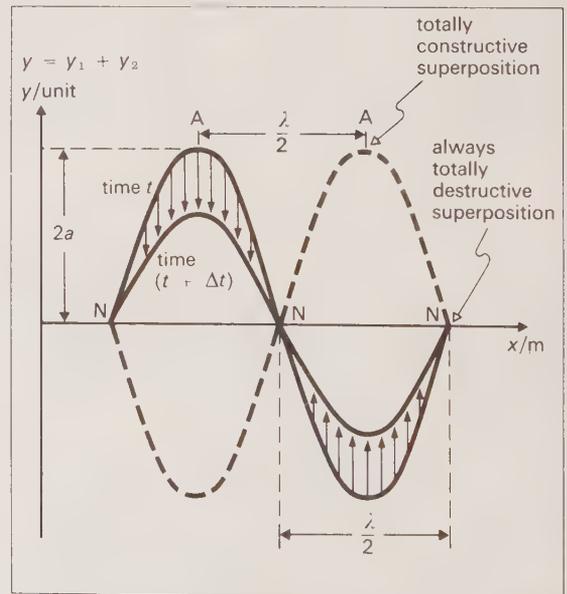


Fig. 14.3. Graphical representation of a stationary wave motion drawn at times t and $(t + \Delta t)$.

Comparison of stationary and progressive wave motion

	<i>Stationary</i>	<i>Progressive</i>
<i>amplitude</i>	Varies according to position from zero at the nodes (permanently at rest) to a maximum of $2a$ at the antinodes .	Apart from attenuation, is the same for all particles in the path of the wave.
<i>frequency</i>	All the particles vibrate in s.h.m. with the same frequency as the wave (except for those at the nodes which are at rest).	All particles vibrate in s.h.m. with the frequency of the wave.
<i>wavelength</i>	$2 \times$ distance between a pair of adjacent nodes or antinodes.	Distance between adjacent particles which have the same phase.
<i>phase</i>	Phase of all particles between two adjacent nodes is the same. Particles in adjacent segments of length $\lambda/2$ have a phase difference of π rad.	All particles within one wavelength have different phases.
<i>waveform</i>	Does not advance. The curved string becomes straight twice in each period.	Advances with the velocity of the wave.
<i>energy</i>	No translation of energy, but there is energy associated with the wave.	Energy translation in the direction of travel of the wave.

Formation of stationary waves

Frequently one wavetrain is derived from the (nearly) total reflection of the other. The foregoing description applies, but one is also concerned with **boundary conditions**.

The phase of the reflected wave can be found by
 (a) experimental observation (p. 99), or
 (b) working out the boundary condition appropriate to the type of wave and boundary being considered.

Examples of boundary conditions

<i>Wave</i>	<i>Boundary</i>	<i>Boundary condition</i>	<i>There exists at the boundary a(n)</i>	<i>Example on page</i>
transverse string wave	rigid wall	displacement always zero	displacement node	331
longitudinal spring wave	rigid wall	displacement always zero	displacement node	—
sound wave	rigid boundary	displacement always zero	displacement node } pressure antinode }	329
sound wave	open boundary	pressure change effectively zero	displacement antinode } pressure node }	329
electromagnetic wave	conducting metal surface	electric field zero (p. 346)	electric field vector node	289

► 14.3 ANALYTICAL TREATMENT OF STATIONARY WAVES

Suppose the waves of fig. 14.2 are represented by

$$y_1 = a \sin(\omega t - kx) \quad (\text{to the right})$$

and $y_2 = a \sin(\omega t + kx)$ (to the left)

Using the principle of superposition, the equation of the standing wave is given by

$$\begin{aligned} y &= y_1 + y_2 \\ &= a \sin(\omega t - kx) + a \sin(\omega t + kx) \\ &= 2a \cos kx \sin \omega t \end{aligned}$$

which may be written

$$y = \left(\begin{array}{c} \text{location} \\ \text{dependent} \\ \text{term} \end{array} \right) \cdot \left(\begin{array}{c} \text{time} \\ \text{dependent} \\ \text{term} \end{array} \right)$$

Deductions from the equation

The equation represents fig. 14.3.

(a) Every particle oscillates at a pulsation

$$\omega = 2\pi f = 2\pi/T$$

(b) The oscillation amplitude varies with location according to $\cos [(2\pi x)/\lambda]$, since $k = 2\pi/\lambda$.

(c) At the *nodes*

$$\begin{aligned} y &= 0 \quad \text{always} \\ \cos \frac{2\pi x}{\lambda} &= 0 \\ \frac{2\pi x}{\lambda} &= n \left(\frac{\pi}{2} \right), \quad n = 1, 3, 5, \text{ etc.} \end{aligned}$$

Distance between nodes = $\lambda/2$.

(d) At the *antinodes* y reaches a maximum amplitude $2a$. This happens when

$$\begin{aligned} \cos \frac{2\pi x}{\lambda} &= \pm 1 \\ \frac{2\pi x}{\lambda} &= m\pi, \quad m = 0, 1, 2, \text{ etc.} \end{aligned}$$

Distances between antinodes = $\lambda/2$.

(e) There is no net energy transfer.

14.4 INTERFERENCE

The term **interference** is best applied to the result of superposing waves from *identical sources*.

When a *pair* of sources is used, we frequently refer to **Young's experiment** (p. 294).

If an interference pattern (or a system of **fringes**) is to be seen, the sources must be coherent. They must also produce waves of about the *same amplitude*.

Sources are **coherent** if they have

- (a) a constant phase relationship (i.e. are **phase-linked**), which implies
- (b) the *same frequency*.

If the waves are transverse, then they must be either unpolarized, or polarized in the same plane.

What happens when these conditions are not observed is discussed on p. 297.

The interference at a *particular point* is said to be

- (a) **constructive** when the wave disturbances reinforce each other, or
- (b) **destructive** when the disturbances are self-cancelling.

Except where one wave is superposed on another, there is no trace of the other's presence.

Since (amplitude)² \propto intensity, the energy associated with a wave takes no account of phase. Thus for surface water waves the same energy is associated with a double trough as with a double crest.

Description of Young's experiment

Fig. 14.4 represents two-source interference for *any type* of wave motion: the reader should be able to associate it with water waves in a ripple tank experiment.

An arbitrary point X will lie on

(a) a **nodal** line (cancellation) if

$$(AX - BX) = (2m + 1) \frac{\lambda}{2}$$

since the disturbance will then be *permanently zero*,

(b) an **antinodal** line (reinforcement) if

$$(AX - BX) = (2m) \frac{\lambda}{2}$$

The disturbance at X will vary with time: the amplitude will be double that produced by A or B separately. (In each case m is an integer.)

There is a *net translation of energy* away from the sources along the antinodal lines (fig. 14.5).

A 'photograph' of the wave pattern would show that neighbouring antinodal lines consist of travelling waves in antiphase.

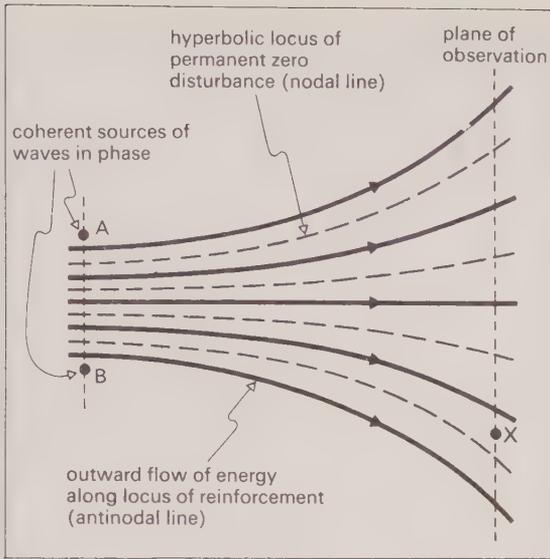


Fig. 14.4. A time-average of the interference pattern obtained in a Young's experiment.

The student should work out what happens to the intensity of the waves and the distance between antinodal lines (the **fringe separation**) when the plane of observation is, first, parallel to AB, and, secondly, perpendicular to AB, using

- (i) different source to plane-of-observation distances,
- (ii) A at different distances from B,
- (iii) sources A and B with various phase relationships (such as a phase difference of π rad),
- (iv) sources A and B of higher or lower frequencies,
- (v) a source A of different frequency from that of B.

► 14.5 ANALYTICAL TREATMENT OF YOUNG'S EXPERIMENT

In this section we will calculate the variation of intensity with position on a screen when the screen is placed parallel to AB.

Suppose the two waves (of equal amplitude) are represented by

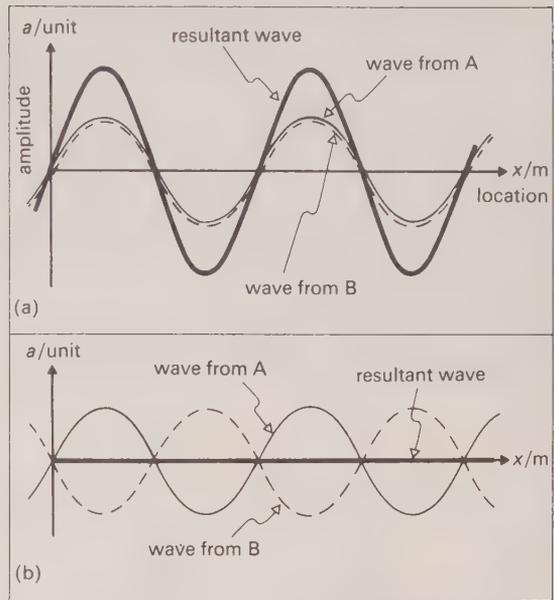


Fig. 14.5. Diagram (drawn for transverse waves) showing the result of superposition (a) along an antinodal line, (b) along a nodal line, for one instant of time.

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

$$y_2 = a \sin(\omega t - kx - \phi)$$

where ϕ is the phase difference between the disturbances at a particular point (produced in this example because the waves come by different routes).

By the superposition principle

$$\begin{aligned} y &= y_1 + y_2 \\ &= 2a \cos \frac{\phi}{2} \sin \left(\omega t - kx - \frac{\phi}{2} \right) \end{aligned}$$

The $[2a \cos(\phi/2)]$ term shows the variations of y 's amplitude with ϕ , i.e. with position on the screen.

The $\sin[\omega t - kx - (\phi/2)]$ term shows how y varies with t for each value of x (the distance from the plane through AB to the plane of the screen).

Intensity variation

$$\cos \frac{\phi}{2} = 0 \quad \text{when} \quad \phi = (2m + 1)\pi$$

Then $y_{\max} = 0$

$$\cos \frac{\phi}{2} = \pm 1 \quad \text{when} \quad \phi = (2m)\pi$$

Then $y_{\max} = 2a$

The amplitude (not displacement) = $2a \cos(\phi/2)$,
so

$$\text{intensity} \propto 4a^2 \cos^2 \frac{\phi}{2}$$

$$I \propto \cos^2 \frac{\phi}{2}$$

This is illustrated by fig. 14.6. Young's fringes are sometimes called 'cos² fringes'.

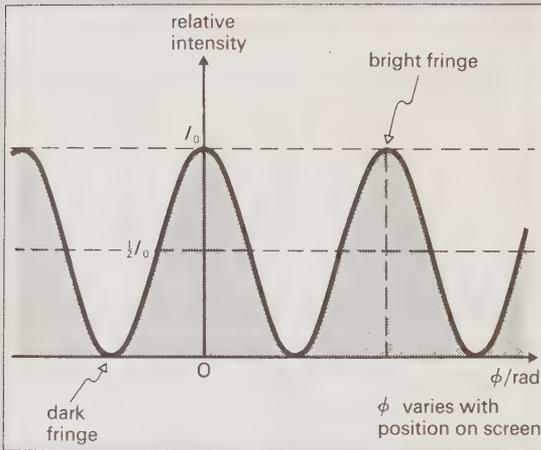


Fig. 14.6. Variation of intensity in a typical interference pattern.

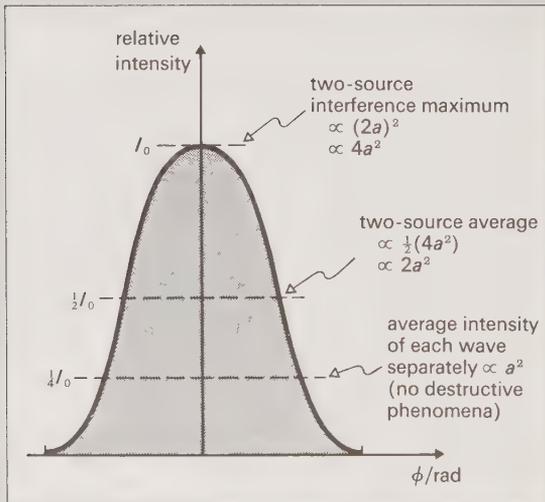


Fig. 14.7. Energy for two sources separately and taken together.

Energy conservation

Fig. 14.7 suggests how energy is conserved.

Intensity and fringe width

Suppose ϕ is the result of a path difference p between waves that originate from sources in phase.

Then
$$\frac{\phi}{2\pi} = \frac{p}{\lambda}$$

$$\therefore \phi = \frac{2\pi p}{\lambda}$$

and
$$I \propto \cos^2 \left(\frac{\pi p}{\lambda} \right)$$

It is shown on p. 295 that

$$p = \frac{ds}{D} \text{ under certain conditions,}$$

where d is the distance along the screen from the central maximum, s is the source separation, and D is the screen-source distance.

Thus
$$I \propto \cos^2 \left(\frac{\pi ds}{\lambda D} \right)$$

Maxima will occur where

$$\cos^2 \left(\frac{\pi ds}{\lambda D} \right) = 1$$

so
$$\frac{\pi ds}{\lambda D} = m\pi$$

and
$$d = \frac{m\lambda D}{s}$$

Thus the fringe width w is given by

$$d_{m+1} - d_m = \frac{\lambda D}{s}$$

$$w = \frac{\lambda D}{s}$$

Note the conditions (p. 295) under which we can use this expression: it *cannot*, for example, be used in ripple-tank and microwave experiments. It is illustrated by fig. 14.8.

14.6 BEATS

Suppose two wavetrains of slightly different frequencies f_1 and f_2 ($f_1 > f_2$), but of equal amplitudes

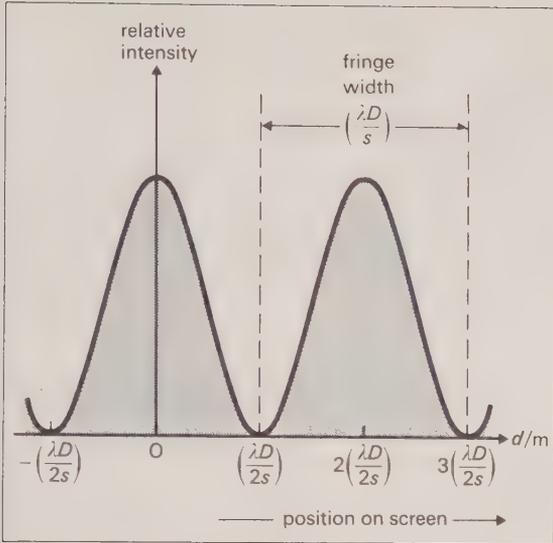


Fig. 14.8. Fringe maxima related to location on the screen.

are superposed while travelling in the same direction at the same speed.

Consider their superposition over a period of time at a fixed point. Fig. 14.9 shows how the result is sometimes reinforcement, and sometimes cancellation, because the sources are not coherent. Fig. 14.10 (overleaf) shows the result of applying the principle of superposition.

Notes

(a) The amplitude varies with time with a frequency

$$\left(\frac{f_1 - f_2}{2}\right)$$

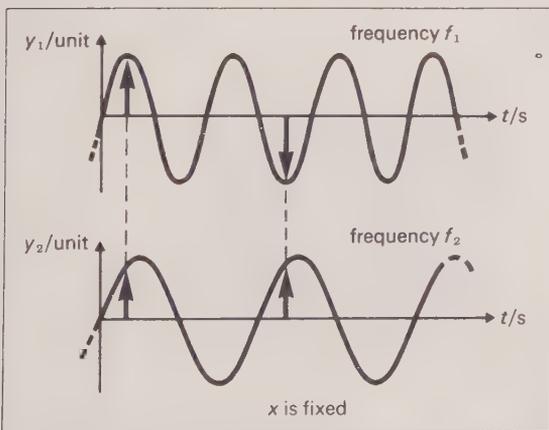


Fig. 14.9. Two wavetrains which give beats when superposed.

(b) The **beat frequency** f_{beat} is the number of times the magnitude of the amplitude reaches a maximum in each second.

$$\begin{aligned} \text{Thus } f_{\text{beat}} &= 2 \times (\text{amplitude frequency}) \\ &= (f_1 - f_2) \end{aligned}$$

More generally

$$f_{\text{beat}} = (f_1 \sim f_2)$$

This result is proved below, and in the next section.

(c) The resultant wave has a (displacement) frequency

$$f = \frac{f_1 + f_2}{2}$$

the average of the superposed waves. Usually f_1 is close to f_2 : then f is roughly equal to either.

Calculation of the beat frequency

Suppose that in time T (the **beat period**), the wave of frequency f_1 completes one cycle more than wave f_2 .

The number of cycles they complete are f_1T and f_2T respectively.

$$\therefore f_1T - f_2T = 1$$

So

$$f_{\text{beat}} = \frac{1}{T} = f_1 - f_2$$

Beats in sound

The ear responds to the energy (intensity) variation.

$$\text{Thus } (\text{loudness}) \propto (\text{amplitude})^2$$

The ear does not distinguish the two halves of fig. 14.10, and hears *two* sounds of maximum intensity in each cycle of period $2/(f_1 - f_2)$. Thus the loudness variation has a frequency $(f_1 - f_2)$.

Beats in sound are discussed further on p. 327.

► 14.7 ANALYTICAL TREATMENT OF BEATS

We aim to find the variation of y with t while x is constant. It is *not* a special case if, for convenience, we choose $x = 0$. (The time variations elsewhere are of the same form.)

Using $y = a \sin(\omega t - kx)$ where $\omega = 2\pi f$, we wish to superpose

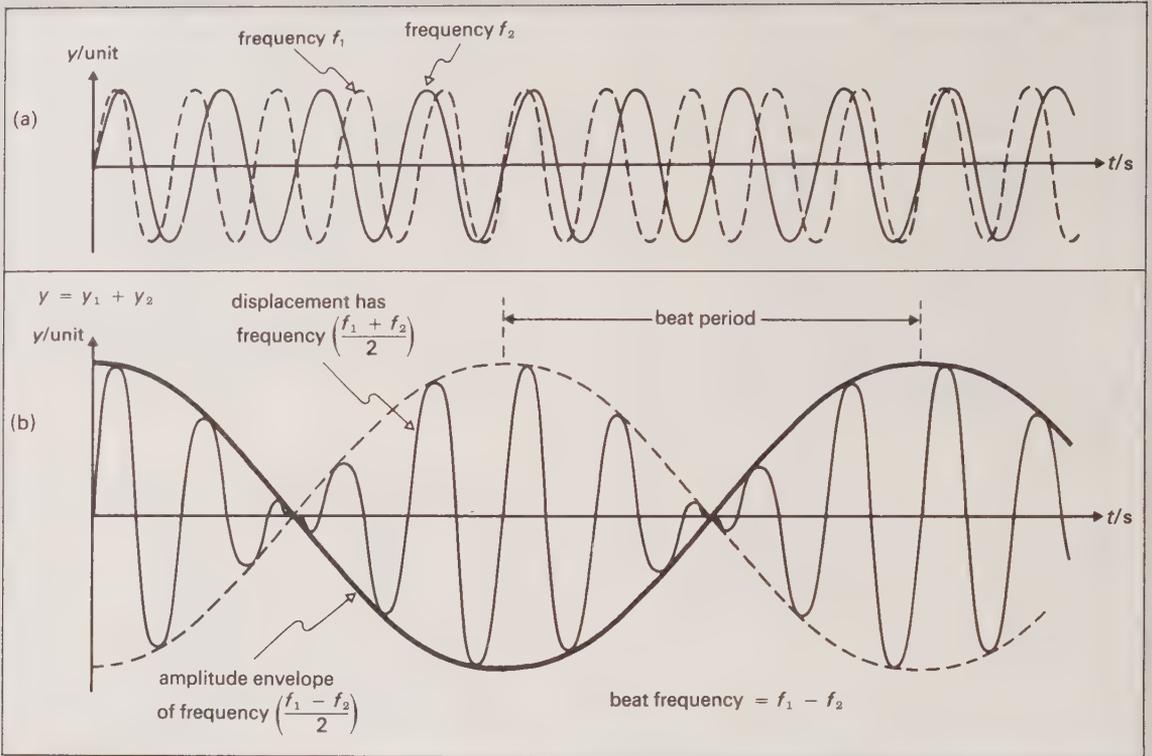


Fig. 14.10. Displacement–time curves for (a) frequencies f_1 and f_2 and (b) their resultant which shows beats.

$$y_1 = a \sin 2\pi f_1 t$$

and

$$y_2 = a \sin 2\pi f_2 t$$

So $y = y_1 + y_2$

$$= a(\sin 2\pi f_1 t + \sin 2\pi f_2 t)$$

$$= \left[2a \cos 2\pi \left(\frac{f_1 - f_2}{2} \right) t \right] \sin 2\pi \left(\frac{f_1 + f_2}{2} \right) t$$

We conclude that the resultant disturbance $y =$

$$\left\{ \begin{array}{l} \text{an amplitude which} \\ \text{varies at frequency} \\ \left(\frac{f_1 - f_2}{2} \right) \\ \text{i.e.} \\ \text{slowly in time} \end{array} \right\} \times \left\{ \begin{array}{l} \text{a displacement} \\ \text{which has frequency} \\ \left(\frac{f_1 + f_2}{2} \right) \\ \text{i.e. varies} \\ \text{rapidly in time} \end{array} \right\}$$

This time variation of y is shown in fig. 14.10.

14.8 DIFFRACTION

Diffraction is the name given to the phenomena occurring when a wave motion passes an obstacle,

or passes through an aperture. The change of wave velocity, or its spreading out, results from the superposition of wavelets that originate from points along the same wavefront.

Suppose the arrangement of fig. 14.11 is established in a ripple tank.

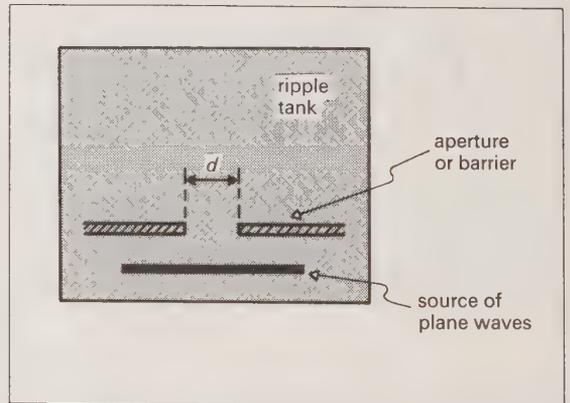


Fig. 14.11. Ripple tank experiment to investigate diffraction.

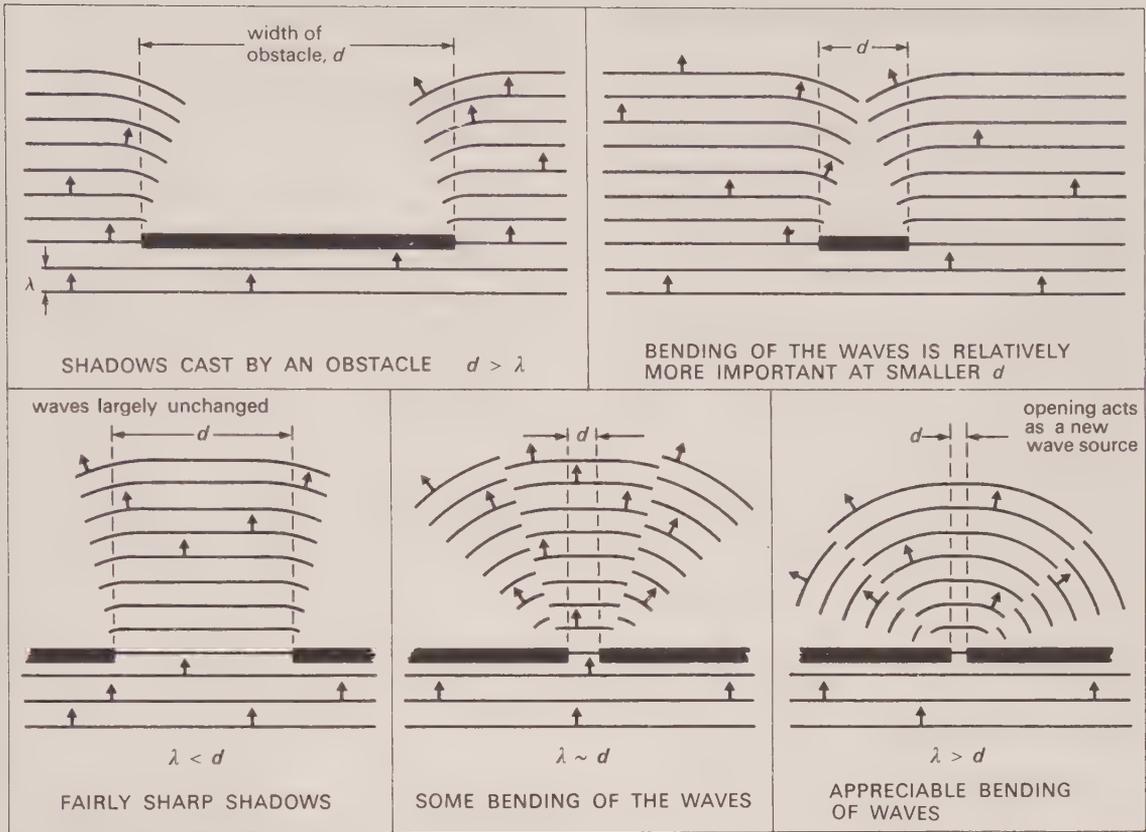


Fig. 14.12. Results of diffraction experiment.

One can adopt one of two procedures:

(a) vary d while using waves of fixed wavelength,

or

(b) vary λ while using an aperture or obstacle of fixed size.

Some results are given in fig. 14.12. These are simplified diagrams: in practice the pattern is complicated by subsidiary maxima and minima (p. 309).

Conclusions

(a) Diffraction is relatively more important when λ becomes large in comparison with d .

(b) When d is very small, the wavelets are nearly semicircular (or, in three dimensions, hemispherical): this corresponds to the idea of a *Huygens* wavelet.

Application: the **diffraction grating**, the principles of which are discussed on p. 304. Such gratings

can be made for all wave-types—the wavelength controls the choice of grating spacing.

Diffraction in light and sound

(A) Sound

The wavelength of sound for a man's voice is about 2 m, and common objects (such as furniture) have smaller linear dimensions than this.

Diffraction (and reflection) enables us to hear sounds from hidden sources.

(B) Light

λ for visible light $\sim 10^{-7}$ m which is a small fraction of the diameter of (say) a window. Diffraction effects are not frequently observed unless deliberately promoted. They are nevertheless very important, and are fully discussed in chapter 39, p. 303.

14.9 THE IMPACT OF WAVES ON OBSTACLES

Consider two extreme cases.

(A) Object large relative to λ

Then fig. 14.12 indicates two of the three effects:

- (1) a shadow is produced behind the obstacle,
- (2) the wave bends round the corners,
- (3) a reflected wave moves back from the obstacle.

For example, under the correct conditions a ball bearing causes light waves from a point source to leave a circular shadow when they are reflected,

but there is also the possibility of a bright spot on the axis (p. 307).

(B) Object very small relative to λ

Two processes operate:

- (1) Diffraction enables the wavefront effectively to reform beyond the object.
- (2) Scattering causes the object to act as a new source of secondary wavelets.

This second process accounts for the colour of the sky. Small particles in the atmosphere scatter blue light (small λ) much more than they do red light (large λ).

15 DOPPLER EFFECT

15.1	CALCULATION OF FREQUENCY CHANGE	119
15.2	EXAMPLES IN PRACTICE	121

▶ 15.3	THE DOPPLER EFFECT IN LIGHT	122
▶ 15.4	SHOCK WAVES	122

15.1 CALCULATION OF FREQUENCY CHANGE

When there is movement of source or observer relative to the medium in which a wave is propagated, then although the wave speed c through the medium is *not* changed, nevertheless the frequency f_o measured by an observer differs from f_s , that emitted by the source.

This is called the **Doppler effect** (1842).

Notation: c for wave speed,
 v for source or observer speed,
 u for speed of medium.

(A) Observer moving

Refer to fig. 15.1. The basic effect here is a change in the number of wave crests received per second. The wavelength λ is unaltered.

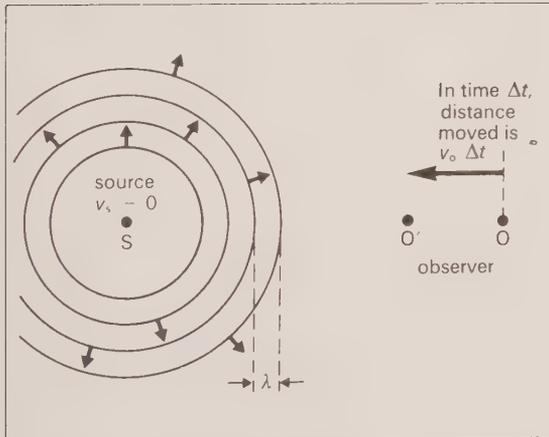


Fig. 15.1. The Doppler effect for a moving observer (relative approach).

Consider a time interval of Δt .

$$f_o = f_s + \left(\frac{\text{number of waves in distance travelled in } \Delta t}{\Delta t} \right)$$

$$= f_s + \left(\frac{v_o \Delta t / \lambda}{\Delta t} \right) = f_s + \left(\frac{v_o}{\lambda} \right)$$

but $f_s \lambda = c$,

$$\text{so } f_o = f_s + v_o \frac{f_s}{c}$$

$$= f_s \left(1 + \frac{v_o}{c} \right) \quad \text{Observer approaches source}$$

$$\text{or } f_o = f_s \left(1 - \frac{v_o}{c} \right) \quad \text{Observer recedes from source}$$

Then

$$f_o = f_s \left(1 \pm \frac{v_o}{c} \right)$$

The equation applies for a receding $v_o < c$. When $v_o > c$, no signal is received.

(B) Source moving

Fig. 15.2 (overleaf) might represent a source moving through a still medium (e.g. a trailing ripple tank vibrator) or a medium moving past a still source (e.g. a paddling swan on a river): *relative movement* is important.

The source movement changes the wave speed relative to the source, which causes

(a) a concentration of wave crests (and consequent decrease in wavelength) in front of the source,

(b) a corresponding increase in wavelength behind.

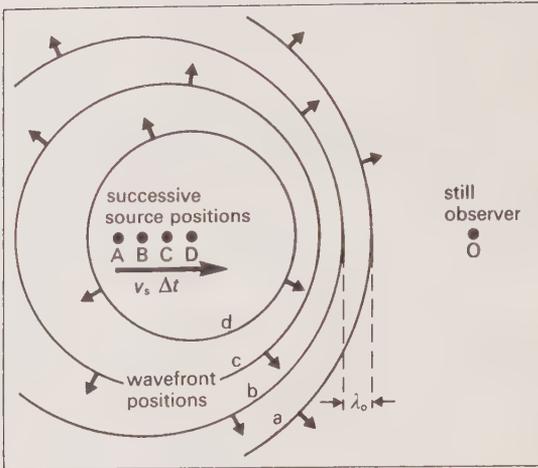


Fig. 15.2. The Doppler effect for a moving source (relative approach).

The basic effect now is a *change of wavelength*. Consider a time interval of Δt .

$f_s \Delta t$ waves are compressed into a length $(c - v_s) \Delta t$.

Hence the apparent wavelength

$$\lambda_o = \frac{\text{distance}}{\text{number of wave crests}} = \frac{(c - v_s) \Delta t}{f_s \Delta t}$$

So $f_o = \frac{c}{\lambda_o} = f_s \left(\frac{c}{c - v_s} \right)$

$= f_s \left(\frac{1}{1 - (v_s/c)} \right)$ Source approaches observer

or $f_o = f_s \left(\frac{1}{1 + (v_s/c)} \right)$ Source recedes from observer

$$f_o = f_s \left(\frac{1}{1 \mp v_s/c} \right)$$

The equation applies for an approach $v_s < c$ (see p. 122).

(C) Medium moving

(a) For sound the effective velocity of the wave motion relative to the source or observer is changed from c to $(c \pm u)$, where u is the resolved part of the medium's velocity along the line joining source and observer.

(b) For light the problem is complicated by the fact that no medium is involved. (See p. 122.)

(D) Source and observer moving

Suppose both movements cause relative approach. Since the source moves,

$$\lambda_o = \frac{c - v_s}{f_s}$$

Since the observer moves in addition,

$$f_o = \frac{c}{\lambda_o} + \frac{v_o}{\lambda_o} = \left(\frac{c + v_o}{c - v_s} \right) f_s$$

(E) Summary

In general

$$f_o = \left(\frac{c \pm v_o}{c \mp v_s} \right) f_s$$

in which upper signs indicate relative approach, and lower signs relative recession.

For sound

$$f_o = \left(\frac{\text{relative speed of sound w.r.t. observer}}{\text{relative speed of sound w.r.t. source}} \right) f_s$$

This takes into account the velocity of the medium.

An approximate relationship

Suppose $v_o = 0$. Then for approach of source

$$f_o = \frac{f_s}{1 - v_s/c}$$

$$\therefore f_o - f_s \left(\frac{v_s}{c} \right) = f_s$$

$$(f_o - f_s) = \Delta f = \left(\frac{v_s}{c} \right) f$$

where $f \approx f_o \approx f_s$ if Δf is small.

Since $f\lambda = c$, and c is constant, differentiating gives

$$f \Delta \lambda + \lambda \Delta f = 0$$

$$\frac{\Delta f}{f} = - \frac{\Delta \lambda}{\lambda}$$

$$\frac{\Delta f}{f} \approx - \frac{\Delta \lambda}{\lambda} \approx \frac{v_s}{c}$$

This is a useful non-relativistic expression which is applicable when $v_s \ll c$. It is therefore particularly useful for electromagnetic waves.

Special cases

Suppose	Then if	f_o becomes
$v_o = 0$	$v_s = c$ (approach)	∞ (shock wave)
	$v_s = c$ (recession)	$f_s/2$
$v_s = 0$	$v_o = c$ (approach)	$2f_s$
	$v_o = c$ (recession)	0 (no signal)

The general ideas of this section apply to all types of wave motion: in detail they apply to waves through material media.

15.2 EXAMPLES IN PRACTICE

(A) Train passing through a station

The value of v_s to be used is the resolved part of the train's velocity along the line joining source to observer. Then f_o varies from

$$f_s \left(\frac{1}{1 - (v_s/c)} \right)$$

through f_s to

$$f_s \left(\frac{1}{1 + (v_s/c)} \right)$$

as the train recedes.

The frequency change is most abrupt as the train passes.

(B) Radar

Detection

The source and receiver are together, the receiver being tuned for frequencies other than f_s . The distance to the moving reflector is found from the time delay before the signal is received.

To measure speed

There are two problems:

(1) The moving object of speed v receives a frequency

$$f' = f_s \left(1 + \frac{v}{c} \right)$$

(2) The moving object is now a moving source emitting at frequency f' .

The detector observes a frequency

$$f_o = f' \left(\frac{1}{1 - (v/c)} \right) = f_s \left(\frac{c + v}{c - v} \right)$$

(C) Rotating bodies

(a) A body like the Sun shows equal but opposing frequency shifts for *Fraunhofer* lines received from the opposite ends of a diameter.

(b) Double stars show a similar effect. (One of the pair may not emit visible light—but the other still shows alternate red and blue shifts.)

(c) *Doppler* effect observations of Saturn's rings show that they are made of discrete bodies (not rigid discs), since different parts have different angular velocities.

(D) The red shift

A retreating light source causes $f_o < f_s$ so a particular spectral line is shifted toward the red part of the spectrum (fig. 15.3). The red shift observations are fundamental to our understanding of the universe. Considered as *Doppler* effects, the red shifts show that the universe of galaxies is expanding in all directions at a rate that increases in proportion to the distance r from the Earth. The speed of recession is given by $v = Hr$, where H is known as the **Hubble constant**.

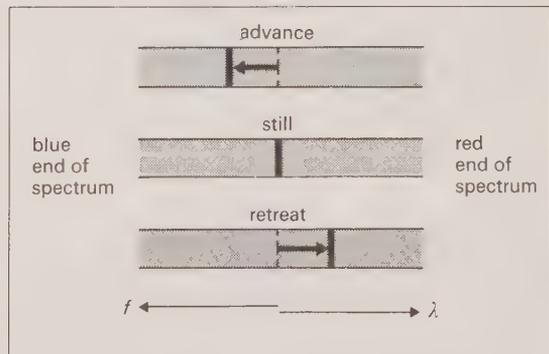


Fig. 15.3. Spectroscopic observation of the expanding universe. (According to relativistic theory, a red shift is also shown by a purely transverse motion.)

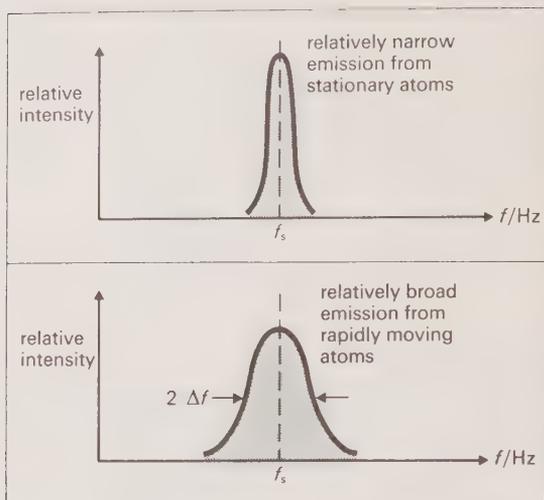


Fig. 15.4. Doppler effect broadens the spectral lines from a discharge tube.

(E) The broadening of spectral lines

Gas molecules emitting light have variable velocities resolved along the line of observation. This results (fig. 15.4) in a broadening of spectral lines.

$$\text{Since} \quad \frac{1}{2} m \overline{v^2} = \frac{3}{2} kT \quad (\text{p. 211})$$

$$\text{and} \quad \frac{\Delta f}{f} = \frac{v_s}{c},$$

high temperatures T result in large values of v_s and so of Δf .

Notes

(a) A substance should therefore be cooled before accurate spectroscopic measurements are made on emitted light.

(b) Measurement of Δf enables T to be calculated. Δf is thus a *thermometric property* which can be used to measure the temperature of a thermonuclear gas plasma when there is present a gas which is not totally ionized.

► 15.3 THE DOPPLER EFFECT IN LIGHT

Suppose an observer approaches a stationary source.

$$\text{Then} \quad f_o = f_s \left(1 + \frac{v_o}{c} \right)$$

If the source approaches a stationary observer

$$\begin{aligned} f_o &= f_s \left(\frac{1}{1 - (v_s/c)} \right) \\ &= f_s \left[1 + \left(\frac{v_s}{c} \right) + \left(\frac{v_s}{c} \right)^2 + \dots \right], \text{ etc.} \end{aligned}$$

When $v_s < c$, then in either case,

$$f_o \approx f_s \left(1 + \frac{v}{c} \right)$$

where v is their *relative velocity*.

We conclude

(a) when v is small, there is no distinction between source and observer movement, but

(b) when v is not negligible compared to c , we seem to have a method of distinguishing source movement from observer movement.

Explanation

(1) For *mechanical waves* (which require a material medium) v_s and v_o can be measured *relative to the medium*, and so can c .

(2) For *electromagnetic waves* no such medium exists, and c is the same (in a vacuum) for any observer whether he and/or the source is moving. (This is a basic postulate of the **special theory of relativity**.)

The analysis above is not applicable, and one should use

$$f_o = f_s \frac{1 + (v_R/c)}{\sqrt{[1 - (v/c)^2]}}$$

in which v_R is the radial component of the velocity of the source as determined by the observer, and v is the magnitude of the velocity. v_R is taken to be positive when the source-observer separation is decreasing.

When $(v/c)^2 \ll 1$, one is justified in using

$$f_o \approx f_s \left(1 + \frac{v_R}{c} \right)$$

for electromagnetic waves.

► 15.4 SHOCK WAVES

Suppose the observer is at rest, but that $v_s > c$. In any time interval the source moves further than the wavefront, and a cone (the **Mach cone**) encloses all the waves emitted by the source (fig. 15.5 on p. 123). A, B, C, etc., are successive source positions.

The envelope of the secondary wavelets is a cone of semi-vertical angle ϕ , where

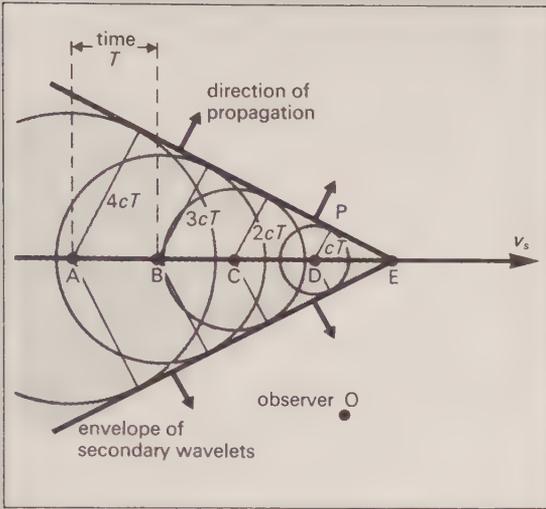


Fig. 15.5. Shock wave constructed by the Huygens method.

$$\sin \phi = \frac{DP}{DE} = \frac{cT}{v_s T} = \frac{c}{v_s}$$

$$\operatorname{cosec} \phi = \frac{v_s}{c}$$

The ratio (v_s/c) is called the **Mach number** (Ma).

For a sound wave, an observer at O hears

- (1) a shock wave from the large pressure difference across the conical wavefront (which has a high energy intensity), then
- (2) later whistling sounds emitted from distant parts.

Shock waves in practice

(A) Sound

The waves (from e.g. a bullet) can be photographed because the large pressure difference causes a significant change of refractive index.

(B) Water

The speed of a fast boat can be calculated by photographing the conical envelope of the bow wave pattern, and measuring ϕ .

(C) Light

When an electrically charged particle moves through a medium at high speed $v_s > c_n$, the speed of light in that medium, a bluish light (**Čerenkov radiation**) is emitted. Čerenkov radiation can be used to measure the speed of the particle.

IV STRUCTURE AND MECHANICAL PROPERTIES OF MATTER

16	STRUCTURE OF MATTER	126
17	ELASTICITY	138
18	SURFACE TENSION	151
19	VISCOSITY	161
20	GRAVITATION	168

16 STRUCTURE OF MATTER

16.1	ATOMS	126
16.2	MOLECULES	127
16.3	INTERMOLECULAR FORCES AND ENERGIES	127
16.4	PHASES OF MATTER	129
16.5	PHASE CHANGES	130
▶ 16.6	THE NATURE OF BONDING IN MATTER	130

16.7	CRYSTALS AND THEIR STRUCTURES	132
16.8	AMORPHOUS OR GLASSY SOLIDS	134
16.9	POLYMERS	134
16.10	BROWNIAN MOTION	135
16.11	MEASUREMENT OF THE AVOGADRO CONSTANT N_A , AND ATOMIC AND MOLECULAR DIAMETERS	135
16.12	KINETIC THEORY	136

16.1 ATOMS

An **atom** is the smallest neutral particle that represents an *element*.

We can visualize atoms as being made up of electrons, protons and neutrons, some of whose properties are given below. (In fact there exist a very large number of other sub-atomic particles whose behaviour does not concern us here.)

Particle	Mass	Charge
electron	9.1×10^{-31} kg	-1.6×10^{-19} C
proton	1.7×10^{-27} kg	$+1.6 \times 10^{-19}$ C
neutron	1.7×10^{-27} kg	0

$$\frac{m_p}{m_e} \approx 1840$$

Atomic diameters lie in the range 1 to 3×10^{-10} m.

Nuclear diameters lie in the range 1 to 7×10^{-15} m.

Every neutral atom, of atomic number Z , has Z electrons grouped round the nucleus in a series of **shells** and **sub-shells**, each of which has a definite energy level and a maximum number of electrons that it can accommodate. The number of electrons in the outermost shell determines the chemical behaviour of an atom because they are less tightly bound to the nucleus. The **valency** of an element is the number of outer-shell electrons involved in chemical reactions and these electrons are called

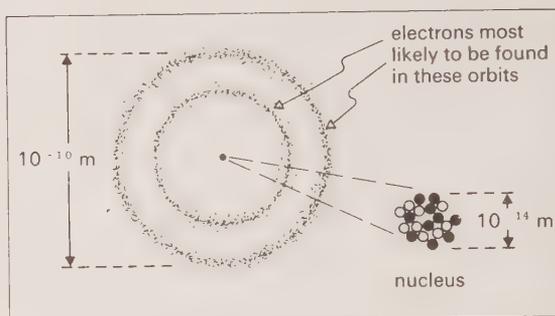


Fig. 16.1. One model of the fluorine atom $^{19}_9\text{F}$ (not drawn to scale).

valence electrons. An **ion** is any charged particle, but usually it represents an atom or group of atoms that has lost or gained one or more electrons from its outermost shell. To a good approximation, ions and atoms can be regarded as hard spheres.

Most *gases* are ionized to a small extent. Their ionization can be increased by

- heating,
- fast-moving charged particles (such as those emitted by radioactive substances),
- exposure to electromagnetic radiation of high energy (e.g. X-rays and γ -rays).

In aqueous *solution*, all acids, bases and salts are either partly or wholly ionized. The fact that water has a high electric permittivity (p. 366) helps to bring this about.

Many *solid* materials are ionized (p. 130).

Metallic atoms and ions

In metals, the atomic radius is always greater than the ionic radius of the same element. The atomic radii do not show significant variation with atomic number and so, in general, the density of metals must increase as the atomic number increases.

16.2 MOLECULES

A **molecule** of a substance is the smallest particle of it which can exist under normal conditions. For example

- (a) an argon molecule is an argon atom,
- (b) an oxygen molecule is a pair of oxygen atoms,
- (c) fig. 16.2 shows a model of a water molecule in the gaseous phase.

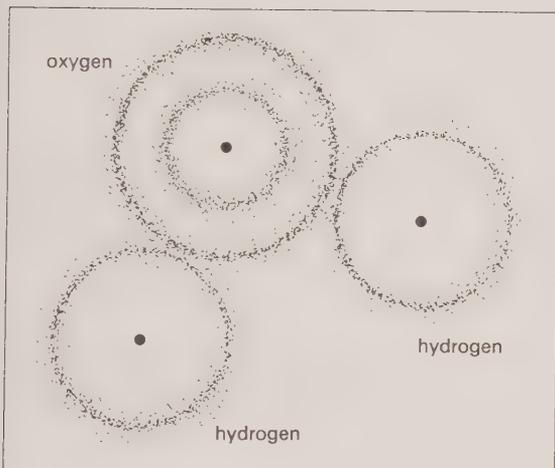


Fig. 16.2. A model of a water molecule H_2O .

For several purposes we can imagine a molecule as though it were *spherical*, as we will do in this section. On occasions we must be more exact: on p. 211 and elsewhere we treat diatomic molecules as though they were *dumb-bell-shaped*.

The idea of a molecule frequently breaks down. Some solids consist of ions and electrons, and it may not be possible to attribute a particular ion to a particular molecule (p. 130). The word *molecule* is frequently used loosely when referring to these ions.

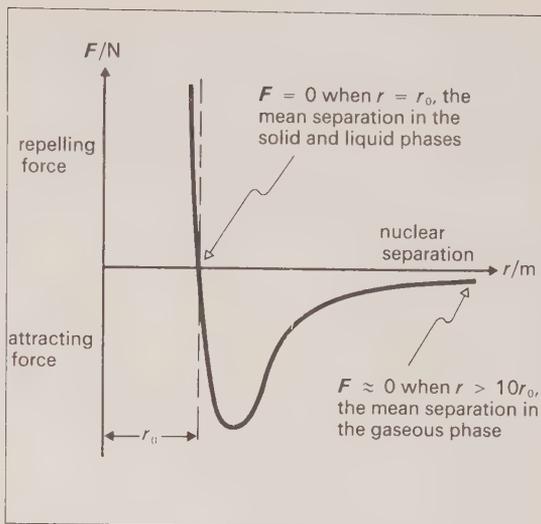


Fig. 16.3. Variation with separation r of the forces F which molecules exert on each other.

16.3 INTERMOLECULAR FORCES AND ENERGIES

Forces

The four *interactions* of nature, which together are the sole cause of all the forces that exist, are discussed on p. 3. It should be stressed that the **electromagnetic interaction**, and *not* the gravitational interaction, is responsible for *intermolecular forces*. (The gravitational forces are too weak to be significant on this scale.) A full understanding can only be obtained from wave mechanics.

Fig. 16.3 contains a great deal of information and should be studied carefully.

A simple explanation

Imagine the molecules of fig. 16.4 (overleaf) to be

- (1) Distant: they are electrically neutral, and there are no electromagnetic forces.
- (2) Closer: molecular distortion results in attraction.
- (3) Closer still: interpenetration of electron shells introduces a repulsive force. The resultant force becomes zero (equilibrium).
- (4) Too close: there is a resultant repulsive force.

Any displacement from the equilibrium position results in attraction or repulsion and thus a continuing *vibration* of amplitude about 10^{-11} m. The vibration is accompanied by a periodic interchange of k.e. and p.e., as each molecule is a simple harmonic oscillator.

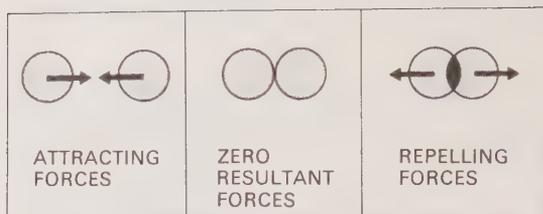


Fig. 16.4. Intermolecular forces for (model) spherical molecules.

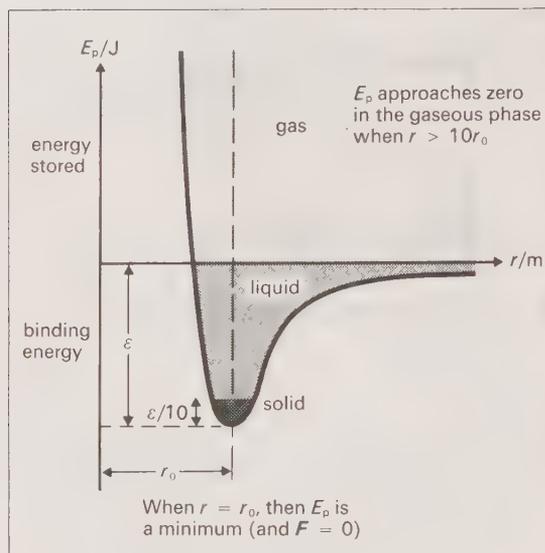


Fig. 16.5. Variation with separation r of the potential energy E_p of a system consisting of a pair of molecules.

Energies

We can classify the phases of matter according to the relative sizes of the **intermolecular attraction energy** ϵ (fig. 16.5) and the mean thermal energy of the molecules (which is a function of the temperature T):

(a) When the thermal energy is a minimum, then r is very close to r_0 . As the thermal energy increases, then so long as it is less than about $\epsilon/10$, the molecular pattern remains *ordered* and we have a **solid**. The thermal motion is restricted to a *vibrational* motion about a *fixed* mean position.

(b) As the thermal energy is increased to about $\epsilon/10$, some bonds are broken and the molecules escape from their fixed positions, as they can now exchange places with their neighbours. Their motion is no longer entirely vibrational (as clusters of molecules migrate) and melting occurs: we have a **liquid**.

(c) When the thermal energy exceeds ϵ , a molecule has enough energy to leave a parent cluster, and escapes with some translational kinetic energy; we now have a **gas**.

The way in which we evaluate ϵ is discussed in section 16.5.

► The effective range of a force

Consider a particle interacting with other identical particles by which it is surrounded. If we exclude short distances the number of interacting particles between the range r and $r + \delta r$ is $\propto r^2$.

Suppose that the potential energy of the interaction falls off as $k \cdot r^{-n}$ (again excluding very small distances), where k is a constant.

(a) If $n > 2$, then the total interaction p.e. between the particle considered and those lying between r and $r + \delta r$ is $\propto r^{2-n}$, and this *decreases* as r increases. Since the particle interacts mostly with near neighbours, the force is said to be **short-ranged**.

(b) If $n < 2$, then the same argument shows that the p.e. of interaction *increases* as r increases (provided the medium continues). The force is said to be **long-ranged**.

Examples

(a) When we are dealing with short-ranged forces the scale is relatively unimportant: thus, the specific latent heat of vaporization is the same for two samples of the same liquid, even if they have different volumes. Since l is determined by the intermolecular forces, it follows that they must be short-ranged.

(b) The properties of material bound by long-ranged forces depend radically on scale, and this is illustrated by the behaviour of stars and planets of different size.

Rates of reaction

In some chemical reactions, molecular combination involves the initial formation of a high-energy complex which breaks down into the final products. If E_A is the **molar activation energy** required,

$$\text{the rate of reaction, } k \propto e^{-E_A/RT}$$

where T is the temperature.

The expression $e^{-E_A/RT}$ is a measure of the proportion of activated-complex molecules compared with reagent molecules. The graph of $\ln k$ against $1/T$ would give a straight line with a negative gradient equal to $-E_A/R$. A high value of E_A would produce a steep gradient and a reaction sensitive to temperature change, although it would proceed slowly.

Phase	Macroscopic behaviour	Molecular order and behaviour	Mean molecular separation	Number of molecules per m ³	Volume of molecules relative to that of substance	Mean number of 'collisions' per second at s.t.p.	Thermal energy E of the molecules
gas near s.t.p.	has variable shape and volume	totally random translational motion	$10r_0$	10^{25}	10^{-3}	10^{10}	$E > \epsilon$
liquid	has variable shape but volume is nearly constant	molecules vibrate within ordered clusters which have some restricted translational motion	r_0	10^{28}	1	10^{14}	$\epsilon > E > \frac{\epsilon}{10}$
solid	both shape and volume are nearly constant	molecules vibrate about a fixed site in the lattice	r_0	10^{28}	1	10^{14}	$E < \frac{\epsilon}{10}$

16.4 PHASES OF MATTER

There are three principal phases of matter whose characteristics are compared in the table above.

(A) The gaseous phase

Intermolecular forces are short-ranged forces (they are only effective at a range of a few molecular diameters). Since the average molecular separation is about $10r_0$, the molecules only interact significantly on impact or very close approach. Different molecules at any instant have different speeds, according to the *Maxwellian distribution* (p. 198), but a typical average random speed is 10^2 m s^{-1} . (A quantitative account of the dynamics of gas molecules is given on p. 199.)

A **plasma** is made by heating gas atoms and molecules to a high temperature. At a temperature of (say) 10^4 K , some of the originally neutral particles will have acquired enough thermal energy (about 10^{-18} J) to cause them to ionize. The plasma consists of the resulting mixture of neutral particles, positive ions and negative ions (many of which are electrons).

(B) The liquid phase

The molecules *vibrate* about a *moving* centre of vibration, because thermal motion causes a con-

stant change of relative positions. There are *no* long-ranged linkages, and although there is *no* long-ranged order, the molecules gather in clusters.

(C) The solid phase

A solid is characterized by a fixed shape (which means that it has a modulus of rigidity) as it can *permanently* resist a shear stress (p. 140). This is because each molecule or ion has a permanent linkage with its neighbours, and so retains a fixed position in the overall pattern (there *is* long-ranged order).

The molecular or ionic pattern can be one of two types:

(a) **Ordered patterns** exist in crystalline substances, and have a periodic repetition of the **unit cell** (specified group of particles) in three dimensions.

(b) **Disordered** arrangements exist in supercooled liquids, such as glass. They are less stable, and therefore rare.

The ordered pattern has the lower potential energy, and so substances tend to be crystalline (since this is more stable).

Fig. 16.6 (overleaf) indicates the behaviour of particles in the different phases, as predicted from data fed into a computer.

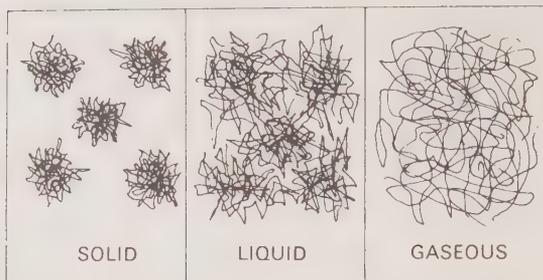


Fig. 16.6. Models of particle behaviour in solid, liquid and gaseous phases.

16.5 PHASE CHANGES

A discussion in terms of the intermolecular attraction energy ϵ is given on p. 225, while the macroscopic view leads to a definition of specific latent heat of transformation.

Condensation

If a molecule makes an *elastic* collision with a surface, it is bound to escape. If the collision is *inelastic*, the molecule may lose sufficient translational k.e. to prevent subsequent escape. The probability that this will happen increases as the temperature decreases (since the molecular k.e. is then reduced), and molecules will then be attracted into a vibratory motion on the surface.

Solidification

As a liquid is cooled, the molecular vibrational energy is reduced until the molecules can no longer change relative positions. Removal of energy (the latent heat) has converted the disordered pattern into an ordered one, and the potential energy of the new structure is smaller.

When the solid is heated the breaking up of the ordered lattice can be detected near the melting point by a sharp change of slope in the curve which plots density against temperature. This indicates an increase in the number of sites in the lattice which are empty (these are called **vacancies**).

Sublimation

This is the process that occurs when, on gaining heat, the particles of a solid evaporate directly without passing through the intermediate liquid phase.

The estimation of ϵ

Consider a system in which each particle makes effective contact with n nearest neighbours. n is called the **coordination number**.

(a) **Sublimation.** Assume the molecule in a solid interacts with nearest neighbours only. Then the binding energy per molecule is $\frac{1}{2}n\epsilon$. The $\frac{1}{2}$ factor ensures that we do not count each bond twice. The molar latent heat of sublimation

$$L_{s,m} = \frac{1}{2}N_A n\epsilon$$

The molar latent heat of vaporization $L_{v,m}$ would be slightly less.

(b) **Surface tension.** A similar argument shows that surface energy should be about $\frac{1}{2}L_{v,m}$, so

$$(\sigma A) \approx \frac{1}{2}(\frac{1}{2}N\epsilon)$$

where N molecules are contained in a surface area A .

The value of ϵ can be estimated from these expressions by assuming values for n and (N/A) . Its size is determined by the nature of the interparticle bond. Typical values might be 10^{-19} J for an ionic or covalent bond, but only 10^{-22} J for a *van der Waals* bond.

► 16.6 THE NATURE OF BONDING IN MATTER

In this section we discuss the types of binding interactions which cause atoms and molecules to attract at relatively large distances, and which join atoms together to form molecules and other structures.

(A) Strong bonds

These bonds result from the exchange or sharing of electrons, and occur because the bonded system has a lower p.e. than the separate atoms.

(a) In the **covalent bond** *electrons are shared* (at least in part) between different atoms. The bond energy is electrostatic in origin. In diamond (fig. 16.7) the individual carbon atoms are linked by covalent bonds to form a rigid crystal lattice.

(b) In the **ionic bond** *electrons are transferred* from one neutral atom to another, and the resulting positive and negative ions attract each other by *Coulomb* forces. In sodium chloride (fig. 16.8) Na^+ and Cl^- ions are arranged so that the ions of each type form a cubic pattern.

In some substances (such as dry HCl) the bonds are *partly ionic* and *partly covalent* (the electron largely belongs to one 'ion', but is to some extent shared by both).

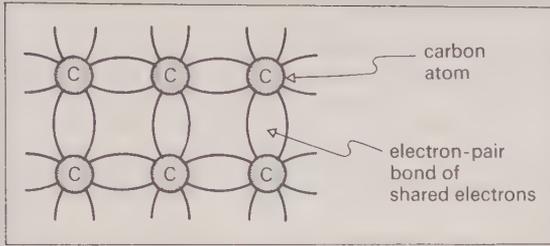


Fig. 16.7. Representation in two dimensions of the covalent bonds between carbon atoms in a diamond crystal.

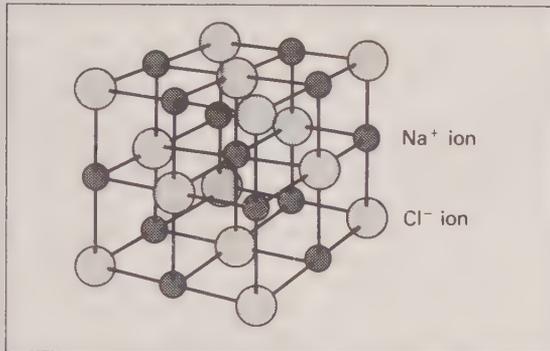


Fig. 16.8. The cubic lattice of Na⁺ and Cl⁻ ions in an NaCl crystal.

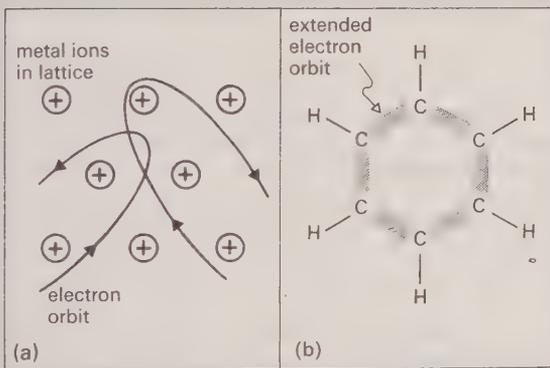


Fig. 16.9. (a) Metallic bonding and (b) bonding within the benzene molecule.

(c) **Metal bonds** and π -bonds occur where electrons have greater freedom of movement, i.e. are less obviously tied to their parent atoms than they are in (a) and (b). The lower energy which the electrons then possess enables them to contribute to binding over a larger region of the structure.

Examples

(a) The special bonding in *metals* (fig. 16.9(a)) results from the extended orbits of the conduction electrons.

(b) The stability of the *benzene molecule* (fig. 16.9(b)) comes from the extended orbits of the π -bond electrons.

(B) Weak bonds

These bonds result from the redistribution of electrons *within a molecule*, and enable it to interact with other molecules. The molecules retain their chemical identity.

Van der Waals forces

These are weak attractive forces acting over short distances outside a molecule and enable it to attract both polar and non-polar neighbours. The forces which link the chains of polymers such as polyethylene are of this type.

There are three effects which produce attraction.

(a) **Orientation effect.** Fig. 16.10(a) shows how *two permanent* electric dipoles will attract one another if their relative orientations are favourable. In practice, orientations do produce an attractive force, but its size decreases with increasing temperature as the molecular distribution becomes more random.

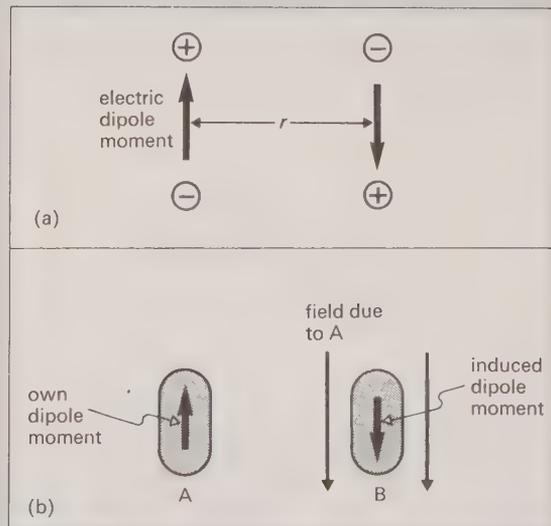


Fig. 16.10. Origin of van der Waals forces: (a) illustrating attraction between permanent electric dipoles and (b) illustrating attraction between temporary electric dipoles.

(b) *Induction effect.* One molecule which has a permanent electric dipole moment will **polarize** (distort) a neighbouring molecule, causing it to have an *induced* anti-parallel dipole moment. The attractive force which results is independent of temperature.

(c) *London effect.* This effect causes an attraction between a pair of molecules *both* of which are *non-polar*. These forces, which have been called **dispersion forces**, are for most gases more important than those resulting from the orientation and induction effects.

In fig. 16.10(b) molecule A sets up the electric field typical of a dipole because of a *random* displacement of its electron cloud. (All molecules have a fluctuating dipole moment of this kind.) Molecule B, finding itself in the field set up by A, acquires an *induced* moment (as in effect (b)).

Notes

(a) The pair of dipole moments always cause attraction, whatever the direction of A's random moment.

(b) The average moment of either molecule is zero over a time which is long relative to the orbital period of an electron.

Hydrogen bonding

Hydrogen bonds are basically electrostatic interactions between molecules whose surface regions have different electron densities. These bonds are particularly important in determining the properties of water (fig. 16.11) and the structure of biological substances such as DNA (deoxyribonucleic acid).

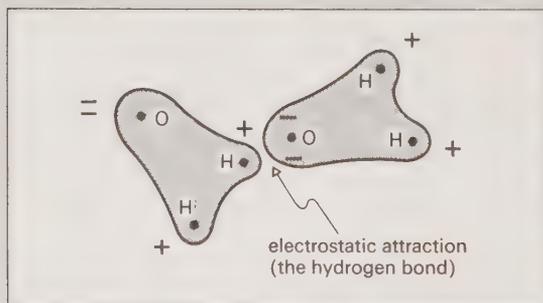


Fig. 16.11. Hydrogen bonding between water molecules.

We would expect the value of the specific heat of sublimation to be closely related to the strength of bond in the substance concerned. Experiment shows that it is highest for ionic and covalent solids, very high for most metals, but comparatively small for solids bound by *van der Waals* forces.

Repulsive forces

The interactions discussed so far in this section all lead to attractive forces, and these *attractive* forces draw atoms and molecules together. Matter does not contract indefinitely because when there is interpenetration of different atoms' electron clouds they exert very strong *repulsive* forces on each other.

16.7 CRYSTALS AND THEIR STRUCTURES

The crystalline state

The vital distinction between a crystal and an amorphous solid lies in the former's long-range order, that is its *regularly repeated* arrangement of atoms, ions or molecules. This essential inner regularity is partly reflected in the external habit, and although the shapes of a particular crystal species are not exactly reproduced from crystal to crystal, the *angle* between specified faces is.

A crystal can be made from single atoms (as in diamond), ions (as in NaCl), or molecules (as in ice). The **unit cell** of a particular lattice is the smallest and simplest unit from which the three-dimensional periodic pattern can be built. It may contain several atoms or molecules. The choice of the unit cell is not unique, but is usually made so that its symmetry corresponds to that of the crystal. Its importance to X-ray crystallography is discussed on p. 311. The physical properties of solids are determined by the nature of the unit cells and their geometrical arrangements. Crystal lattices contain *line defects* known as **dislocations** and *point defects* involving **vacancies**, **interstitial atoms** and **substitutional atoms**. The degree of imperfection is an important factor. Most crystalline solids (e.g. igneous rocks and most metals) are **polycrystalline**, consisting of numbers of small crystals (**crystallites**) or **grains**, of arbitrary shape packed closely together. In igneous rocks and metallic alloys, the grains are of different substances (e.g. granite contains grains of quartz, feldspar and mica). The grains are separated by **grain boundaries** which are regions of higher energy than the grain interiors. A metal with small grains has greater strength than the same metal with large grains because the smaller grains have a larger total area of grain boundaries, which is more effective at inhibiting the movement of dislocations. The concentration of impurities within grain boundaries and the corrosion caused there by chemical action can result in a metal being weakened. Grain boundaries and the long-range order within a grain can be demonstrated conveniently using a bubble raft.

Some crystal structures

The structure of a crystal is determined by

- (i) the kind of bond(s) between its particles, and
- (ii) the size and shape of those particles.

Thus in metals the non-directional nature of the π -bonds results in a close-packed structure, whereas a more open structure is found in covalent solids, since the bonding is directional.

(a) **Face-centred cubic** packing is shown in fig. 16.8. The cube has a particle at each corner, and in addition one at the centre of each face. (This pattern can be made in fig. 16.12 by positioning the top layer such that each of its spheres lies above a gap in the bottom layer.)

Examples: Al and Cu.

(b) **Hexagonal close-packing** is shown in fig. 16.12. The structure is built up from layers of hexagons, with particles in alternate layers in identical positions.

Examples: Mg and Zn.

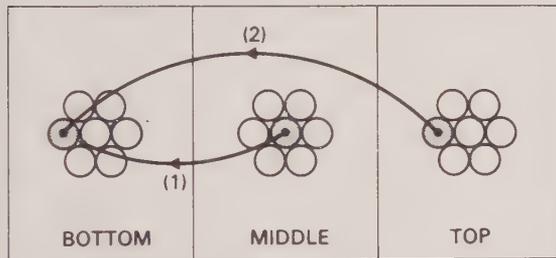


Fig. 16.12. Building spheres into a hexagonal close-packed structure: (1) place middle layer to superpose dot on dot, and (2) place top layer so as to superpose bottom layer exactly.

(c) **Body-centred cubic** packing has a particle at each corner of a cube, and one at the centre.

Examples: Fe and Na.

(d) **Tetrahedral structures** have one particle at each of the corners of a regular tetrahedron, and one at the centre. The structure is exemplified by the covalent bonding of diamond.

In each case a structure is aiming at maximum stability by minimizing the crystal's energy density. This is done by

- (i) keeping the crystal electrically neutral,
- (ii) preserving the directional nature of covalent bonds,
- (iii) minimizing the strong *Coulomb* repulsion between like ions and, as far as is consistent with these requirements, by
- (iv) packing the ions as densely as possible.

Alloys and metallic solid solutions

Alloys have more than one constituent element and possess metallic properties. Within the crystal structure, the different constituents are distributed arbitrarily and their proportions vary between certain limits. However, the atom positions are fixed as in a pure crystalline substance. This is known as a **metallic solid solution**. In an **interstitial** solid solution, small atoms can fit into some of the interstices that exist in a close-packed structure of larger atoms without distorting the lattice significantly, e.g. iron-carbon alloy (*steel*). In a **substitutional** solid solution, the atomic diameters of the two metals forming an alloy must be similar, e.g. copper-zinc alloy (*brass*). The added metal atoms can be substituted for the original metal atoms in identical lattice positions, but there is usually a solubility limit because of the lattice distortions produced and because of changes in the free electron density if the metal valencies differ. The lattice distortions in metallic solid solutions often give alloys special properties of hardness, strength and resistivity that are preferable to those of the pure metals concerned.

► The energies of crystals

A *single crystal* is an ideal solid in that its particles are in a well-defined environment, and this enables the energy to be computed relatively simply.

(a) **Ionic crystals.** The interionic forces are (i) long-ranged electrostatic (*Coulomb*) repulsion and attraction, and (ii) the shorter-ranged repulsive forces between the electron clouds. If the charges are $\pm e$, and a structure contains N ions, the total p.e. is given by

$$E_p = -\left(\frac{Ne^2}{4\pi\epsilon_0}\right)\frac{A}{x} + \frac{B}{x^9}$$

in which x is the side of the unit cell, and A and B are constants. (A is called the **Madelung constant**, a characteristic of a particular ionic structure.) The binding energy is dominated by the *Coulomb* term, and is approximately equal to it.

(b) **Molecular crystals.** Fig. 16.5 shows, for a pair of molecules, the p.e. curve resulting from the long-ranged attractive forces and the shorter-ranged repulsive forces.

The total energy E_p of the structure is the sum of the energies of each molecular pair. It can be expressed in the form

$$E_p = -\left(\frac{A}{r^6} + \frac{B}{r^{12}}\right)$$

where A and B are constants. The equilibrium value of E_p is a minimum, and so the equilibrium separation r_0 of the particles can be found by putting

$$\frac{dE_p}{dr} = 0$$

This, in turn, enables E_p to be calculated.

16.8 AMORPHOUS OR GLASSY SOLIDS

A few minerals such as opal and limonite are **amorphous** or non-crystalline and consist of particles (possessing short-range order) packed together in a disorderly arrangement, therefore possessing no long-range order. **Glasses** are common amorphous substances and they behave like solids with exceptionally high viscosity at normal temperatures. When a glass changes from the solid to the liquid state, there is no sharp melting-point, no sudden volume change and no absorption of latent heat. The solid and liquid structures are therefore very similar—the solid can be regarded as a liquid in which the disordered structure has been frozen. Some physical properties, such as linear expansivity and specific heat capacity, show a significant change over a narrow range of temperatures centred on the **glass transition temperature**, T_g . This is because there is a great difference in the diffusion rates of molecules in the liquid and solid states. **Devitrification**, which involves the change from glass to very small crystals (crystallites), can occur over long time periods in both man-made glass and natural glass such as obsidian (formed when viscous volcanic lava is cooled rapidly).

16.9 POLYMERS

These materials possess an *intermediate* state of order between the crystalline and the amorphous states. Cellulose, lignin, proteins, cotton, silk and wool are some natural polymers, while synthetic examples include perspex, polyvinyl chloride (PVC), polystyrene, synthetic rubbers, polyester resin and fibres like rayon and nylon. They consist of huge molecules (**macromolecules**), each composed of a long flexible polymer chain of vast numbers ($>10^4$) of basic atom groups called **monomers**. The simplest example, a polythene (polyethene) molecule, is an irregularly twisted and tangled chain of c. 25 000 CH_2 groups with CH_3 groups at each end, giving a total length of about $3 \mu\text{m}$. Even very weak polymer solutions have high viscosity because the chains may cut across

liquid layers travelling at different speeds, thereby causing internal tension forces.

Crystalline polymers

These contain crystalline regions where the polymer chains are parallel, as well as amorphous disordered regions. They have sharp melting points and their partial order is revealed by X-ray diffraction patterns. Polythene is a **non-oriented** example because the axes of the crystalline regions are not aligned. Pure polythenes are extensively crystalline, which results in greater hardness, greater strength and higher melting points. A fibre such as cotton is an **oriented** example because the crystalline regions are aligned approximately with the fibre axis. This alignment gives it high tensile strength which can be made greater by increasing the crystalline proportion. The amorphous regions give flexibility to the fibres which is important for textiles.

Amorphous polymers

In spite of the relatively long-range order along each chain, X-ray diffraction patterns reveal amorphous structures in these polymers showing that the chains must be tangled up at random. Polymers such as perspex and rubber change from a *glassy* state to a *rubbery* state as the temperature passes through the glass transition temperature T_g , and are called **thermoplastics**. They become rigid again on cooling below T_g . When glassy, the chains are in fixed orientations with the monomers vibrating about mean positions. When rubbery, the weak *van der Waals* forces that act between chains allow the monomers to rotate so that the orientation of the chains can alter. T_g for natural rubber is -80°C (*rubbery* at room temperature), and for perspex is 100°C (*glassy* at room temperature).

Thermosetting

Thermosetting involves mixing chemicals together and heating them strongly so that cross-links are formed between the chains. The new material cannot be softened on heating but remains hard and rigid until it finally decomposes. Rubbers (or **elastomers**) show remarkable elasticity because their long chains are linked at certain points by heat treatment involving the addition of a small proportion of sulphur (a process called **vulcanization**). These cross-over links ensure that the rubber is a solid rather than a viscous liquid like natural rubber, and they allow the tension forces to be passed between the unfolding chains when the rubber is stretched. X-ray diffraction

photographs show that unstretched rubber is amorphous. However, when it is stretched, a clear diffraction pattern can be observed showing that a significant degree of order results from the stretching. If the sulphur proportion is $>30\%$, far more links are formed between the chains so that hard ebonite is formed and the structure is said to be **reticulated**. Bakelite (the first plastic) is made by heating phenol with methanal (formaldehyde). Ebonite and bakelite are termed **thermosetting plastics**.

16.10 BROWNIAN MOTION

The concept of pressure (developed for an ideal gas in section 24.5) breaks down for very small areas because of the randomness of the molecular bombardment. This is illustrated in the **Brownian movement**.

We deduce that the smoke particles of fig. 16.13 (typically about 10^{-6} m across) are being bombarded by the molecules of the air. We cannot see the air molecules, so they must be very small. To produce the observed effects they must, if small, be moving at high speeds.

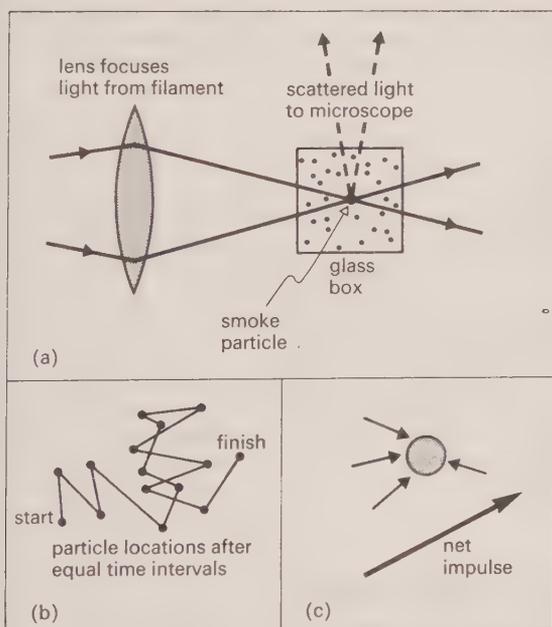


Fig. 16.13. The Brownian movement: (a) the experimental arrangement, (b) the observation, and (c) the explanation (not to scale).

Orders of magnitude

(i) A typical smoke particle might move about 10^{-6} m (its own diameter) during one second.

(ii) The size and direction of the resultant force on the particle changes significantly during a time interval of about 10^{-10} s.

According to the principle of equipartition of energy (p. 211)

$$\frac{1}{2}mc^2 \text{ (molecule)} = \frac{1}{2}MC^2 \text{ (particle)}$$

so if $M \gg m$, it follows that $\bar{c}^2 \gg \bar{C}^2$

Because these are mean square speeds the speed of a particular molecule at a particular instant may be very small.

Notes

(a) The study of the path taken by a particular particle is called the **random (drunkard's) walk problem**, and was carried out by *Einstein*. It leads to a rough method of measuring the *Avogadro constant* N_A .

(b) The average displacement of a particle starting from an origin on successive journeys is zero, even though the average distance travelled from the origin is not zero. This is because all directions of travel are equally likely.

(c) On a macroscopic scale *Brownian* motion explains

(i) the quivering movement of the suspension of a delicate mirror galvanometer, which sets a limit on the attainable sensitivity,

(ii) the random noise of an amplifier (the **Johnson noise**) caused by the varying voltages which accompany random electron movement within a metal.

Both effects are reduced by lowering the temperature.

(d) We regard *Brownian* movement as indirect experimental confirmation that fluids consist of small particles moving at high speeds.

16.11 MEASUREMENT OF THE AVOGADRO CONSTANT N_A , AND ATOMIC AND MOLECULAR DIAMETERS

(A) The Avogadro constant N_A

X-ray diffraction

The ions or atoms in crystals are arranged in regular planes separated by equal distances. This causes them to diffract X-rays in the same way that a ruled grating diffracts visible light (p. 304).

The wavelength of X-rays ($\sim 10^{-10}$ m) can be measured by a separate experiment in which they are diffracted by a ruled grating of known spacing.

The spacing of the planes in the crystal is calculated from a knowledge of the angles of diffraction and the wavelength of the X-rays, and from this N_A can be deduced simply.

Radioactivity

Radium emits α -particles whose rate of emission can be measured by a scintillation counter. The volume of helium that collects in a given time can be measured at a known temperature and pressure. This enables the mass of the helium to be calculated, and so the mass m_α of an α -particle can be found.

Since the molar mass of helium is 4.0026×10^{-3} kg mol⁻¹, N_A is calculated from

$$N_A[\text{mol}^{-1}] = \frac{4.0026 \times 10^{-3} \text{ kg mol}^{-1}}{m_\alpha [\text{kg}]}$$

The value obtained is only approximate.

Electrolysis

The **Faraday constant** F is numerically equal to the charge required to liberate one mole of singly charged ions in electrolysis, and is measured in coulombs mole⁻¹. Its value (p. 10) can be calculated from measurements on the electrical deposition of copper. Then N_A is calculated from

$$N_A = \frac{F}{e}$$

in which the electronic charge e can be measured separately by *Millikan's oil drop experiment* (p. 472). Historically, this experiment was used to measure e (rather than N_A) many years before *Millikan's oil drop experiment*. Nowadays, the diffraction method for N_A , combined with the electrolysis determination of F , gives a value of e which is many orders of magnitude more precise than *Millikan's value*.

Brownian motion

The **Boltzmann constant** k can be determined from a study of *Brownian motion*. We use it to find N_A from

$$N_A = \frac{R}{k}$$

where R is the universal molar gas constant (p. 197).

(B) Atomic diameters

These can be found by direct measurement using X-rays (above). The values obtained agree with

the orders of magnitude predicted from a knowledge of N_A combined with the densities of the substances being used.

(C) Molecular diameters

Approximate methods

(a) Let a drop of a fatty acid (such as oleic acid) whose volume is known fall gently onto a water surface dusted with lycopodium powder. The drop spreads out into a layer, the extent of which is shown by the powder. The film thickness t cannot be less than one molecular diameter, where

$$t = \frac{\text{volume of drop}}{\text{area of layer}}$$

In practice the acid is usually dissolved in alcohol, which evaporates or dissolves in the water.

(b) When mica is cleaved the steps of cleavage can be measured by an interferometer. They are found always to be multiples of a particular distance.

(c) The thickness of a soap film can be measured, and shown to be as small as 10^{-8} m.

More exact methods

(a) The molecular diameter r_0 can be estimated from measurements on *deviations from the gas laws*, and from critical data. The constant b in *van der Waals's equation* (p. 234) can be written

$$b = \frac{2}{3} \pi N_A r_0^3$$

so that if b is measured by experiment, then r_0 can be evaluated.

(b) The *kinetic theory* of gases relates measurements of viscosity, diffusion and thermal conductivity (the *transport phenomena* (p. 203)), from which it is possible to calculate r_0 .

(c) r_0 can also be estimated from the *dielectric properties* of liquids whose molecules are permanent electric dipoles.

(d) The *scanning tunnelling microscope* (STM) can resolve individual atoms in a variety of environments, including air and water, and is even capable of determining energy states on an atom-by-atom basis.

16.12 KINETIC THEORY

Kinetic theory is that branch of physics based on the assumption that matter consists of molecules. It treats the following topics, to which reference should be made for details:

(A) Change of phase

Evaporation

Vapour pressure

Latent heat

(p. 227)

(p. 227)

(p. 225)

(B) Pressure

Liquid and solid pressure

Gas pressure, and deductions from

$$p = \frac{1}{3} \rho \overline{c^2}$$

(p. 72)

(p. 199)

(C) Change of shape and volume

Elasticity

Hooke's Law

(p. 138)

(p. 139)

(D) Transport phenomena

Diffusion—transport of mass

Viscosity—transport of momentum

Electrical conduction—transport of electric charge

Thermal conduction—transport of heat

(p. 201)

(p. 203)

(p. 378)

(p. 236)

(E) Thermal expansion

(p. 187)

(F) Surface energy and surface tension (p. 152)

In each example the predictions of the theory are in satisfactory agreement with experiment. Sometimes the agreement is quantitative. Thus the assumptions of the kinetic theory are indirectly confirmed.

17 ELASTICITY

17.1	ELASTICITY, STRESS AND STRAIN	138
17.2	THE ELASTIC MODULI AND HOOKE'S LAW	138
17.3	YOUNG'S MODULUS E	139
17.4	BULK MODULUS K	140
17.5	SHEAR MODULUS G	141
17.6	MEASUREMENT OF YOUNG'S MODULUS	141
17.7	BEHAVIOUR OF A WIRE UNDER STRESS	142
17.8	PROPERTIES OF MATERIALS	142

17.9	STRESS-STRAIN GRAPHS	143
17.10	DISLOCATIONS	144
17.11	FATIGUE AND CREEP FRACTURES	145
17.12	STRESSED WIRES	146
17.13	POISSON'S RATIO μ	147
17.14	COMPOSITE MATERIALS	147
▶ 17.15	ISOTHERMAL AND ADIABATIC BULK MODULUS	148
▶ 17.16	THE TORSIONAL PENDULUM	148
17.17	BENDING MOMENT M AND SHEARING FORCE F_s	149

17.1 ELASTICITY, STRESS AND STRAIN

When a pair of opposed external forces is applied to a body, it changes the relative positions of the body's molecules. This calls restoring forces into play, and accounts for the **elasticity** of a body—its ability to return to its original form after the distorting forces have been removed.

Fluids possess only **volume elasticity**.

Solids possess in addition **rigidity**, since they can also resist a change of shape.

Solids also possess **tensile elasticity**, ability to resist longitudinal elongation and compression. (This property is a consequence of a solid having volume and rigidity elasticities.)

Stress is a measure of the cause of a deformation and is defined by

$$\text{stress} = \frac{\text{force}}{\text{area}}$$

There are three kinds of stress, each of which has dimensions $[ML^{-1}T^{-2}]$, and unit Pa ($N\ m^{-2}$).

Strain is a measure of the extent of deformation, and is defined by

$$\text{strain} = \frac{\text{change of dimension}}{\text{original dimension}}$$

There are also three kinds of strain each of which is a pure number, with neither dimensions nor a unit.

17.2 THE ELASTIC MODULI AND HOOKE'S LAW

An **elastic modulus** is defined by the equation

$$\text{modulus} = \frac{\text{stress}}{\text{strain}}$$

$$\text{modulus [Pa]} = \frac{\text{stress [Pa]}}{\text{strain [no unit]}}$$

Three kinds of modulus are defined to correspond to the three kinds of stress and strain: each has dimensions $[ML^{-1}T^{-2}]$.

Name of elastic modulus	Nature of stress	Nature of deformation
bulk modulus	change of pressure	change of size but not shape
shear modulus	$\frac{\text{tangential force}}{\text{area}}$	change of shape but not size
Young's modulus	$\left(\frac{\text{tensile or compressive force}}{\text{area}} \right)$	change (usually) of both size and shape

These moduli are discussed in the following sections, but for convenience a table of comparative values is given here.

Substance	Modulus		
	Bulk modulus K/Pa	Shear modulus G/Pa	Young's modulus E/Pa
typical metal	1×10^{11}	4×10^{10}	1×10^{11}
water	2×10^9	—	—
gas at pressure 1×10^5 Pa	1×10^5	—	—

Hooke's Law

Hooke's Law was first enunciated for the deformation of a spring. It is a macroscopic (measurable) result of the fact that the graph of fig. 16.3 is a straight line in the region r close to r_0 .

We now extend Hooke's Law to elastic deformation in other situations by writing

$$\frac{\text{stress}}{\text{strain}} = \text{constant}$$

provided the deformation is small.

Thus under some conditions we can treat the elastic moduli as constants: when Hooke's Law does not apply they remain as useful concepts which can still be used (as average values).

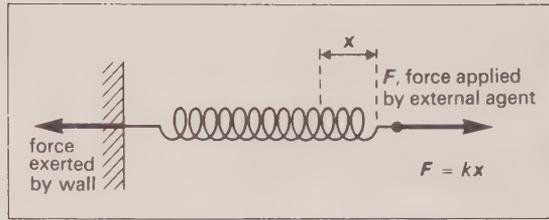


Fig. 17.1. Hooke's Law for a spring.

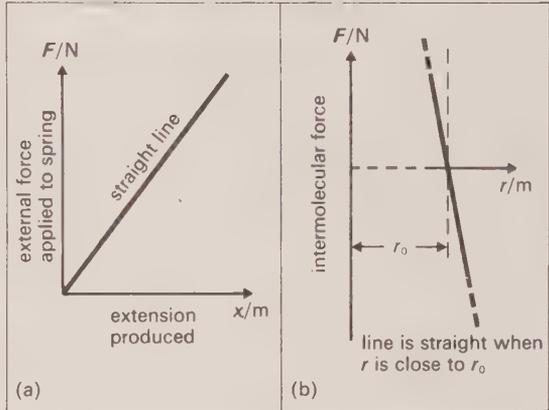


Fig. 17.2. (a) Macroscopic representation, and (b) microscopic cause, of Hooke's Law.

It is important to note that the behaviour described by Hooke's Law makes it possible for deformed bodies to vibrate with s.h.m.

17.3 YOUNG'S MODULUS E

Refer to fig. 17.3 (overleaf).

$$\text{Stress } \sigma = \frac{\text{tensile force}}{\text{area of cross-section}} = \frac{F}{A}$$

$$\text{Strain } \epsilon = \frac{\text{change of length}}{\text{original length}} = \frac{\Delta l}{l_0}$$

Young's modulus E is defined by the equation

$$E = \frac{\sigma}{\epsilon} = \frac{F/A}{\Delta l/l_0}$$

Consider now the *microscopic* scale. Let the applied force F increase the mean separation of adjacent planes of atoms by Δr in the direction of F . If k is the interatomic force constant, the restoring force experienced by each displaced atom is given by $F = k \Delta r$. As there are $1/r_0^2$ atoms m^{-2} , the stress to separate the planes = $k \Delta r/r_0^2$.

The strain = $\frac{\Delta r}{r_0}$ and

$$E = \frac{k \Delta r/r_0^2}{\Delta r/r_0} = \frac{k}{r_0}$$

$$E = \frac{k}{r_0}$$

Notes

(a) k is the gradient, dF/dr , where the graph of fig. 17.2(b) passes through r_0 .

(b) For metals, $E \sim 10^{11}$ Pa, $r_0 \sim 10^{-10}$ m, $k \sim 10$ N m^{-1} .

(c) When the temperature is increased, r_0 increases and dF/dr decreases (fig. 16.3). E must therefore decrease.

(d) By using $f = (1/2\pi)\sqrt{(k/m)}$ and taking the atomic mass $m \sim 10^{-25}$ kg, the *natural atomic frequency of vibration* is shown to be $\sim 10^{12}$ Hz.

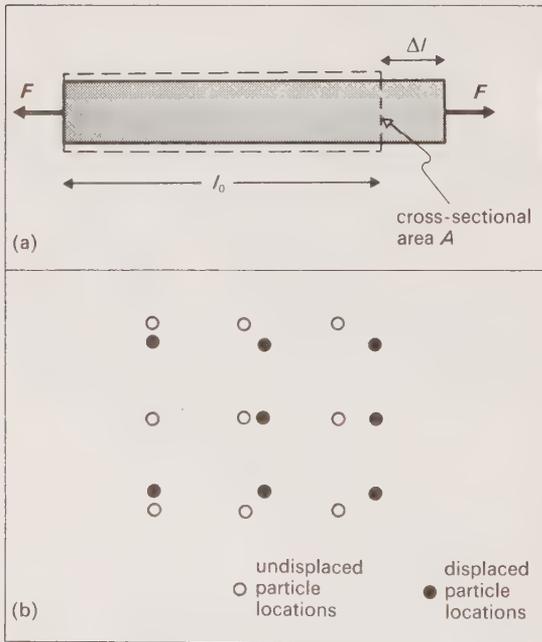


Fig. 17.3. Illustration of Young's modulus on (a) a macroscopic, and (b) a microscopic scale (both exaggerated).

17.4 BULK MODULUS K

$$\text{Stress } \sigma = \frac{\text{change of normal force}}{\text{area}} = \Delta p$$

(the change of hydrostatic fluid pressure)

$$\text{Strain } \theta = \frac{\text{change of volume}}{\text{original volume}} = \frac{\Delta V}{V_0}$$

Bulk modulus K is defined by the equation

$$K = \frac{\sigma}{\theta} = \frac{-\Delta p}{\Delta V/V_0}$$

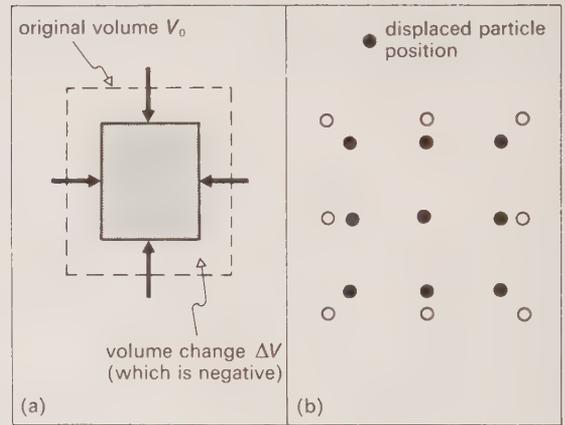


Fig. 17.4. To illustrate bulk modulus on (a) a macroscopic, and (b) a microscopic scale (both exaggerated).

Notes

(a) An increase in p produces a diminution in V . The negative sign is introduced to make K positive.

(b) $1/K$ is called **compressibility** (κ). A material is easily compressed if it has a *small* bulk modulus.

(c) K is approximately constant for the condensed phases (*Hooke's Law*).

(d) We can calculate K for gases, as shown on p. 148.

(e) For a solid material (at low temperatures to avoid thermal effects),

$$K = (\text{constant}) \times \left(\frac{\text{molar binding energy}}{\text{molar volume}} \right)$$

17.5 SHEAR MODULUS G

Shear modulus is sometimes called **rigidity modulus**, and is possessed only by *solids*, since the shape of a fluid is determined by its container.

$$\text{Stress } \tau = \frac{\text{tangential force}}{\text{area over which it is applied}} = \frac{F}{A}$$

$$\text{Strain } \gamma = \text{angle of shear } \gamma = \frac{\Delta x}{y}$$

and has no dimensions.

Shear modulus G is defined by the equation

$$G = \frac{\tau}{\gamma} = \frac{F/A}{\gamma}$$

$$G[\text{Pa}] = \frac{F[\text{N}]/A[\text{m}^2]}{\gamma[\text{rad}]}$$

(This equation should be compared with the defining equation for the *coefficient of viscosity* (p. 163)).

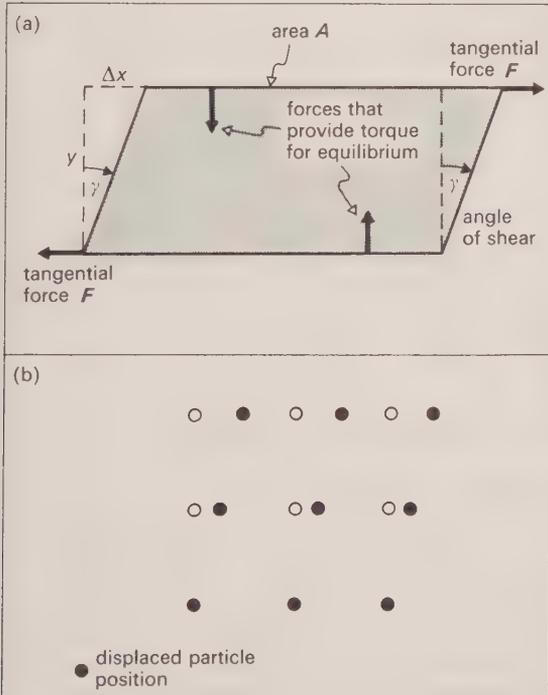


Fig. 17.5. To illustrate shear modulus on (a) a macroscopic, and (b) a microscopic scale (both exaggerated).

Examples of shearing

(a) The restoring torque T exerted by a wire twisted through angle θ is given by

$$T = c\theta$$

The **suspension** (or **torsion**) constant c is numerically equal to the restoring torque per unit angular displacement. See

(i) moving-coil galvanometer (p. 419),

(ii) *Boys's* experiment (p. 177).

(b) The torsional pendulum (p. 148).

(c) The extension of a helical spring involves the *twisting* of segments of the wire.

17.6 MEASUREMENT OF YOUNG'S MODULUS

In a measurement of E (fig. 17.6) we might use quantities of the following orders of magnitude:

$$F \sim 10^2 \text{ N} \quad (F \text{ is not expressed in kgf})$$

$$A \sim 10^{-1} \text{ mm}^2 \quad (\text{the mean value from several micrometer readings})$$

$$\Delta l \sim 2 \text{ mm}$$

$$l_0 \sim 2 \text{ m}$$

Obviously the values of A and Δl need especially careful measurement.

Δl should be measured during loading and unloading, using the same time delays, to reduce *hysteresis* (see fig. 17.7, in which the effect is exaggerated).

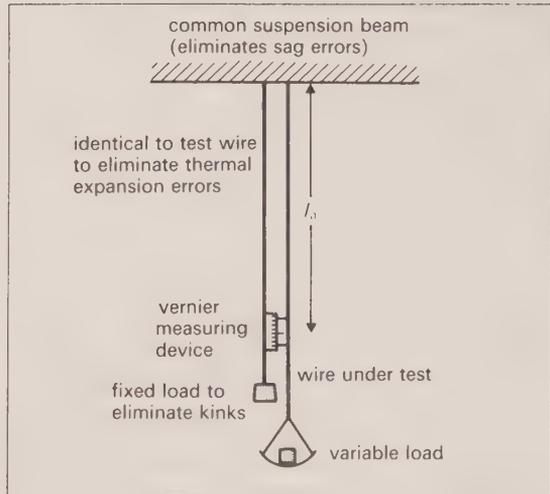


Fig. 17.6. Experimental arrangement for the measurement of E .

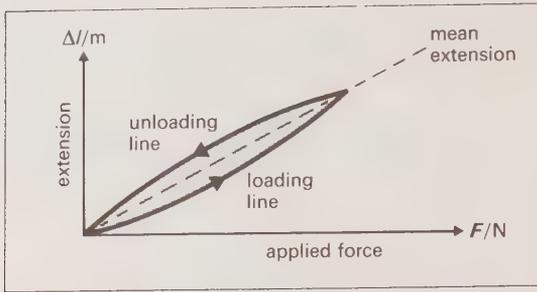


Fig. 17.7. Calculation of mean values of Δl for a wire taken through a hysteresis loop.

$$E = \frac{l_0}{A(\Delta l/F)}$$

In

l_0 and A are known, and the mean value of $(\Delta l/F)$ can be found from the graph.

17.7 BEHAVIOUR OF A WIRE UNDER STRESS

Fig. 17.8 shows a load–extension curve: the stress–strain curve is similar, but not identical, because l and A change during the experiment. The extension, rather than the load, is usually plotted along the x -axis.

(a) After the **limit of proportionality** the wire no longer obeys *Hooke's Law*, but still returns to its original form when the load is removed. It remains *elastic* as long as the strain is restricted to the stretching of the interatomic bonds.

(b) After the **elastic limit**, the wire shows a **permanent set**, and never regains its original shape and size. The deformation has been *dissipative* (p. 204).

(c) At the **yield point** there is a marked change in internal structure brought about by the slipping of crystal planes over one another without losing contact. The close-packing of the atoms is maintained by the metallic bonding.

(d) In the **plastic region** small increases in load produce marked increases in extension because of flow processes (fig. 17.9(b)).

(e) **Ductile fracture** occurs when the elongation and narrowing that accompanies the sliding of the atomic planes results in a constriction or *neck* of small cross-sectional area A . The **breaking stress** σ is then exceeded because $\sigma = F/A$ and so the tensile forces pull the wire apart.

(f) Before the constriction forms, further re-arrangement of the crystal imperfections (p. 144) means that any further increase of strain needs increased stress. This process is called **work-hardening** or **strain-hardening**.

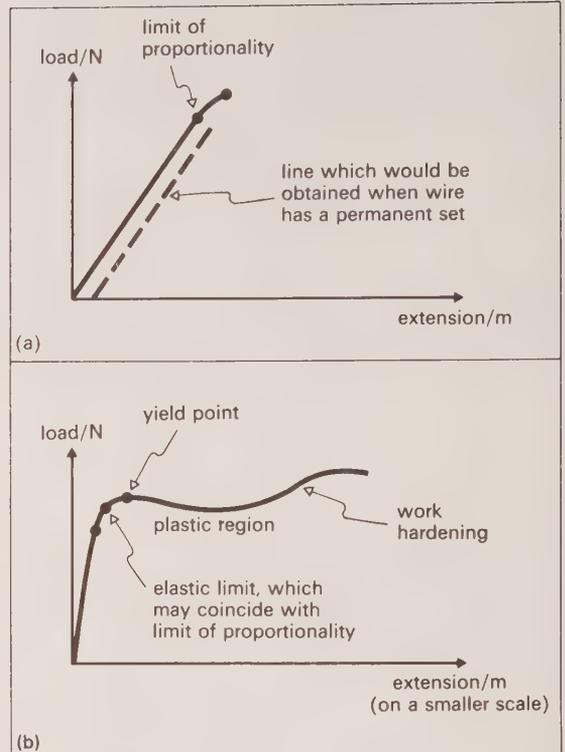


Fig. 17.8. Typical load–extension graphs for a ductile material.

Similar processes take place when a rod is compressed. At the breaking point the rod is *fractured* by compressive forces.

17.8 PROPERTIES OF MATERIALS

Stiffness and strength

A material is *stiff* if a large applied stress is required to produce a small strain, i.e. if it has a high *Young* modulus. A material is *strong* if a large stress is needed to break it. The **ultimate tensile stress** σ_u (or **ultimate strength** or **breaking stress**) is calculated from the maximum force applied and the *original* cross-sectional area. The ratio of the applied tensile force to the *original* cross-sectional area is known as the **nominal stress** and is clearly less than the true stress. In fig. 17.10,

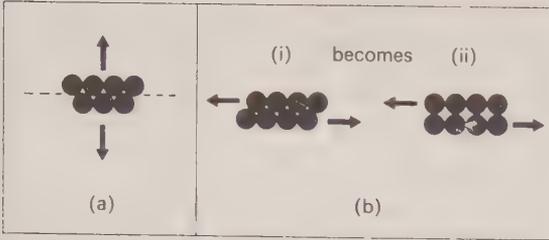


Fig. 17.9. Breakdown of solid lattice by (a) brittle fracture, and (b) plastic flow.

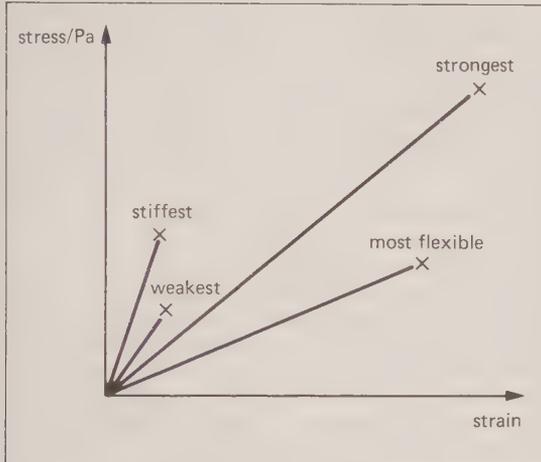


Fig. 17.10. Graphical illustration of stiffness and strength.

(a) the steeper the graph, the *stiffer* the material,
(b) the greater the breaking stress, the *stronger* the material.

The following table shows some approximate values for E and σ_u .

Material	E/GPa (stiffness)	σ_u/MPa (strength)
carbon fibre	410	2000
steel	210	2000
copper	130	200
glass	75	70
wood	12	60
polythene	1	30
rubber	0.007	30

Some materials, such as cast iron, concrete, plaster and masonry, are strong in compression but weaker in tension because of the effect of surface cracks. By contrast, wood is strong in tension but weaker in compression because the cell walls tend to collapse.

Toughness and brittleness

A **tough** material, such as steel or polythene, can absorb large amounts of strain energy and will deform plastically before breaking. Such materials have the ability to resist the propagation of cracks and are extremely important to engineers. A **brittle** material, such as glass or concrete, breaks over a very short period of time without plastic deformation.

Ductility and malleability

A **ductile** material, such as a metal, can be drawn into a wire or rolled into a thin sheet. This change of shape results from plastic deformation due to planes of atoms sliding over each other. A **malleable** material can be hammered, pressed or bent into a different shape.

Hardness

The **hardness** of a material is a measure of its resistance to scratching. It can be assessed from the load required to produce a particular size of indentation in the material. **Mohs' hardness scale**, from 1 to 10, uses ten natural minerals of increasing hardness, such as calcite (3) and quartz (7). Each selected mineral can scratch those below it on the scale.

Resilience

Samples of **resilient** materials, such as some rubbers and plastics, can absorb large quantities of strain energy below their elastic limits. Structures of high resilience can absorb the energy of mechanical blows without permanent damage, e.g. packaging, rubber fenders, helical springs in railway buffers. Examples from Nature include the ability of a spider's web to cope with impacts of flies, and the ability of trees and plants to survive in the wind. The engineering design of a structure has to find the right balance between resilience, stiffness and strength.

17.9 STRESS-STRAIN GRAPHS

In fig. 17.11 overleaf, graphs (a) to (d) show the stress-strain relationships for different materials. The scales of each graph should be noted carefully.

(A) Single metallic crystal

These '**whiskers**' are single crystals a few mm in length and a few μm in diameter. They can withstand much higher σ_u and strain values than a metal wire. This means that they are much *stronger* although they have the same stiffness. They break

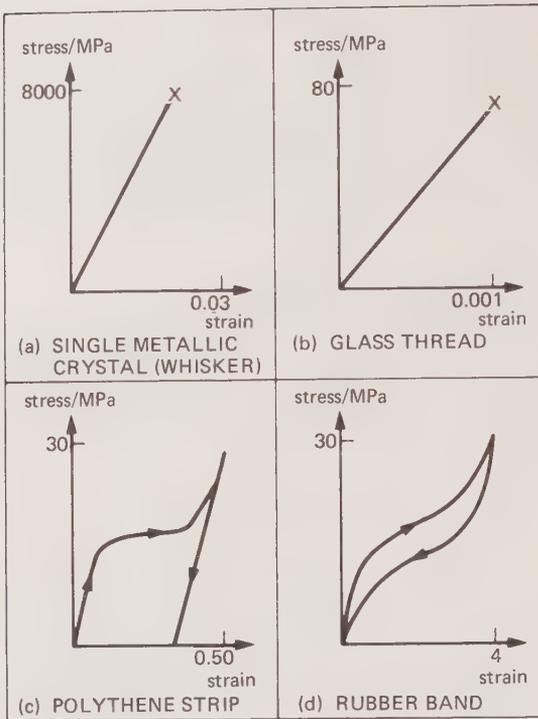


Fig. 17.11. Stress-strain graphs.

close to the theoretical values of σ_u and are thought to have almost perfect crystalline structure.

(B) Glass

A glass thread, drawn from a glass rod, behaves elastically when loaded until it exhibits **brittle fracture**. In theory, brittle fracture should occur when the applied stress reaches the calculated value to pull one layer of molecules away from an adjacent layer (see fig. 17.9(a)). The observed strengths are many times less than the theoretical values because of minute surface cracks cutting across the direction of the applied stress. These cracks act as **stress-multipliers** so that the stress at the tip of each crack is much greater than the applied stress. As the crack spreads, the stress at the tip increases further with the result that fracture takes place quickly. Freshly-drawn fibres of glass can be even stronger than steel because any cracks are parallel to the fibre axis. However, chemical action and handling soon reduce their strength. The spreading of cracks is inhibited if a material is in compression and so glass and concrete are often *pre-stressed* to increase their toughness. *Laminated glass* has great strength because it

consists of layers of pre-stressed glass alternating with layers of transparent adhesive which restrict the propagation of cracks. In ductile materials, such as metals, the progress of cracks is restricted by plastic deformation at the tip of each crack.

(C) Polythene

A strip of polythene behaves elastically, with constant stiffness, until it *necks* with a sudden increase in strain for hardly any increase in stress. The narrowed part stays the same width but increases in length because the polymer chains are unfolding along the axis of the strip. This results in a large number of crystals being aligned with the axis. After the unfolding, the chains themselves are being separated and so the strip becomes stiff again and reasonably elastic. A large permanent strain results when the stress is removed.

(D) Rubber

A rubber band behaves elastically and can show very large reversible strains, although extreme stretching can result in a small permanent set. The sizeable area of the *hysteresis loop* represents the considerable dissipation of heat energy during one loading-unloading cycle. Rubber used in, for example, car tyres should have a loop of small area so that temperature increases are small. When the strain is large, the cross-sectional area of the band is correspondingly small so that the real stress is much greater than that calculated from the original area. This means that rubber has greater stiffness than is apparent from the graph. At low stresses the polymer chains become untangled relatively easily leading to large strains. Once the chains have been straightened, the stiffness of the rubber increases because further strain requires separation of the molecules themselves. Unlike most other materials, the stiffness of rubber increases with temperature because the polymer chains become further entangled on heating.

17.10 DISLOCATIONS

The *theoretical* minimum shearing stress for crystal planes to slip over each other may be about 100 times greater than the measured value. The implication of this is that crystal structures must contain **imperfections** or **defects**. In 1934 *Taylor* suggested that the slip between adjacent atomic planes was equivalent to the movement of a dislocation. This would require only very small

movements of individual atoms and therefore relatively small stresses. A **dislocation** is an imperfection in the crystal lattice and may consist of an incomplete plane of atoms, the edge of which is the **dislocation line**. The movement of a dislocation is usually a combination of an **edge dislocation** (displacement perpendicular to the dislocation line) and a **screw dislocation** (displacement parallel to the dislocation line). Significant movement along a slip plane in a crystal requires great numbers of dislocations, many of which are produced by the multiplication of initial internal dislocations when stress is applied. Direct observation of dislocations is possible using *electron microscope* photographs and X-ray diffraction techniques. Surface dislocation lines on etched crystals can be seen under an *optical microscope*. An elementary study of dislocations can be made using a bubble raft.

When a metal is cold-worked by rolling or stretching, many dislocations are created. These dislocations interact with each other and with point defects and grain boundaries so that their progress is impeded. Consequently, the metal becomes stiffer and stronger and is said to be **work-hardened**. If a metal is heated and allowed to recrystallize slowly, it is said to be **annealed**. This process reduces many of the lattice distortions and so the metal may now be deformed more easily.

17.11 FATIGUE AND CREEP FRACTURES

(A) Fatigue fracture

Metal components often operate under conditions of fluctuating stress, e.g. repeated impact, twisting, bending or alternating tension and compression. **Fatigue fracture** occurs when a small crack forms, often where the stress is high, and spreads slowly through the metal until the sound part of the component fails suddenly as a result of the reduced area. If the same maximum stress were applied steadily, the material would be safe from fracture. Fatigue accounts for most metal failures in, for example, gears, axles and springs. Aircraft designers must consider the effects of varying stress due to changing altitude.

In fig. 17.12, graph (1) is typical for ferrous metals with corrosion and also for most non-ferrous metals and alloys. Graph (2) is typical for many ferrous metals without corrosion and implies that there is a safe range of *stress amplitude*. The fatigue limit usually increases with the strength of the metal and with fall in temperature.

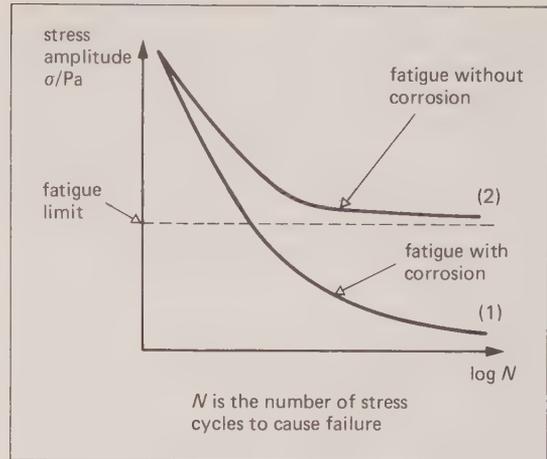


Fig. 17.12. Stress–endurance curves.

(B) Creep fracture

Fig. 17.13 shows that when a constant stress is applied to a material sample, an almost *instantaneous* strain results. **Creep** is the *continuous* deformation that occurs from prolonged static stress caused by tension, compression, bending, torsion or shear forces. At low temperatures, the **creep rate** $d\epsilon/dt$ decreases to a nearly constant low value as the work-hardening makes dislocation movement increasingly difficult. During **steady-state**

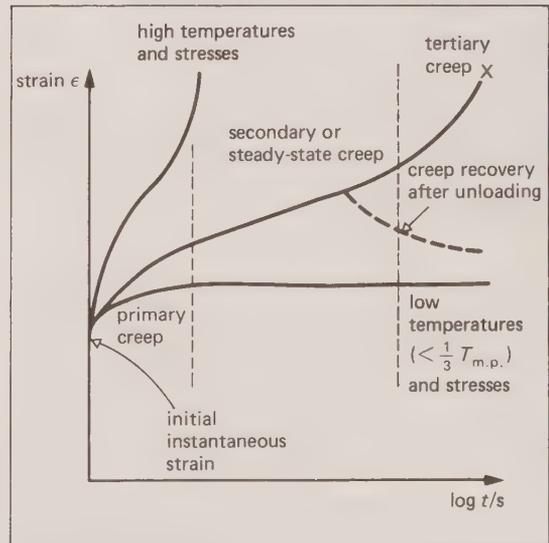


Fig. 17.13. Creep curves.

creep at higher temperatures, thermal energy effects such as solid-state diffusion balance the work-hardening effects. In **tertiary creep**, $d\varepsilon/dt$ increases due to the growth of cracks, and the sample eventually fractures. At a given temperature, $d\varepsilon/dt$ increases significantly with increasing applied stress. The high-temperature operation of machine components requires **creep-resistant** materials. These are usually special metal alloys (e.g. Ni, Ti, Al) which do not suffer from oxidation at high temperatures. Creep at room temperature can be observed in lead roofs and pipes, and in materials such as textiles, rubber and tarmac. The very low 'geological' creep rates of about 10^{-14} s^{-1} (c. 0.10 in 10^6 years), found within the Earth's crust at high pressures and temperatures, lead to ductile behaviour in most rocks. The higher creep rates that can exist in surface rocks lead to brittle fracture and earthquakes.

If E_D is the **molar self-diffusion energy** for molecules to move past their neighbours in a solid material, then for a given stress,

$$\frac{d\varepsilon}{dt} \propto e^{-E_D/RT}$$

where T is the temperature. The expression $e^{-E_D/RT}$ is a measure of the proportion of molecules which have sufficient **activation energy** for creep at a given applied stress. The graph of $\ln d\varepsilon/dt$ against $1/T$ would give a straight line with a negative gradient equal to $-E_D/R$.

17.12 STRESSED WIRES

(A) Energy stored in a stretched wire

If we apply a force F to a wire, then

$$F = \frac{EA}{l_0} \cdot x$$

when the wire is extended x . The energy W stored in the wire is the work done *on* the wire during extension, where

$$W = \int_0^{\Delta l} F dx$$

W is represented by the area under the line of fig. 17.14, so

$$\begin{aligned} W &= \frac{1}{2} F_{\max} \Delta l \\ &= \frac{1}{2} \left(\frac{F_{\max}}{A} \right) \cdot \left(\frac{\Delta l}{l_0} \right) \cdot (Al_0) \\ &= \frac{1}{2} \text{stress} \times \text{strain} \times \text{volume} \end{aligned}$$

When a wire is deformed, the energy W stored elastically in volume V is given by

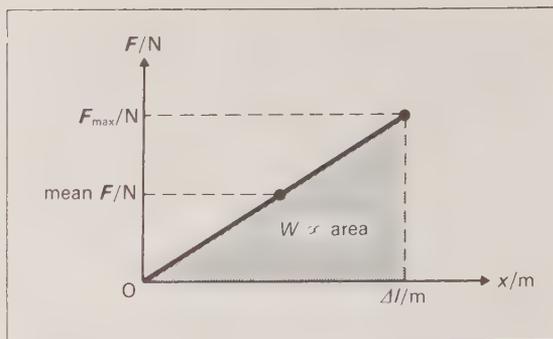


Fig. 17.14. Potential energy stored by a stretched wire.

$$\frac{W}{V} = \frac{1}{2} \text{stress} \times \text{strain}$$

$$\frac{W}{V} \left[\frac{\text{J}}{\text{m}^3} \right] = \frac{1}{2} \text{stress} \left[\frac{\text{N}}{\text{m}^2} \right] \times \text{strain} \left[\frac{\text{m}}{\text{m}} \right]$$

If the deformation of a wire is not perfectly elastic, the energy dissipated when the specimen is taken through a cycle such as that of fig. 17.7 is represented by the shaded area. (See also p. 461.)

The quantity W/V is called the **energy density** or **resilience** of the wire and is a measure of the strain energy stored in tension or compression, without permanent deformation. It is also given by the following expression:

$$\frac{W}{V} = \frac{1}{2} \varepsilon^2 E = \frac{1}{2} \frac{\sigma^2}{E}$$

If the wire experiences *shear* stress and strain, the energy density or resilience is given by

$$\frac{W}{V} = \frac{1}{2} \gamma^2 G = \frac{1}{2} \frac{\tau^2}{G}$$

The work done in stretching a specimen up to the point of fracture is closely related to the **heat of sublimation** of the substance, and so this can be used to estimate the theoretical strength (always far greater than the strength in practice). The heat of sublimation also gives a very good evaluation of the *molecular attraction energy* ε (p. 128).

(B) Tension in a cooling wire

Suppose a wire of natural length l_0 is heated through $\Delta\theta$, and prevented from contracting, as is the steel rim of a locomotive wheel when it is being fitted. The heating produces an extension Δl , where

$$\Delta l = l_0 \alpha \Delta\theta \quad (\text{def. of } \alpha)$$

but
$$F = EA \frac{\Delta l}{l_0} \quad (\text{def. of } E)$$

so the tension in the wire, when thus stretched, would be given by

$$F = EA\alpha \Delta\theta$$

and the stress F/A would be $\sigma = E\alpha \Delta\theta$.

In polycrystalline materials, different values of α lead to intergranular stresses and subsequent distortion. This is an important factor in processes such as the weathering of rocks and the thermal fatigue of tin-based alloy bearings.

17.13 POISSON'S RATIO μ

When a rod is stretched, its cross-sectional area decreases as the length increases, and there is usually a change of volume.

Poisson's ratio μ is defined by the equation

$$\mu = \frac{-\text{transverse strain}}{\text{longitudinal strain}}$$

Thus
$$\mu = \frac{-\Delta r/r_0}{\Delta l/l_0}$$

where r might represent the radius of a cylindrical rod. The negative sign is inserted so that μ should be a positive number. It has no dimensions and no unit. For a cylindrical rod

$$V = \pi r^2 l$$

$$\Delta V = \pi r^2 \Delta l + 2\pi r l \Delta r$$

For zero volume change on stretching,

$$\Delta V = 0$$

so
$$\mu = \frac{-\Delta r/r_0}{\Delta l/l_0} = \frac{1}{2}$$

It should be noted that the three moduli of elasticity used in this book are *not independent*. When one analyses waves travelling through elastic media, μ is a convenient parameter to relate these three moduli, and express the modulus appropriate to a particular situation.

17.14 COMPOSITE MATERIALS

These are solid systems of well-defined, distinct components. The interactions between adjacent

components lead to composite materials having their own characteristic properties, which may combine the desirable qualities such as lightness, stiffness and strength of the separate components. Modern engineering applications often use materials that combine the high *stiffness/weight ratio* of a material such as glass with the high *strength/weight ratio* of a material such as resin. Some important examples are described below.

(A) Natural wood

This consists of strong, brittle fibres of cellulose chains in a soft, amorphous matrix of lignin. This structure gives it strength but allows relatively easy splitting in the direction of the fibres. Plywood is difficult to split because it consists of alternate layers with fibre directions at right angles.

(B) Reinforced plastics

Glass-reinforced plastic (GRP), or *fibreglass*, consists of glass fibres embedded in a thermosetting plastic resin which sticks to them and inhibits the spreading of cracks. The material has great strength and many applications, e.g. boat hulls and car bodies. *Carbon-fibre-reinforced plastics* (CFRP) have significantly greater strength/weight and stiffness/weight ratios than steel, and have many specialized uses in spite of their high cost. The main disadvantage of plastics is that they decompose above about 250°C.

(C) Reinforced metals

These materials can be used at higher temperatures, with very high strength, by using metallic matrices for the fibre reinforcement, e.g. aluminium reinforced by silica or boron fibres, nickel reinforced by carbon fibres.

(D) Concrete

Hard concrete consists of a mixture of sand and stones bound together firmly by a matrix of small crystals. Although it is very strong under compression, it is about ten times weaker in tension. *Reinforced concrete* uses steel rods, which bind strongly with the concrete over a wide temperature range, giving it strength in both compression and tension. In *pre-stressed concrete* the steel rods are in a state of tension so that the concrete is always in compression thereby reducing the risk of cracking.

► 17.15 ISOTHERMAL AND ADIABATIC BULK MODULUS

(A) Isothermal bulk modulus

From p. 140

$$K_{\text{iso}} = \frac{-\Delta p}{\Delta V/V_0} = -V_0 \frac{dp}{dV}$$

Under isothermal conditions an ideal gas obeys Boyle's Law, $pV = \text{constant}$.

Differentiating $-V \left(\frac{dp}{dV} \right)_{\text{iso}} = p$

Thus the *isothermal* bulk modulus of an ideal gas is equal to its pressure.

(B) Adiabatic bulk modulus

Under certain conditions (p. 209) the pressure and volume of an ideal gas are related by

$$pV^\gamma = \text{constant} \\ p = \text{const } V^{-\gamma}$$

Differentiating

$$\left(\frac{dp}{dV} \right)_{\text{ad}} = -\gamma \text{const } V^{-\gamma-1} \\ -V \left(\frac{dp}{dV} \right)_{\text{ad}} = \gamma p$$

The *adiabatic* bulk modulus of an ideal gas

$$K_{\text{ad}} = \gamma p$$

K_{ad} controls the speed of audible sound waves through the gas (p. 328).

► 17.16 THE TORSIONAL PENDULUM

(A) An elemental cylinder

In fig. 17.15, angle of shear = γ .

Effective area A over which tangential force is applied is half the shaded area:

$$A = \pi r \delta r$$

By geometry $XY = r\theta \approx l\gamma$ if γ is small. Then

$$G = \frac{F/A}{\gamma} = \frac{F/\pi r \delta r}{r\theta/l}$$

The torque of the couple is

$$F2r = \frac{G\theta 2\pi r^3 \delta r}{l}$$

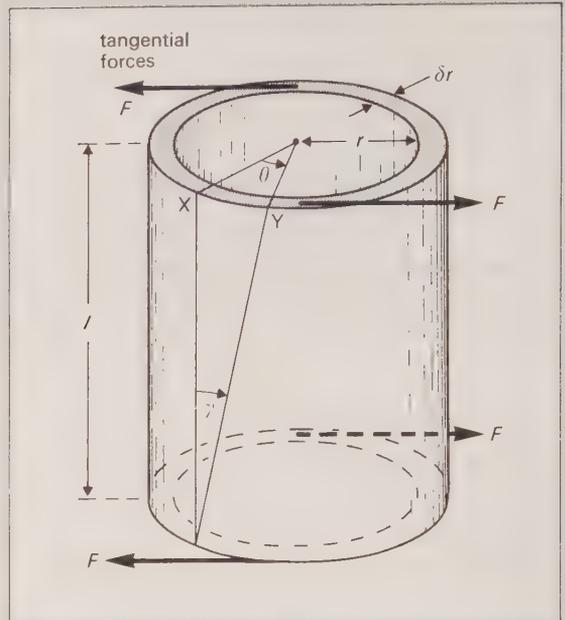


Fig. 17.15. Equal but opposite torques applied to a cylindrical shell cause an angle of shear γ .

Note that the shear is zero at the centre ($r = 0$) and increases to a maximum at the surface.

(B) A solid cylinder

Total torque to produce a twist θ in fig. 17.16,

$$T = \int_0^a \frac{G\theta 2\pi r^3 dr}{l} = \frac{\pi G a^4}{2l} \theta = c\theta$$

The torque exerted by the cylinder is given by $T = -c\theta$, because it is a restoring couple and acts so as to decrease θ . c is the **torsion constant**.

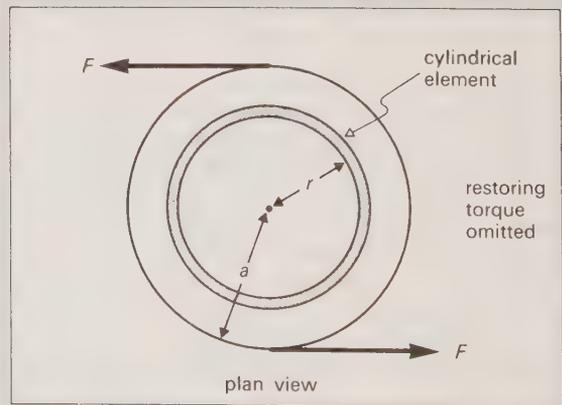


Fig. 17.16. Cylindrical element as part of a solid cylinder.

(C) Application to the pendulum

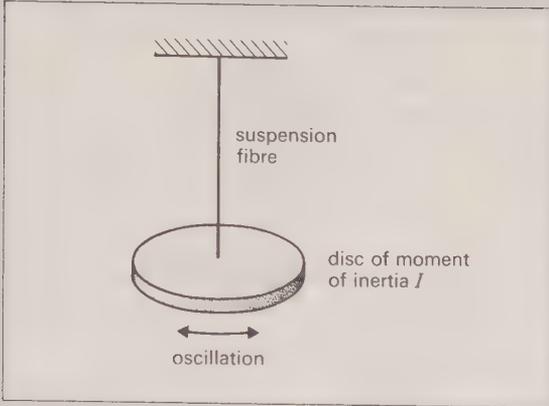


Fig. 17.17. The torsional pendulum.

Use $T = I\ddot{\theta}$ (p. 57)

$$\frac{-\pi G a^4}{2l} \theta = I\ddot{\theta}$$

$$\ddot{\theta} = \frac{-\pi G a^4}{2I l} \theta$$

Thus the motion is simple harmonic, and has a time period

$$2\pi \sqrt{\frac{2I l}{\pi G a^4}}$$

This result can be used for the measurement of G , the shear modulus.

17.17 BENDING MOMENT M AND SHEARING FORCE F_s

In fig. 17.18(a), the longitudinal filaments in the upper half of the bending beam will be stretched and those in the lower half will be compressed. The **neutral surface**, which passes through the centre of mass, is the plane in which the filaments experience zero stress and strain. The stress and strain in the other filaments are directly proportional to the distance of the filaments from the neutral surface. The **neutral axis** is the line of intersection between the neutral surface and any lateral plane through the beam. The shaded area WXYZ is any cross-section so that the beam is divided into two portions which must exert an equal and opposite system of forces on each other if the beam is in equilibrium. Each system reduces to a vertical force (**shearing force F_s**) in the plane of WXYZ and a couple (**bending moment M**). The beam bends until the couple produced by

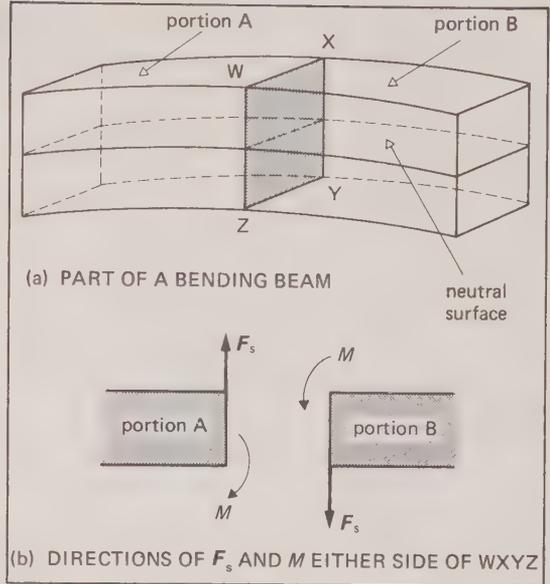


Fig. 17.18. Bending of a beam.

the tension and compression forces at WXYZ is exactly equal and opposite to the applied bending moment acting at this section. The magnitude of F_s at WXYZ is equal in size to the resultant of all the vertical forces acting either on portion A or on portion B, and its direction is opposite to this resultant. The value of M at this section is the sum of the moments of the forces acting either on portion A or on portion B of the beam. As a convention, the equilibrium of the left-hand portion (portion A) can be considered and the values of F_s and M taken to be positive when they act in the directions shown in fig. 17.18(b). Graphs which show the variation of these quantities with distance along the main axis of a beam are **shearing force diagrams** and **bending moment diagrams** respectively. Fig. 17.19 (overleaf) shows two simple situations. In both cases it can be seen that

$$\frac{dM}{dx} = F_s$$

and

$$\frac{dF_s}{dx} = \frac{d^2M}{dx^2} = 0$$

When the beams are loaded uniformly, F_s is directly proportional to x and so each M graph becomes a curve.

The *Young modulus E* (or *stiffness*) of the material of the beam can be determined from the maximum deflection y_{max} of the beam. If l is the **second**

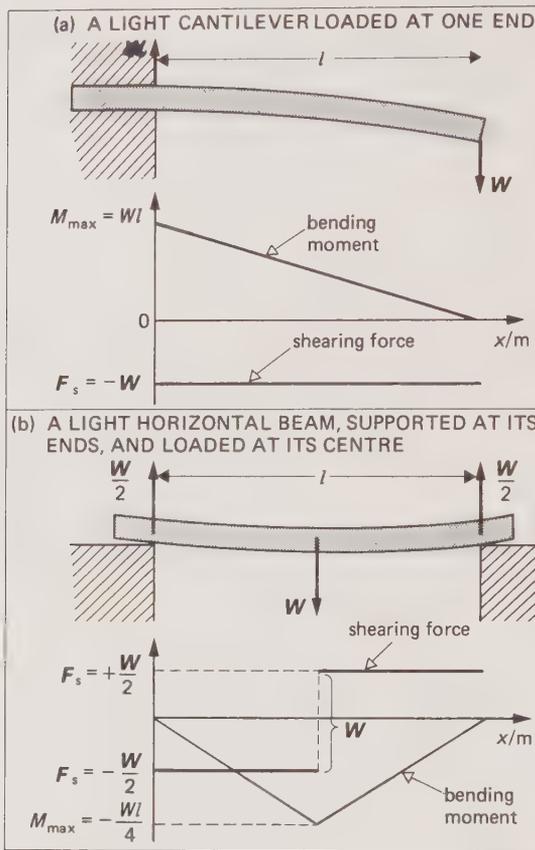


Fig. 17.19. Bending moment and shearing force diagrams.

areal moment (see fig. 17.20) of the cross-sectional area of the beam about a neutral axis in the plane of cross-section, then for

(a) a light cantilever loaded at one end,

$$y_{\max} = \frac{1}{3} \frac{Wl^3}{EI}$$

(b) a light beam, supported at its ends and loaded at its centre,

$$y_{\max} = \frac{1}{48} \frac{Wl^3}{EI}$$

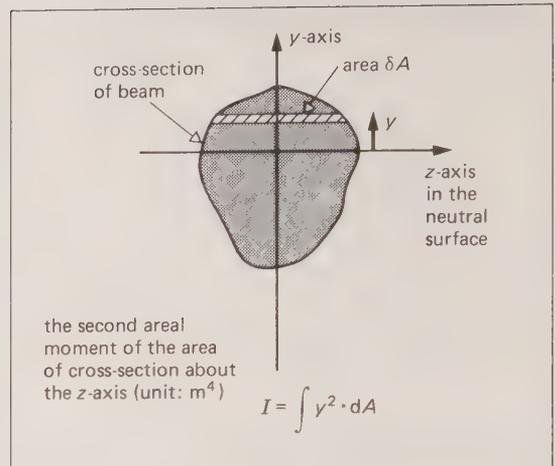


Fig. 17.20. The second areal moment.

For a loaded cantilever, E may also be calculated from its period of vibration,

$$T = 2\pi \sqrt{\frac{ml^3}{3EI}}$$

where m is the mass at the end of the cantilever.

The quantity EI (N m^2) is known as the **flexural rigidity** of the beam. A high value leads to small deflections, a low value of resilience and high frequencies of vibration. From the above expressions for a cantilever,

$$T = 2\pi \sqrt{\frac{y_{\max}}{g}}$$

which leads to a determination for the value of g .

The weight (and cost) of beams can be reduced, without altering their strength significantly, by using I-, Z- or T-shaped sections instead of rectangular ones. This is because material near to the neutral surface experiences only low stress and therefore a considerable proportion is not required. The circular cross-section of pipes and tubes has the advantage of equal strength in all directions perpendicular to their axes.

18 SURFACE TENSION

18.1	DEFINITIONS	151	▶ 18.6	SURFACE ENERGY AND THE SPHERICAL SOAP BUBBLE	156
18.2	DISCUSSION OF MOLECULAR BEHAVIOUR	152	18.7	CAPILLARY ACTION	157
18.3	THE SHAPE OF LIQUID SURFACES	153	18.8	SOME SURFACE TENSION PHENOMENA	157
18.4	EXCESS PRESSURE INSIDE A SPHERICAL LIQUID SURFACE	155	18.9	MEASUREMENT OF THE ANGLE OF CONTACT θ	158
18.5	GENERAL RESULT FOR EXCESS PRESSURE	156	18.10	MEASUREMENT OF SURFACE TENSION γ	159

18.1 DEFINITIONS

Surface tension γ

Refer to fig. 18.1.

The surface tension γ is defined by the equation

$$\gamma = \frac{F}{\Delta l}$$

$$\gamma \left[\frac{\text{N}}{\text{m}} \right] = \frac{F[\text{N}]}{\Delta l[\text{m}]}$$

γ is a force per unit length.

$$[\gamma] = \frac{[\text{MLT}^{-2}]}{[\text{L}]} = [\text{MT}^{-2}]$$

γ decreases as the temperature increases, approaching zero near the critical temperature. At this temperature the liquid and vapour become indistinguishable (see p. 231).

Substance	$\gamma/\text{N m}^{-1}$	T/K
diethyl ether	1.7×10^{-2}	293
soap solution	2.5×10^{-2}	290
water	5.9×10^{-2}	373
water	7.3×10^{-2}	290
mercury	47×10^{-2}	293

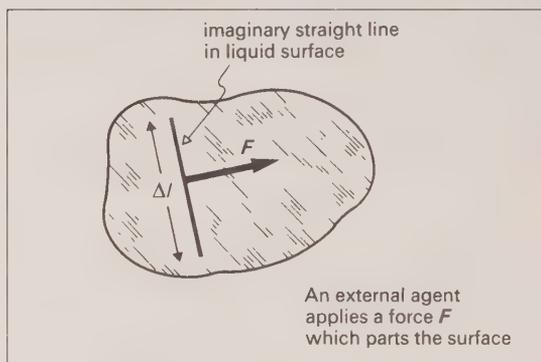


Fig. 18.1. Definition of surface tension γ . (Note that F is drawn in the plane tangential to the surface.)

Free surface energy σ

Suppose that under *isothermal* conditions mechanical energy W is required to create an additional area A of surface.

The **free surface energy** σ is defined by the equation

$$\sigma = \frac{W}{A}$$

$$\sigma \left[\frac{\text{J}}{\text{m}^2} \right] = \frac{W[\text{J}]}{A[\text{m}^2]}$$

σ represents work done per unit area and is numerically and dimensionally equal to γ .

$$[\sigma] = \frac{[MLT^{-2}L]}{[L^2]} = [MT^{-2}]$$

► Under *adiabatic* conditions (such that no heat enters or leaves the system) the temperature decreases. Under *isothermal* conditions heat is added to the surface to restore the original temperature. Thus

$$\left(\begin{matrix} \text{additional energy} \\ \text{of film} \end{matrix}\right) = \left(\begin{matrix} \text{work} \\ \text{done} \end{matrix}\right) + \left(\begin{matrix} \text{added thermal} \\ \text{energy} \end{matrix}\right)$$

so total surface energy $> \sigma$.

The decrease of temperature under adiabatic conditions leads to an increased γ , and thus an increased force to oppose the increase of area—an example of *Le Chatelier's Principle*.

To show $\gamma = \sigma$ numerically

The mechanical work done by the agent in fig. 18.2 is

$$F \Delta x = (2\gamma a) \Delta x$$

where the 2 takes account of the two surfaces of the film.

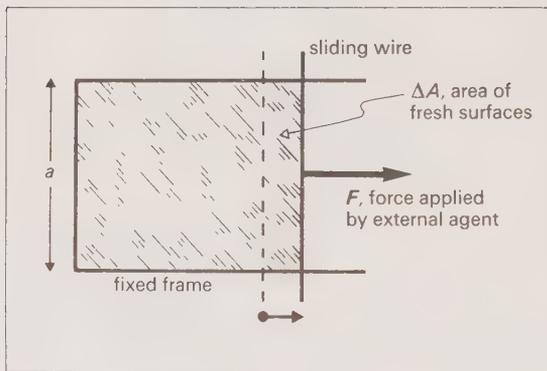


Fig. 18.2. An external agent increases the area of a soap film.

Increase in free surface energy is

$$\sigma \Delta A = \sigma(2a \Delta x)$$

Under *isothermal* conditions we equate

$$2\gamma a \Delta x = \sigma 2a \Delta x$$

so

$$\sigma = \gamma$$

The angle of contact θ

θ is the angle between the tangent planes to the two surfaces, measured *through the liquid* (fig. 18.3). Even for a given liquid and solid, the value of θ is not constant (p. 158).

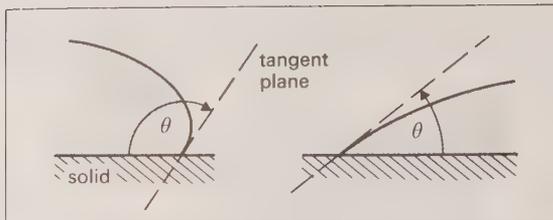


Fig. 18.3. To define angle of contact θ .

18.2 DISCUSSION OF MOLECULAR BEHAVIOUR

For our microscopic examination of surface tension, we can consider either forces or energies, as we did for the macroscopic definitions.

(A) Discussion in terms of forces

A liquid molecule experiences short-ranged forces which are both attractive and repulsive, but the repulsive forces have the shorter range. Fig. 18.4 brings out these points.

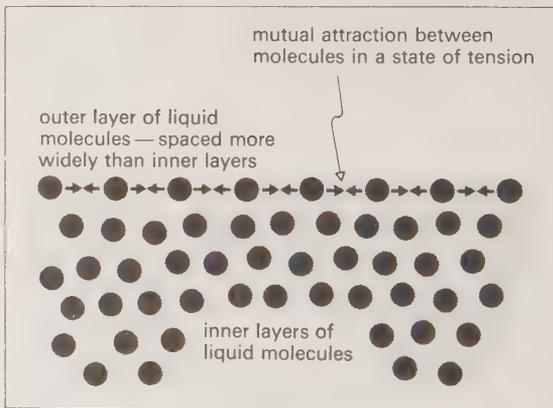


Fig. 18.4. Attractive forces between surface liquid molecules which are in a state of tension.

(a) The top layer of molecules does not experience any interaction from above. Molecules which try to leave this layer by moving upwards experience a downward force which may prevent them from doing so. This force cannot persist and the returning molecules experience abrupt, halting collisions. The persisting force field, in which

molecules experience unbalanced downward attractive forces towards the surface, is only a few molecular diameters thick.

(b) The individual molecules in the top layer are more widely spaced than those in the bulk of the liquid. Therefore the attractive forces that they experience from either side from their close neighbours cause them to be in a *state of tension* (p. 35). If an attempt is made to part the liquid surface, and thereby to break on *one side only* the bonds that cause the tension force, then the molecules at that point experience a resultant attractive force from the molecules on *the other side*.

The effect is appropriately called **surface tension**, since the increased lateral spacing of the surface molecules causes the surface to behave like a stretched membrane. (But note that because we are dealing with a liquid, an increase in area brings more molecules into the surface layer, so γ does *not* increase as the area increases.)

It is called a **contractile** property and is generally restricted to liquids. When external conditions allow, there is spontaneous contraction of the surface area with movement of surface molecules into the interior of the liquid.

(B) Discussion in terms of energy

Fig. 18.4 shows that surface molecules have an increased separation: fig. 16.5 then indicates that their mutual intermolecular p.e. has become *less negative*, i.e. has increased. This point is illustrated in a different way by fig. 18.5.

The parting of the two hemispheres involves work being done by an external agent to part

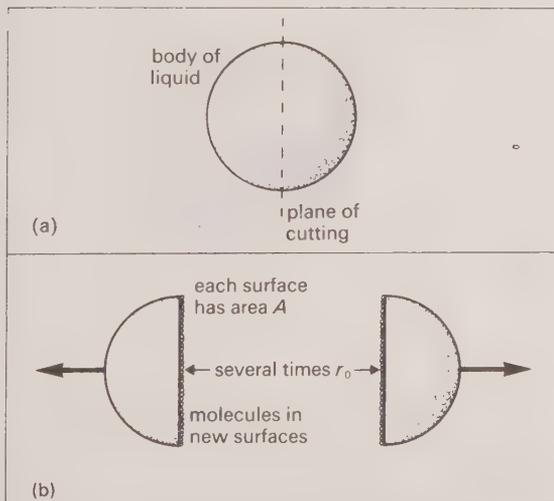


Fig. 18.5. A method of estimating σ : in (a) we have a spherical drop, which in (b) has been cut into two halves.

molecules which attract one another during separation. The work done by the agent appears as the molecular p.e. associated with molecules in the new surface, i.e. as **surface energy**.

► Suppose each molecule in the new surface has lost about three near neighbours. Then the work W done by the agent is given by

$$W = \left(\begin{array}{c} \text{number of} \\ \text{lost} \\ \text{neighbours} \end{array} \right) \times \left(\begin{array}{c} \text{energy for separation} \\ \text{of each} \\ \text{molecular pair} \end{array} \right) \\ \times \left(\begin{array}{c} \text{number of} \\ \text{molecular pairs} \\ \text{involved} \end{array} \right) \approx 3\varepsilon \left(\frac{A}{r_0^2} \right)$$

where ε and r_0 were defined in fig. 16.5 and the area occupied by a molecule in the surface has been taken as r_0^2 . The new area created is $2A$.

Using typical values $r_0 \sim 10^{-10}$ m and $\varepsilon \sim 10^{-21}$ J, we find

$$\sigma = \frac{W}{2A} = \frac{3\varepsilon}{2r_0^2} \sim 10^{-1} \text{ J m}^{-2}$$

The fact that the result is of the right order of magnitude lends support to the kinetic theory (p. 136).

A liquid surface has p.e. Any system takes up, in equilibrium, that configuration for which the p.e. becomes a minimum. This is a useful tool, provided one remembers that other types of p.e. may be involved.

18.3 THE SHAPE OF LIQUID SURFACES

Some useful ideas

(a) The surface of a liquid in equilibrium is parallel to the tension forces established in the surface, which are perpendicular to the resultant attractive force experienced by a molecule displaced above the surface. If this were not so the attractive forces would have a resolved part along the surface which would accelerate the surface molecules.

(b) The gravitational force on an individual molecule is entirely negligible compared to the intermolecular forces.

(c) The pull of the Earth on a liquid drop assumes greater importance as the drop size increases.

$$\text{The gravitational p.e.} \propto \text{mass} \propto r^3$$

$$\text{whereas} \quad \text{surface p.e.} \propto \text{area} \propto r^2$$

for a spherical drop.

(A) Absence of solids and gravitational forces

In two dimensions a cross-section of the liquid surface will be a circle (fig. 18.6).

In three dimensions the surface will be a sphere. Since this is the shape for which a given volume has the smallest area, it follows that a given liquid mass has smallest surface p.e. This situation holds approximately for

- (a) a soap bubble (whose mass is negligible),
- (b) very small droplets, and
- (c) globules suspended by a fluid of density equal to their own, or in a state of free-fall.

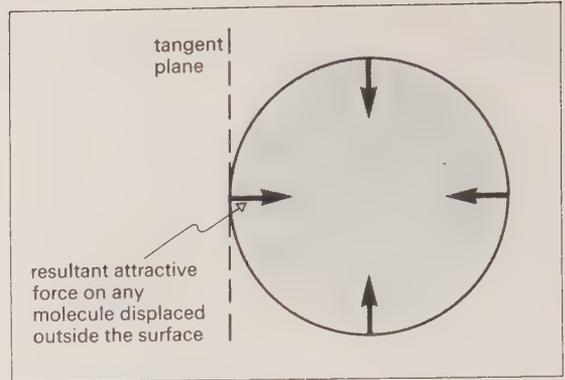


Fig. 18.6. A liquid surface shown in two dimensions. (Repulsive forces not shown.)

(B) When gravitational forces are appreciable

In fig. 18.7 the drop p.e. is made up by

- (a) surface p.e., $A\sigma$, plus
- (b) gravitational p.e., mgh .

The centre of gravity is pulled down until, in the first instance, the decrease of gravitational p.e. exceeds the increase of surface p.e. Equilibrium is attained when any further movement would make the decrease of one equal the increase of the other. The total p.e. is then a minimum.

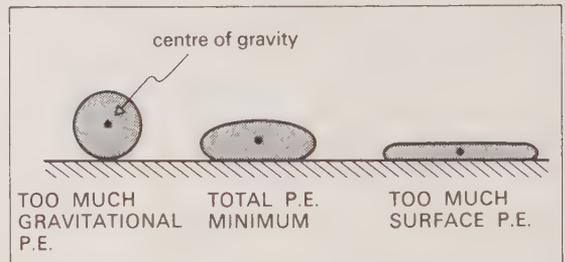


Fig. 18.7. Possible shapes for a mercury drop.

(C) The effect of a solid

Although solids do not possess contractile surface tension, they do affect the tension forces in liquid surfaces which are in interfacial contact with them. Only a liquid can store contractile p.e. in its surface when an applied force increases the surface area, because molecules can move from its interior to fill any gaps as they are created. Consider the following forces:

S = the adhesive force between solid and liquid surfaces, directed normally across the interface (S will depend upon the nature of both the solid and the liquid).

L = the cohesive force within the liquid surface,

M = the modified cohesive force within the liquid surface resulting from contact with the solid surface.

It can be shown from Dupré's relationship that $L - M = S$.

In fig. 18.8, a liquid is shown with a flat surface in contact with a solid wall. If molecule X is in equilibrium, it must experience balanced attractions and repulsions on all sides. As there is no distorting force upwards, there can be none acting

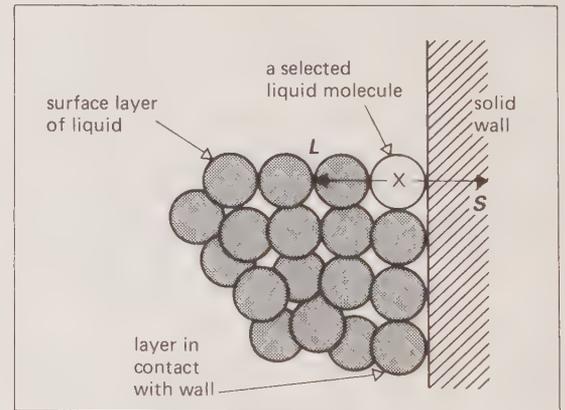


Fig. 18.8. A flat liquid surface in contact with a solid wall.

downwards, so that the tension forces in the vertical layer of liquid molecules in contact with the wall have been reduced to zero. Therefore, from the above relationship,

$$M = L - S = 0 \quad \text{and} \quad L = S.$$

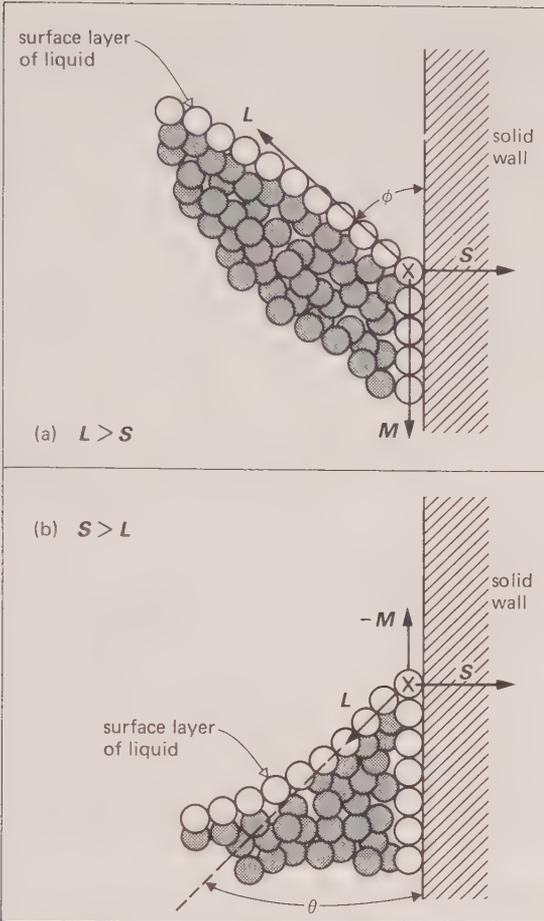


Fig. 18.9. Curved liquid surfaces in contact with a solid wall.

Fig. 18.9(a) shows that, when $L > S$, the liquid must be curved at the contact with the wall. For equilibrium of molecule X,

$$S = L \sin \phi \quad \text{and} \quad M = L \cos \phi = L - S$$

The surface tension forces in the layer next to the wall have been reduced, but not to zero. Fig. 18.9(b) shows that when $S > L$, the adsorbed layer of molecules tends to *increase* its area, giving a *negative* surface tension force or *extending thrust*. For equilibrium of molecule X,

$$S = L \sin \theta \quad \text{and} \quad -M = L \cos \theta = S - L$$

As $\theta = (\pi - \phi)$, the above results can be summarized by the equation

$$S = L(1 + \cos \theta)$$

This enables S to be calculated as long as $S < 2L$. It is clear that the relative magnitudes of L and S control the shape of a liquid surface, and its

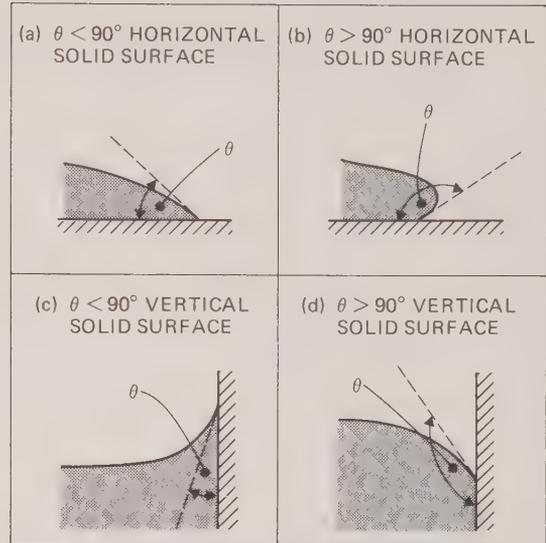


Fig. 18.10. Liquid and solid surfaces in contact.

behaviour. Fig. 18.10 shows some shapes of liquid surfaces in contact with horizontal and vertical solid surfaces.

The angle of contact θ

θ was defined on p. 152. A liquid for which θ is small is said to *wet* the particular surface it adjoins, and *spreads* over that surface.

A **wetting agent** (such as a detergent) greatly *reduces* θ .

A **waterproofing agent** aims to prevent wetting by *increasing* θ .

A **meniscus** is a consequence of θ having a particular value. The gravitational p.e. becomes relatively less significant when the vessel's dimensions are *small*. A *fine* tube shows greater **capillary effects** than a wide tube.

18.4 EXCESS PRESSURE INSIDE A SPHERICAL LIQUID SURFACE

Fig. 18.11 (overleaf) shows a spherical cap in equilibrium under the action of two sets of forces.

(A) The surface tension forces

The force at a particular point is tangential to the surface. The resultant of all these forces

(a) has a zero resolved part in the plane of the base, but

(b) has a resolved part perpendicular to this plane

$$\gamma(2\pi R \sin \alpha) \cos(90^\circ - \alpha) = \gamma 2\pi R \sin^2 \alpha$$

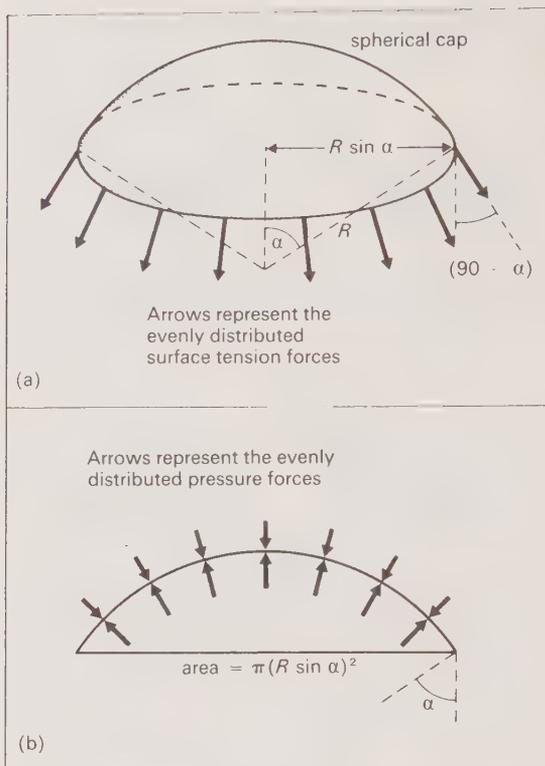


Fig. 18.11. Forces acting on a spherical cap: (a) due to surface tension, and (b) due to (excess) pressure.

(B) The pressure force

To maintain equilibrium the pressure inside the surface exceeds that outside, by Δp , say.

This pressure difference causes a resultant force which

(a) has a zero resolved part in the plane of the base, but

(b) has a resolved part perpendicular to this plane of

$$\Delta p \times (\text{area of base}) = \Delta p \times \pi(R \sin \alpha)^2$$

Equating

$$\Delta p \times \pi R^2 \sin^2 \alpha = \gamma 2\pi R \sin^2 \alpha$$

so
$$\Delta p = \frac{2\gamma}{R}$$

Notes

(a) Δp is independent of α : therefore this result holds for all surfaces which form part of a sphere, e.g. a hemisphere.

(b) A film has two surfaces, and there is a pressure difference across each. Therefore a spherical soap film shows an excess pressure of $4\gamma/R$.

(c) Note that small soap bubbles have a larger Δp than big bubbles.

18.5 GENERAL RESULT FOR EXCESS PRESSURE

Suppose a curved surface has principal radii of curvature R_1 and R_2 . Then it can be shown by using the principle of virtual work that

$$\Delta p = \gamma \left(\frac{1}{R_1} + \frac{1}{R_2} \right)$$

The pressure on the concave side of the surface exceeds that on the convex side.

(a) A **synclastic** surface is one for which the principal centres of curvature are found on the same side of the surface, e.g. a sphere.

(b) An **anticlastic** surface is one for which they are on opposite sides, e.g. a saddle-shaped surface, such as that formed by separating two coaxial funnels joined by an initially cylindrical film.

For a synclastic surface, R_1 and R_2 have the same sign. For an anticlastic surface R_1 and R_2 are allotted opposite signs.

Special cases

- (i) The soap bubble and liquid drop.
- (ii) The cylindrical film

$$\Delta p = 2\gamma \left(\frac{1}{R} + \frac{1}{\infty} \right) = \frac{2\gamma}{R}$$

and liquid jet

$$\Delta p = \gamma \left(\frac{1}{R} + \frac{1}{\infty} \right) = \frac{\gamma}{R}$$

(iii) For an anticlastic surface showing no pressure difference, such as that in (b), $\Delta p = 0$ means that R_1 and R_2 are numerically equal, but of opposite sign.

► 18.6 SURFACE ENERGY AND THE SPHERICAL SOAP BUBBLE

Suppose the radius of a spherical soap bubble be increased from r_0 to $(r_0 + \delta r)$.

If r_0 represents the equilibrium radius, the net

energy change is zero. Thus, using the principle of virtual work,

$$\left(\begin{array}{c} \text{work done by} \\ \text{pressure} \\ \text{forces} \end{array} \right) + \left(\begin{array}{c} \text{work done by} \\ \text{surface tension} \\ \text{forces} \end{array} \right) = 0$$

$$\Delta p \delta V = \text{increase in surface energy} \\ = \sigma \delta A$$

assuming the change to be reversible (p. 204).

$$\text{Using } A = (4\pi r^2) \times 2 \quad V = \frac{4}{3}\pi r^3 \\ \delta A = (8\pi r \delta r) \times 2 \quad \delta V = 4\pi r^2 \delta r \\ \Delta p 4\pi r_0^2 \delta r = \sigma 16\pi r_0 \delta r$$

whence
$$\Delta p = \frac{4\sigma}{r_0}$$

18.7 CAPILLARY ACTION

This is typical of a situation which may be analysed by a discussion of

- (a) forces (indicated on p. 152),
- (b) energy, or
- (c) pressures, as is done here.

In fig. 18.12 we suppose the liquid surface to be part of a sphere and the sides of the tube to be nearly vertical so that the angles marked can be taken as equal.

$$\text{Pressure at A} = \text{pressure at B (flat surface)} \\ = \text{pressure at C} \\ = \text{pressure at D} + h\rho g$$

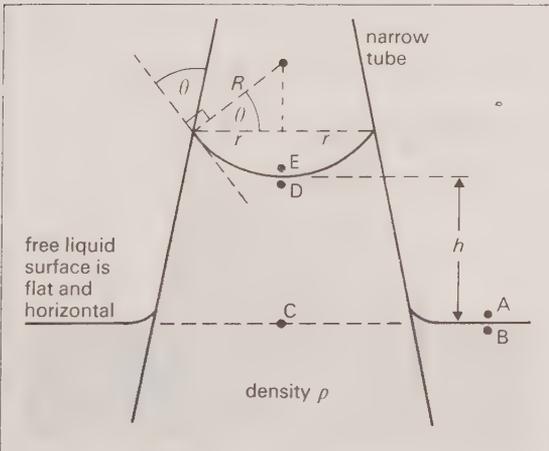


Fig. 18.12. Calculation of capillary rise by a pressure argument.

$$\text{Pressure at E} = \text{pressure at D} + \frac{2\gamma}{R} \\ = \text{pressure at D} + \frac{2\gamma \cos \theta}{r}$$

$$\text{Pressure at A} = \text{pressure at E} \\ + hg \times (\text{gas density}) \\ \approx \text{pressure at E}$$

so
$$\frac{2\gamma \cos \theta}{r} = h\rho g$$

Notes

- (a) $h \propto \gamma$, and $h \propto 1/r$.
- (b) If h and r are measured, then γ can be found.
- (c) When $\theta < 90^\circ$, h is +ve
 $\theta = 90^\circ$, h is 0
 $\theta > 90^\circ$, h is -ve.

A negative h corresponds to **capillary depression** as is shown by mercury, for which θ can be 140° .

(d) Since we assume *here* that the value of θ is constant, the orientation of the liquid surface changes according to the shape of the containing tube.

18.8 SOME SURFACE TENSION PHENOMENA

In this section we mention briefly surface phenomena, some of which the reader should try to explain using basic principles.

(a) If two *parallel plates* are pressed together so as to sandwich a thin film of liquid between them, a large force is needed to pull them apart. (Think in terms of the *small* pressure that results *within* the liquid.)

(b) If a *narrow tube* contains a large number of liquid drops, it may not be possible to clear the tube by blowing at one end. (Think in terms of the pressure difference across *each* drop when distorted.)

(c) Imagine *two bodies floating* in a liquid with which they make angles of contact of θ_1 and θ_2 respectively.

Then if

- (1) θ_1 and θ_2 are both less than 90° , or
- (2) θ_1 and θ_2 are both more than 90° ,

then the bodies are drawn together. But if

- (3) $\theta_1 > 90^\circ > \theta_2$, or vice versa, then the

bodies are forced apart.

(Think in terms of the *pressure differences* across the respective menisciuses.)

(d) *Dependence of equilibrium vapour pressure on surface curvature*

Compare situations (a) and (b) in fig. 18.13.

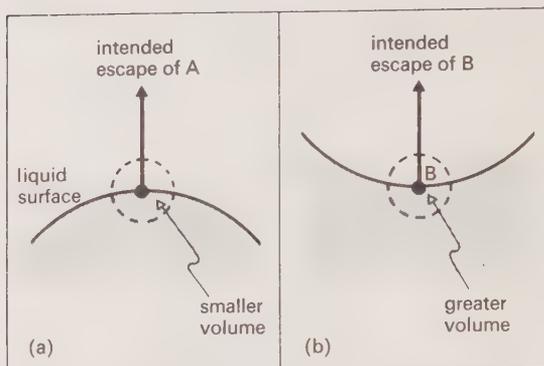


Fig. 18.13. Effect of surface curvature on probability of molecular escape: (a) convex surface, and (b) concave surface.

On attempted escape, molecules A and B experience attraction from the molecules within the spherical surface drawn. The net attractive force on B exceeds that on A. The relative ease of A's escape means there will be more molecules in the vapour, and so the vapour pressure above the convex surface will exceed that above a comparable plane surface, which in turn exceeds that above a concave surface.

Alternatively one may consider the excess pressure within the convex surface as being responsible for a force which encourages A's escape, and vice versa for B.)

(e) The effect of surface tension encourages

- (1) the *evaporation* of small droplets, and
- (2) the *collapse* of small vapour bubbles.

(1) implies that a vapour may be considerably *supersaturated* before condensation occurs (unless some suitable nucleus is available). (2) implies that a liquid may be considerably *superheated* before boiling occurs, in the same way.

(1) and (2) are important in the design of the cloud and bubble chambers respectively (p. 520).

18.9 MEASUREMENT OF THE ANGLE OF CONTACT θ

θ is controlled by the previous history of the liquid surface (fig. 18.14), and is very much affected by the cleanliness of the surface involved.

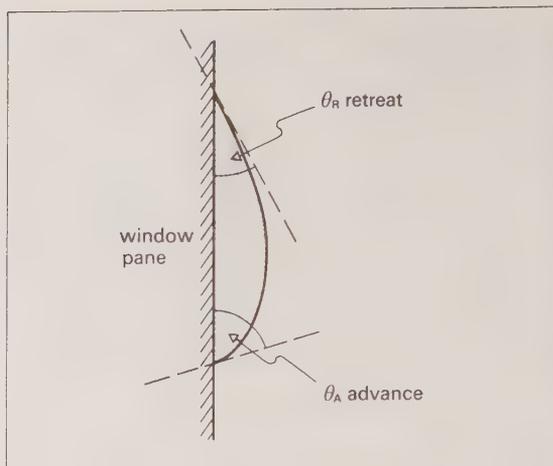


Fig. 18.14. A raindrop illustrating the advancing and retreating angles of contact.

(A) Round-bottom flask method

The liquid is poured into a round-bottom flask (fig. 18.15) until the meniscus is horizontal.

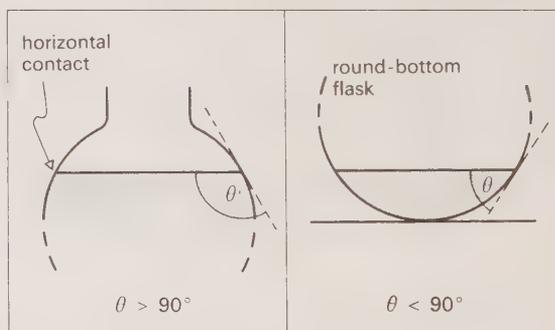


Fig. 18.15. The round-bottom flask method for θ .

(B) Rotating plate method

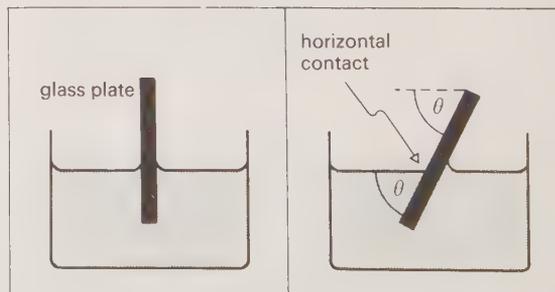


Fig. 18.16. The rotating plate method for θ .

18.10 MEASUREMENT OF SURFACE TENSION γ

No method will produce reliable results unless *absolute cleanliness* is observed throughout.

(A) Chemical balance method

(a) *Liquid which forms a stable film* (e.g. soap solution). Refer to fig. 18.17.

Balance the beam with the soap film in position. Puncture the film, and remove a mass m from the other pan to restore balance.

Weight removed = downward pull of film
(two surfaces)

$$mg = 2\gamma l$$

(b) *For a liquid* (which may not form a stable film) one may pull a cleaned microscope slide or a wire ring from the surface. The method is useful for *comparing* γ for different liquids.

In practice a *torsion* balance is often used in place of a beam balance.

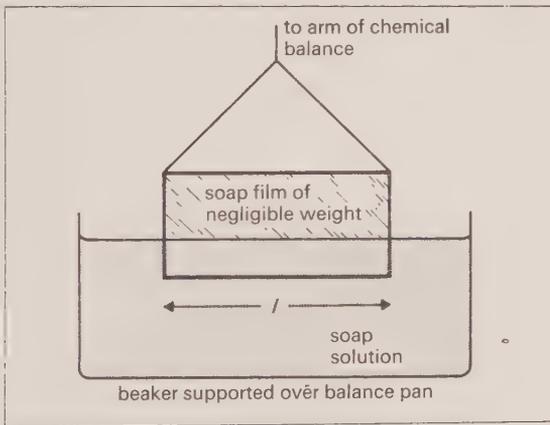


Fig. 18.17. Measurement of γ by direct weighing.

(B) Capillary rise method

Use
$$\gamma = \frac{rg\rho h}{2 \cos \theta}$$
 (p. 157)

This is the international standard method.

Notes

- (a) The tube must be wetted so that θ is known.
- (b) θ must be measured separately.
- (c) h can be measured by travelling microscope (as can r).
- (d) r is measured (after breaking the tube) at the *meniscus level*.

Because mercury shows a depression and is not transparent the method of fig. 18.18 must be adopted.

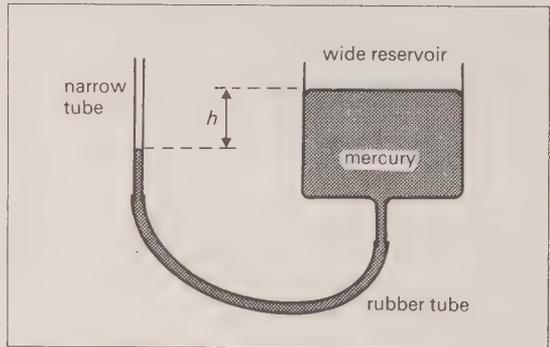


Fig. 18.18. Measurement of capillary depression for mercury.

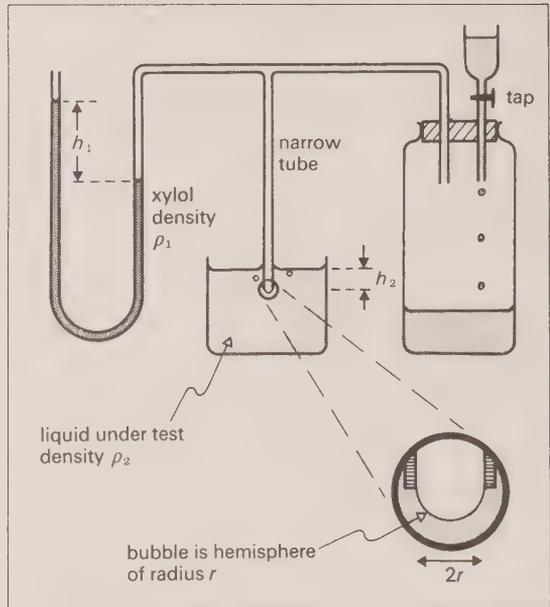


Fig. 18.19. Apparatus for Jaeger's method.

(C) Jaeger's bubble method

Fig. 18.19 shows the apparatus for Jaeger's method, which has the following advantages:

(a) It is easily adapted to investigate how γ varies with temperature.

(b) Contamination is reduced because a fresh liquid surface is formed continually.

(c) A knowledge of the angle of contact is not needed.

h_1 is read just as the bubble breaks away from the narrow tube—this is its maximum value, and corresponds to maximum bubble excess pressure, and so minimum r .

$$\text{Since } A + h_2\rho_2g + \frac{2\gamma}{r} = A + h_1\rho_1g$$

(where A is the atmospheric pressure),

$$\frac{2\gamma}{r} = (h_1\rho_1 - h_2\rho_2)g$$

The proper value of r is near to the radius of the narrow tube. It can be found by calibration (using a liquid whose γ has been found by the capillary rise method).

(D) Other methods

These include

- (a) measurements on large sessile drops,
- (b) measurement of wavelength of ripples (e.g. on a molten metal),
- (c) the drop weight method.

19 VISCOSITY

19.1 THE NATURE OF FLUID FLOW	161
▶ 19.2 DEFINITIONS	162
▶ 19.3 FORCES ACTING ON A LIQUID LAYER	163
▶ 19.4 TURBULENCE AND REYNOLDS'S NUMBER (Re)	163

▶ 19.5 DIMENSIONAL METHODS	164
▶ 19.6 STOKES'S LAW AND TERMINAL VELOCITY	165
▶ 19.7 RESISTANCE OFFERED BY FLUIDS	166
▶ 19.8 EXPERIMENTAL DETERMINATION OF η	167

19.1 THE NATURE OF FLUID FLOW

(A) Viscous and non-viscous flow

Non-viscous flow

When the flow is *laminar*, then the flow pattern can be described by streamlines and tubes of flow. If, in addition, no ordered energy is converted to random internal energy by work being done against viscous forces, then the *Bernoulli Equation* applies (p. 75). It represents the law of conservation of energy adapted for this particular situation.

Viscous flow

Although a fluid at rest cannot *permanently* resist the attempt of a shear stress to change its shape, viscous forces can be brought into play which oppose the *relative motion* between different layers of the fluid. **Viscosity** is thus an internal friction between different layers of a fluid moving with different velocities.

For a viscous fluid, steady external forces produce a steady rate of flow, with no acceleration of any part of the fluid.

(B) Microscopic (molecular) description

Liquids

A liquid differs from a gas in that while neither can permanently resist a shear stress, a liquid does resist both compressive and tensile stresses.

In fig. 19.1, the molecular approach is accompanied by a decrease of intermolecular p.e. and an increase in molecular k.e. If the molecular k.e. becomes disordered, a temporary bond is formed,

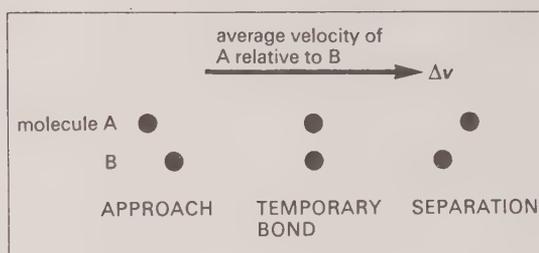


Fig. 19.1. A pair of molecules in adjacent liquid layers with different velocities.

which means that an external agent must do work if the molecules are later to be separated. The work done by the external agent equals the increase in random (internal) energy.

(a) A *temperature increase* means that the molecules have a greater thermal speed, which in turn allows a smaller time in which the gain of molecular k.e. can be disordered (the temporary bond is less strong). *The viscosity of most (but not all) liquids decreases with temperature.*

(b) A *pressure increase* brings the molecules marginally closer, which results in an *increase* in viscosity.

Gases

It is important not to confuse the *drift* momentum of a gas molecule with the momentum associated with its random *thermal* motion.

In fig. 19.2 the *slow* molecule moving up across the boundary CD acquires a larger drift velocity, i.e. gains drift momentum, and so has experienced a force to the right (*Newton II*). This means

it has exerted a force to the left (*Newton III*) which attempts to retard the faster layer.

Similarly the *faster* molecule moving down across CD exerts a force to the right on the slower layer into which it moves.

(a) A *temperature increase* means that the molecules have a greater thermal speed, which increases the rate at which they cross the boundary CD: *the viscosity of a gas increases with temperature.*

(b) Advanced analysis predicts that *the viscosity of a gas should be independent of pressure*, provided the pressure is not too small. This has been verified experimentally, and is strong support for the kinetic theory.

A quantitative treatment is given in chapter 24.

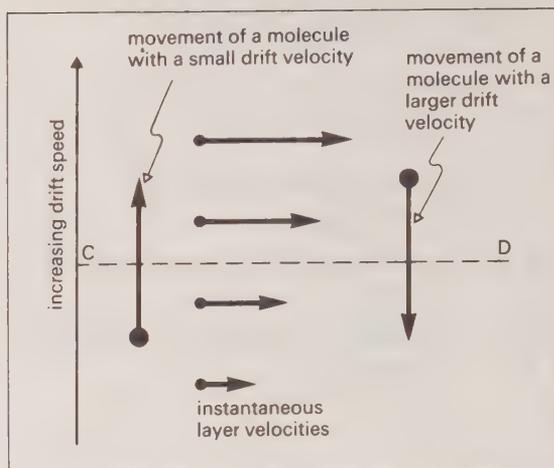


Fig. 19.2. Explanation of viscous forces in a gas.

► 19.2 DEFINITIONS

Tangential stress τ

This is sometimes called **shear**, or **shearing stress**.

Refer to fig. 19.3. The **tangential stress** τ is defined by the equation

$$\tau = \frac{F}{A}$$

$$\tau[\text{Pa}] = \frac{F[\text{N}]}{A[\text{m}^2]}$$

Velocity gradient

In fig. 19.4 the average velocity gradient between the two upper layers is $\delta v/\delta y$. As $\delta y \rightarrow 0$, the particles shown are in adjacent layers.

The **velocity gradient** is defined by the equation

$$\text{velocity gradient} = \frac{dv}{dy}$$

$$\text{velocity gradient} \left[\frac{1}{\text{s}} \right] = \frac{\delta v[\text{m/s}]}{\delta y[\text{m}]}$$

Notes

(a) It has dimensions $[\text{T}^{-1}]$.

(b) Here v increases as y increases, and this makes dv/dy positive.

(c) The velocity gradient is sometimes referred to as the **rate of shear**.

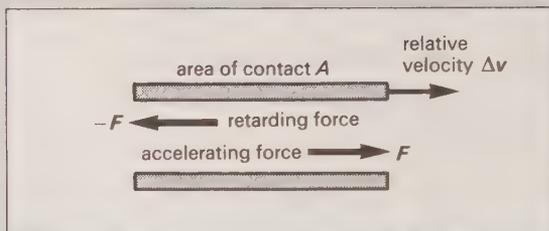


Fig. 19.3. Adjacent layers of a fluid undergoing viscous flow. According to Newton's third law, the two forces F have the same magnitude.

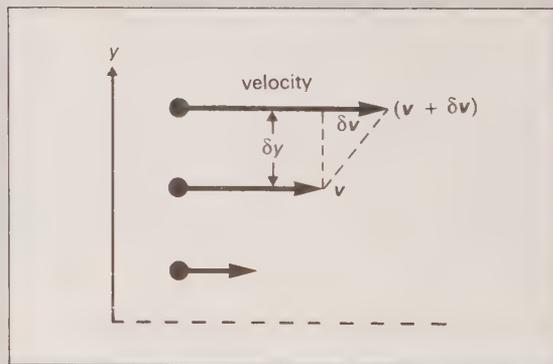


Fig. 19.4. To define velocity gradient.

Coefficient of viscosity η

Fig. 19.5 shows the graph obtained when one plots tangential stress against velocity gradient.

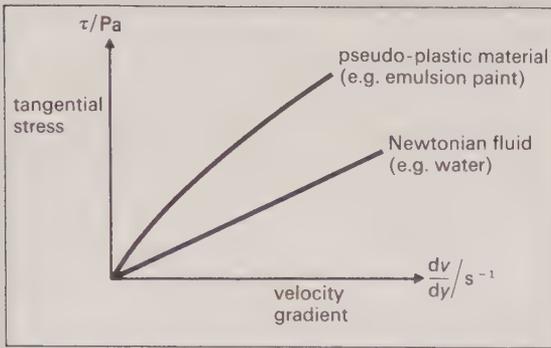


Fig. 19.5. Newtonian and non-Newtonian fluids.

For an ideal gas one can predict a straight line.

Experimentally many pure liquids also give a straight line—they are called **Newtonian** liquids.

We define the **coefficient of viscosity** η by the equation

$$\eta = \frac{\text{tangential stress}}{\text{velocity gradient}} = \frac{F/A}{dv/dy}$$

$$\eta \left[\frac{\text{kg}}{\text{m s}} \right] = \frac{F[\text{N}]/A[\text{m}^2]}{\frac{dv}{dy} \left[\frac{1}{\text{s}} \right]}$$

Notes

(a) The unit of η is $1 \text{ N s m}^{-2} = 1 \text{ kg m}^{-1} \text{ s}^{-1} = 1 \text{ Pa s}$.

(b) The dimensions of η are $[\text{ML}^{-1} \text{T}^{-1}]$.

(c) The coefficient of viscosity of a fluid is analogous to the shear modulus G of a solid:

$$G = \frac{\text{stress}}{\text{strain}} = \frac{\tau}{\gamma}$$

$$\eta = \frac{\text{stress}}{\text{rate of strain}} = \frac{\tau}{\dot{\gamma}}$$

The velocity gradient dv/dy is analogous to the rate of change $\dot{\gamma}$ of the angle of shear (p. 140).

(d) Orders of magnitude (at 293 K):

Substance	$\eta/\text{Pa s}$
air	1.8×10^{-5}
water	1.0×10^{-3}
glycerol	8.3×10^{-1}
golden syrup	1.0×10^2

(e) η is very sensitive to changes of temperature. For pitch

$$\eta = 5 \times 10^{10} \text{ Pa s at 273 K, and}$$

$$\eta = 1 \times 10^1 \text{ Pa s at 373 K.}$$

η^{-1} is called the **fluidity** ϕ of the liquid and indicates the tendency to flow. If E_{visc} is the molar energy required to push past neighbouring molecules,

$$\phi \propto e^{-E_{\text{visc}}/RT}$$

where T is the temperature. $e^{-E_{\text{visc}}/RT}$ is a measure of the fraction of molecules having more than this energy. E_{visc} is about 30% of the energy required for vaporization (involving complete removal of a molecule from its neighbours). The graph of $\ln \phi$ against $1/T$ would give a straight line with a negative gradient equal to $-E_{\text{visc}}/R$. A high value of E_{visc} would lead to a steep gradient and low fluidity (high viscosity).

(f) It is seen from fig. 19.5,

(i) that for *Newtonian* fluids η has a unique value (at a particular temperature),

(ii) that pseudo-plastic materials have an *apparent* coefficient of viscosity whose value decreases as the tangential stress is increased. They are very important in practice, e.g. in the paint industry.

► 19.3 FORCES ACTING ON A LIQUID LAYER

In fig. 19.6 (overleaf) the liquid element experiences three forces:

(a) F_1 is the drag force exerted by a slower layer underneath,

(b) F_2 is the force in the flow direction exerted by a faster layer above,

(c) P is the force that results from the pressure difference Δp across the ends of the fluid element.

For steady (non-accelerated) flow

$$P + F_2 + F_1 = 0$$

Note that $F_1 > F_2$, showing that the velocity gradient is not constant.

► 19.4 TURBULENCE AND REYNOLDS'S NUMBER (Re)

Streamline flow occurs only when the rate of flow is small. The **critical speed** is the speed of bulk flow at which the flow becomes turbulent (fig. 19.7, overleaf).

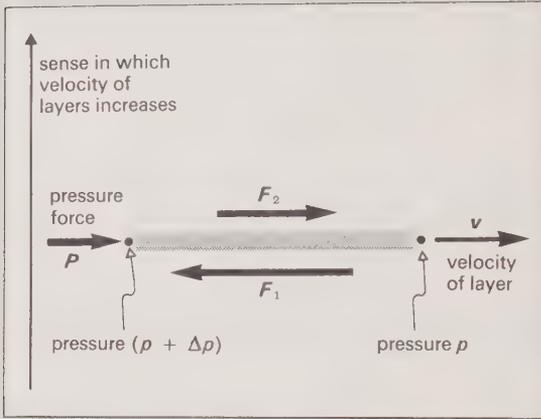


Fig. 19.6. The forces acting on a liquid layer.

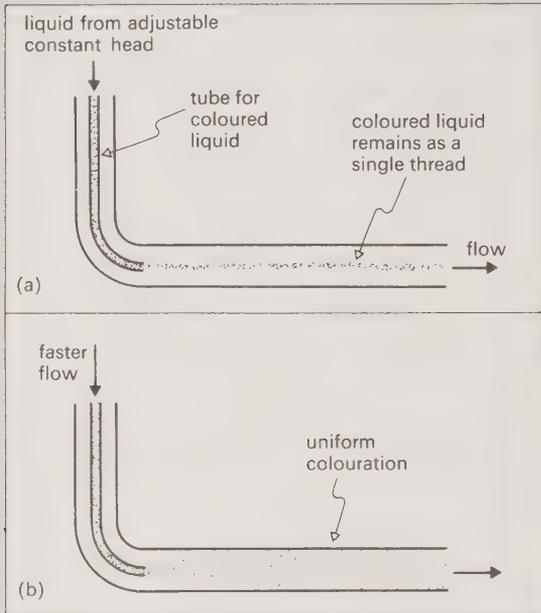


Fig. 19.7. Demonstration of turbulence: (a) streamline flow, and (b) turbulent flow.

Suppose we put $2r =$ tube diameter, $\rho =$ liquid density and $v =$ speed of bulk flow, defined as rate of volume flow divided by cross-sectional area.

Then we define **Reynolds's number** (Re) by

$$(Re) = \frac{2vr\rho}{\eta}$$

It is a *dimensionless* combination of v , $2r$, ρ and η

and (obviously) has no unique value. It is found empirically that for circular tubes, when

- (a) $(Re) < 2200$, the flow is laminar,
- (b) $(Re) \sim 2200$, the flow is unstable,
- (c) $(Re) > 2200$, the flow is usually turbulent.

(Re) is thus a convenient parameter for measuring the stability of flow. The exact value of (Re) for instability depends on the shape of the entrance to the tube.

► 19.5 DIMENSIONAL METHODS

Examples (A), (B) and (C) refer to fig. 19.8.

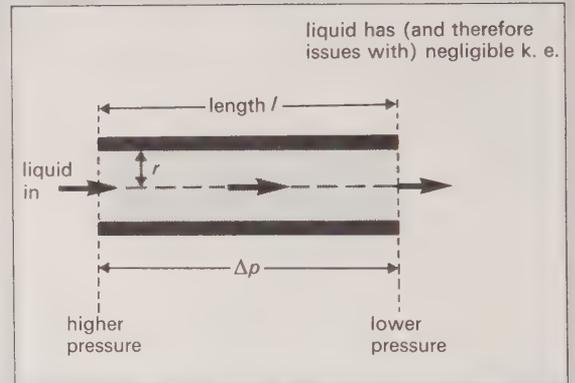


Fig. 19.8. Streamline liquid flow through a horizontal cylindrical capillary tube.

(A) Maximum speed of flow v_{max}

Suppose the speed at the centre is given by

$$v_{max} = k(\eta)^x \left(\frac{\Delta p}{l}\right)^y (r)^z$$

in which

- (a) k is a dimensionless constant,
- (b) $\Delta p/l$ is called the **pressure gradient**,
- (c) we have *not* inserted ρ since the liquid is not accelerated (its mass is not relevant).

Equating dimensions

$$[LT^{-1}] = [ML^{-1}T^{-1}]^x \left[\frac{MLT^{-2}}{L^2L} \right]^y [L]^z$$

from which

$$\begin{aligned} x &= -1 \\ y &= +1 \quad \text{so} \quad v_{max} = k \frac{r^2}{\eta} \left(\frac{\Delta p}{l}\right) \\ z &= +2 \end{aligned}$$

Full analysis shows that

$$v_{\max} = \frac{1}{4} r^2 \left(\frac{\Delta p}{l} \right)$$

(B) Rate of volume flow V/t

Suppose $V/t = k(\eta)^x \left(\frac{\Delta p}{l} \right)^y (r)^z$

Equating dimensions as before, we find

$$x = -1$$

$$y = +1 \quad \text{so} \quad \frac{V}{t} = k \frac{r^4}{\eta} \left(\frac{\Delta p}{l} \right)$$

$$z = +4$$

Note that the rate of volume flow is proportional to

(a) pressure gradient,

(b) r^4 : this seems reasonable because both the cross-sectional area and the average liquid speed are $\propto r^2$,

(c) η^{-1} : this means that a small η gives a large rate of flow. Full analysis shows that $k = \pi/8$, so

$$\frac{V}{t} = \frac{\pi r^4}{8 \eta} \left(\frac{\Delta p}{l} \right)$$

which is called **Poiseuille's Equation**. It can be used for the measurement of η .

(C) Critical speed v_c

Turbulence starts to set in when the speed of bulk flow = v_c . Suppose

$$v_c = k(\rho)^x (\eta)^y (2r)^z$$

in which

(a) ρ is included because turbulence involves a change of momentum,

(b) η is included because a very viscous liquid is less likely to form eddies,

(c) $2r$ is a linear dimension of the channel (e.g. the diameter of a tube).

Equating dimensions shows

$$x = -1$$

$$y = +1 \quad \text{so} \quad v_c = k \left(\frac{\eta}{2\rho r} \right)$$

$$z = -1$$

It follows that

$$k = \frac{2v_c r \rho}{\eta}, \quad \text{but} \quad (Re)_c = \frac{2v_c r \rho}{\eta}$$

so $k = (Re)_c$: the constant is the value that *Reynolds's number* has when the flow is unstable (turbulence about to occur).

$$v_c = (Re)_c \left(\frac{\eta}{2\rho r} \right)$$

(D) Stokes's law

The dimensional part of this law can be deduced by the same method—see the next section.

► 19.6 STOKES'S LAW AND TERMINAL VELOCITY

(A) Derivation of Stokes's law

In fig. 19.9 a viscous drag exists because there exists a velocity gradient between the boundary layer and more distant points in the viscous medium.

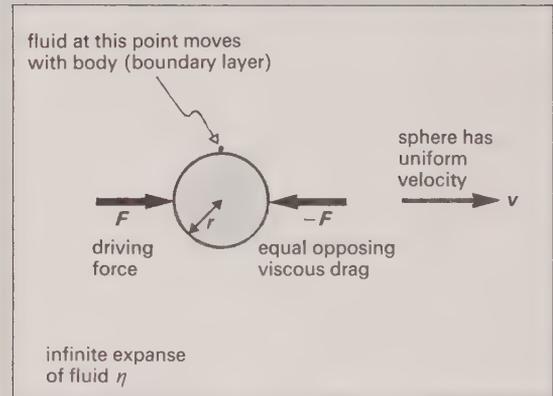


Fig. 19.9. A sphere moving at constant velocity through an infinite viscous fluid.

Suppose $F = k(r)^x (\eta)^y (v)^z$

in which we have not considered ρ because we are not concerned with net changes in the liquid's momentum, even on a local scale, if the flow is streamline.

Equating dimensions, we find

$$x = y = z = +1$$

So $F = k\eta r v$

Full analysis shows that

$$F = 6\pi\eta r v$$

an equation which represents **Stokes's law**.

The resistive force experienced by the body is more correctly given by $-6\pi\eta r v$.

Notes

(a) The equation only applies at small speeds, such that

(i) the conditions are streamline,

(ii) there is no slipping between the liquid and the sphere (the effect of adhesive forces).

(b) Corrections are necessary for a (real) medium of limited extent.

(c) The equation must be modified when the size of the particles of the medium becomes significant relative to the size of the sphere. (It cannot be applied without correction to Millikan's experiment, see p. 471.)

(B) Terminal velocity

When a sphere has reached a terminal velocity of magnitude v_t , as shown in fig. 19.10,

$$G = U + F$$

$$G - U = F$$

$$\frac{4}{3}\pi r^3(\rho - \sigma)g = 6\pi\eta r v_t$$

So
$$v_t = \frac{2r^2(\rho - \sigma)g}{9\eta}$$

Note that $v_t \propto r^2$.

This equation is applicable (with corrections) to

(a) the oil drops in Millikan's experiment,

(b) small water droplets in clouds. (It cannot be used for large rain drops, for which $v_t > v_c$, and so turbulence sets in.)

► 19.7 RESISTANCE OFFERED BY FLUIDS

(A) Low speeds

When the relative velocity is less than the critical velocity, the flow is *streamline*, and *viscous forces* are responsible for the resistance (fig. 19.11). This is sometimes called *skin friction*.

The force
$$F \propto r$$

$$\propto \eta$$

$$\propto v$$

but F is independent of ρ (Stokes's law).

Note that F would be zero in a non-viscous fluid (if it were not for the turbulence that must then ensue).

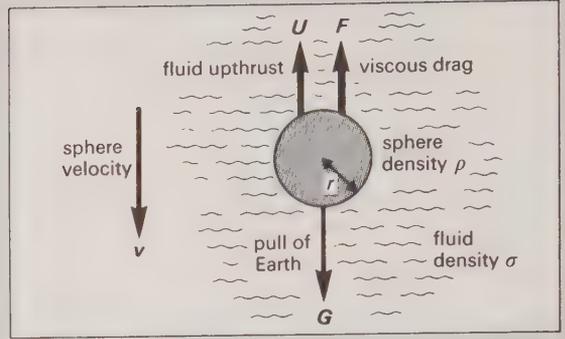


Fig. 19.10. A falling sphere.

(B) High speeds

When the relative velocity exceeds the critical velocity, then streamlines no longer exist and the flow becomes turbulent. The ordered energy (of the fluid flow or body) is converted to the random kinetic energy associated with eddy formation (fig. 19.12).

The force
$$F \propto r^2$$

$$\propto \rho$$

$$\propto v^2$$

but F is independent of η .

If we assume any one of these relationships we can use the method of dimensions to show how F is related to the other three quantities.

We can also say

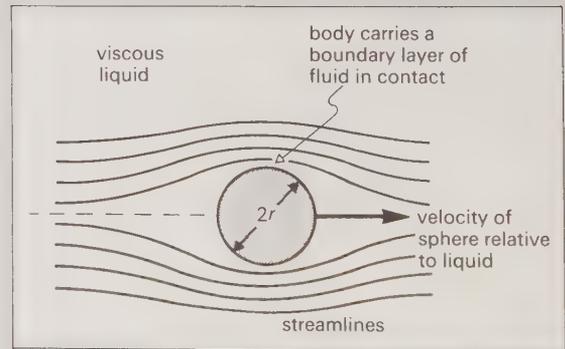


Fig. 19.11. Relative movement of a sphere and a viscous liquid (streamline conditions).

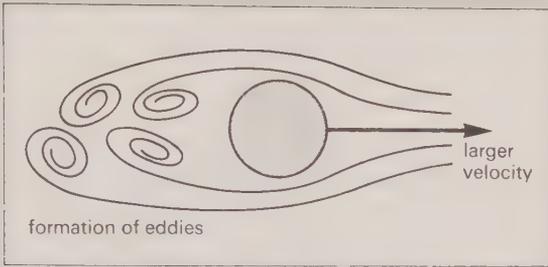
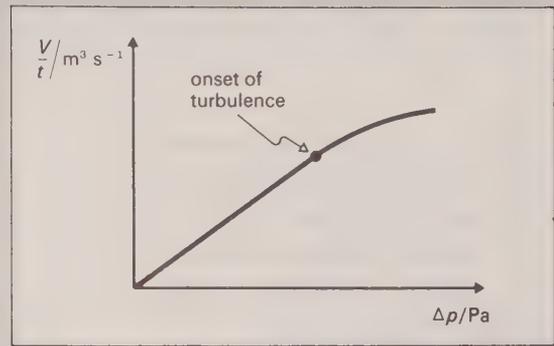


Fig. 19.12. Turbulent conditions.

Fig. 19.13. Ensuring streamline conditions when η is being measured.

$$\begin{aligned}
 F &\propto (\text{rate of change of liquid momentum}) \\
 &\propto \left(\frac{\text{mass of liquid whose velocity}}{\text{is changed in unit time}} \right) \cdot (\text{velocity change}) \\
 &\propto (\rho r^2 v) v \\
 &\propto r^2 \rho v^2
 \end{aligned}$$

The critical value of *Reynolds's number*, and hence the critical velocity, are determined by the *shape* of the body. The purpose of *streamlining* is to reduce the onset of turbulence at the trailing edge.

(C) Intermediate speeds

In (A) and (B) we take the effects of ρ and η respectively to be negligible. In practice F will depend (to some extent) on both ρ and η .

► 19.8 EXPERIMENTAL DETERMINATION OF η

(A) Poiseuille's tube method

$$\text{Use } \frac{V}{t} = \frac{\pi}{8} \left(\frac{\Delta p}{l} \right) \frac{r^4}{\eta}$$

(a) A preliminary experiment indicates the onset of turbulence (fig. 19.13).

For example when $l \sim 0.2$ m and $r \sim 1$ mm, the flow is unstable if Δp is about 9×10^2 Pa (equivalent to a head of water of 90 mm).

(b) A constant-head apparatus is needed.

(c) The k.e. of emerging liquid must be near zero.

(d) r^4 is needed for the calculation, so great care must be taken in measuring r . The mercury weight method does give an average value for r^2 . (Cf. p. 159.)

(e) The liquid temperature must be constant. η varies very rapidly with changes of T .

(B) Stokes's law method

$$\text{Use } F = 6\pi\eta r v$$

(a) The method is useful for viscous liquids like glycerol, for which the flow is still streamline at the terminal velocity v_t .

(b) r must be very much less than the diameter of the container, otherwise the walls will affect the shape of the streamlines. (For the average measuring cylinder we want $r \leq 2$ mm.)

(c) Temperature control is important.

(d) The falling sphere must be freed from clinging air bubbles.

Shore's AL

20 GRAVITATION

<p>20.1 HISTORICAL BACKGROUND 168</p> <p>20.2 NEWTON'S LAW OF UNIVERSAL GRAVITATION 169</p> <p>20.3 THE GRAVITATIONAL FIELD 169</p> <p>▶ 20.4 GAUSS'S LAW 170</p> <p>▶ 20.5 THE FIELDS OF SPHERICAL SHELLS AND SPHERES 171</p>	<p>20.6 THE EARTH'S GRAVITATIONAL FIELD 172</p> <p>20.7 GRAVITATIONAL POTENTIAL 173</p> <p>▶ 20.8 THE POTENTIAL DUE TO A SPHERICAL MASS 174</p> <p>20.9 ORBITAL MOTION 174</p> <p>20.10 TUNNELS AND TIDES 176</p> <p>20.11 MEASUREMENT OF G 176</p>
--	--

20.1 HISTORICAL BACKGROUND

The history of the study of gravitation is an important example of how ideas in physics sometimes develop.

(1) *Copernicus* (1543) suggested that the Sun (and not the Earth) was the centre of the solar system.

(2) *Tycho Brahe* (1590s) made accurate observations on the planetary motions.

(3) *Kepler* (1618) summarized, in **Kepler's Laws**, the kinematic facts accumulated by *Brahe*. These laws indicate briefly important regularities in the planetary orbits.

(4) *Newton* (1686) suggested an *hypothesis* to explain why *Kepler's* laws were obeyed—this was the inspired step of genius. The hypothesis was used to forecast the result of possible measurements, such as

- (a) the relationship between the acceleration of free-fall g_0 at the Earth's surface, and that of the Moon in its orbit,
- (b) the timing of eclipses, and
- (c) the positions of undiscovered planets.

(5) When *Newton's* hypothesis had been thus confirmed by experiment, it was accepted as a *theory*. It is commonly expressed by the equation called **Newton's Law of gravitation**.

(6) Further refinement of instrumental technique led to observations, such as those on the precession of Mercury's orbit, which could not be explained in detail by *Newton's* theory.

We now have a *modified theory*, **Einstein's General Theory of Relativity** (1905), of which *Newton's Law* is a special limiting case.

Suppose we consider a small body moving under a gravitational force near a much more massive body. Then for this situation **Kepler's Laws** can be written:

Law I The Law of Orbits. *The path of the small body is a conic section having the large body as one focus. The particular section is determined by the speed of the small body at a given instant.*

Law II The Law of Areas. *The area swept out in a given time by the radius vector joining the large body to the small body is always the same (fig. 20.1).*

Law III The Law of Periods. *For closed orbits (ellipses) the ratio*

$$\frac{(\text{orbital period})^2}{(\text{semimajor axis})^3}$$

is always the same.

Kepler's Laws were originally stated for the solar system, which exemplifies them. They follow from the application of

- (a) *Newton's Laws* of motion and
- (b) *Newton's Law* of gravitation.

Law II is a consequence of the gravitational force being a *central force*.

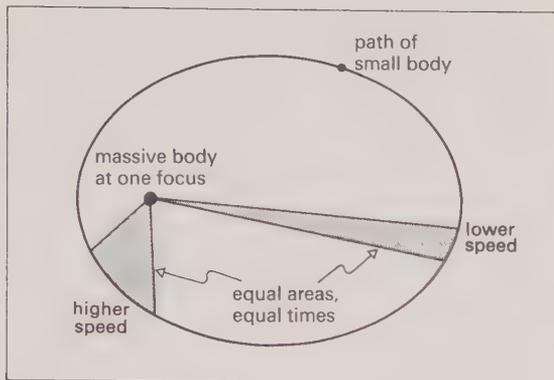


Fig. 20.1. The law of areas.

20.2 NEWTON'S LAW OF UNIVERSAL GRAVITATION

Newton's Law of gravitation (1686) states that

$$F \propto \frac{m_1 m_2}{r^2}$$

where the symbols are defined in fig. 20.2.

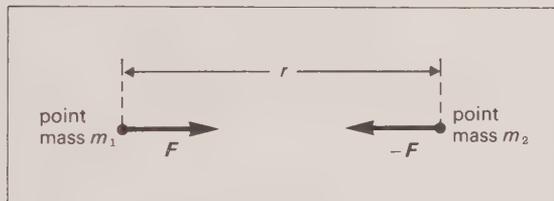


Fig. 20.2. Newton's Law of gravitation.

Notes on the law

(a) Note carefully that the law deals with *point masses*. It can be extended to deal with finite bodies by integration, or by application of *Gauss's Law*.

(b) m represents the **gravitational mass** of the particle. *Inertial* and gravitational masses are identical (or at least proportional), and we will not distinguish between them.

(c) We already have units for F [N], m [kg] and r [m]. We must therefore write

$$F = G \frac{m_1 m_2}{r^2}$$

where G is a constant whose value must be found *experimentally*. (See also p. 337.)

(d) The gravitational force is a **central force**, because it acts along the line joining the points. The gravitational forces that we consider are all *attractive* (p. 336).

The universal constant of gravitation

$$G = \frac{Fr^2}{m_1 m_2}$$

(a) G has a magnitude $6.7 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$ (p. 176).

(b) G has dimensions

$$[G] = \frac{[\text{MLT}^{-2}][\text{L}^2]}{[\text{M}^2]} = [\text{M}^{-1}\text{L}^3\text{T}^{-2}]$$

This shows that G could be expressed in $\text{m}^3 \text{kg}^{-1} \text{s}^{-2}$. This is a less useful form than $\text{N m}^2 \text{kg}^{-2}$.

(c) G is a *universal constant*. Its physical magnitude does not depend on:

- (i) the type of substance considered,
- (ii) the nature of intervening (screening) materials,
- (iii) temperature, etc.

(d) Although G is so small (gravitation being the *weakest interaction*), it becomes an important factor in problems dealing with massive bodies. (Refer to the discussion on the ranges of different types of force (p. 128).)

The small value of G illustrates why gravitation is *not* responsible for interatomic and intermolecular forces (p. 127).

20.3 THE GRAVITATIONAL FIELD

A **gravitational field** is said to exist at a point if a force of gravitational origin is exerted on a test mass placed at that point.

We consider the *interaction* from two viewpoints, as shown in fig. 20.3 (overleaf).

As with electric field (p. 339) and the magnetic field (p. 400) the problem is twofold:

(a) What is the field established by a given mass distribution?

(b) What force is experienced by a particular mass distribution placed in the given field?

In practice there are two factors which simplify the problem enormously:

(i) gravitational forces are mostly negligible unless at least one very massive body is involved, and

(ii) one of the masses concerned is usually very large and possesses spherical symmetry to a high

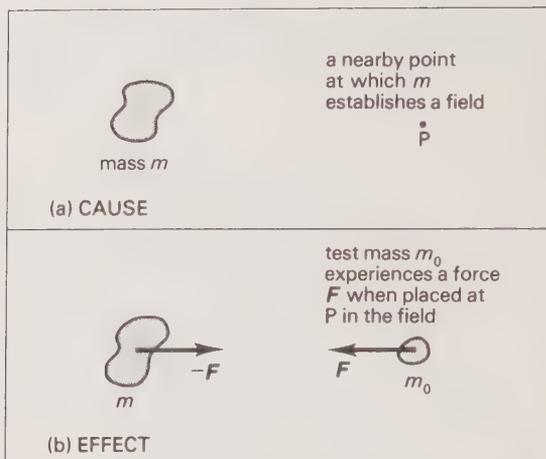


Fig. 20.3. The concept of a gravitational field: (a) the cause of the field, and (b) the effect of the field.

degree, while the other is either similar or so small as to approximate to a particle.

Gravitational field strength

Suppose a test mass m_0 placed at a point in a gravitational field experiences a force F .

The **gravitational field strength** g at that point is defined by the equation

$$g = \frac{F}{m_0}$$

$$g \left[\frac{\text{N}}{\text{kg}} \right] = \frac{F[\text{N}]}{m_0[\text{kg}]}$$

Notes

(a) g represents a vector quantity whose direction is that of the force exerted on the test mass.

(b) We see from *Newton's second law* ($F = ma$) that if the gravitational force is the resultant force on the test mass, then an acceleration g is produced. g has the alternative unit m s^{-2} , and in many circumstances is referred to as the **acceleration due to gravity**.

(c) The gravitational force exerted by the Earth on a small mass is sometimes referred to as its **weight**, but the term is not defined uniquely.

Gravitational field lines

A **gravitational field line** is a convenient fictional concept which can be developed to aid the visualization of a gravitational field. A line is drawn

such that the *tangent* to it at any point gives the direction of g at that point.

Gravitational field lines have the general properties ascribed to electric field lines (p. 340). They are not used widely because most important gravitational fields have the same symmetry (with inward-pointing radial field lines) and are easy to visualize.

A body projected in the Earth's gravitational field follows a trajectory which is very nearly parabolic: this demonstrates that a field line does *not*, in general, display the path taken by a mass in a gravitational field.

► 20.4 GAUSS'S LAW

Gravitational field flux

Suppose a gravitational field g makes an angle θ with the outward normal to a small area ΔA , as shown in fig. 20.4.

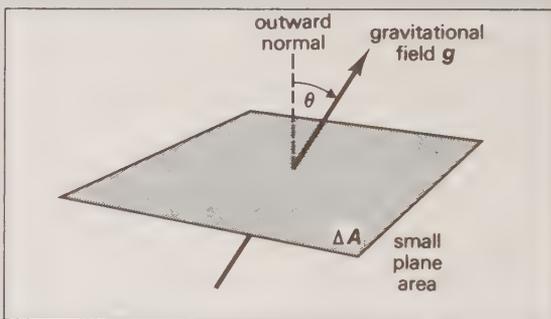


Fig. 20.4. Definition of gravitational field flux ψ_g .

Since ΔA is small, g has effectively the same size and direction at each point on the surface.

The **gravitational field flux** ψ_g through the area ΔA is defined by the equation

$$\psi_g = g \cos \theta \cdot \Delta A$$

$$\psi_g \left[\frac{\text{N m}^2}{\text{kg}} \right] = g \cos \theta \left[\frac{\text{N}}{\text{kg}} \right] \cdot \Delta A[\text{m}^2]$$

ψ_g has the dimensions $[\text{L}^3\text{T}^{-2}]$.

The calculation of ψ_g through an arbitrary surface can become a complex matter (p. 344), but for the situations which we will need to consider g will have the same value at each point on the surface, and θ will usually equal π rad.

Gauss's law

Gauss's law in gravitation states that for any closed surface

$$\psi_g = -4\pi G \Sigma m$$

where Σm represents the total mass enclosed by the surface, and ψ_g is the total gravitational field flux through that surface.

A more detailed discussion of Gauss's law will be found in chapter 44. For the moment we merely note that it is a statement, alternative to Newton's law, for expressing the fact that gravitational force obeys an inverse square law. Its experimental verification is largely indirect, relying on confirmation that the predictions of the law are observed in practice.

The negative sign in the statement of the law is required because we always deal with gravitational forces which are attractive. Field lines point inwards to large isolated masses, causing θ to equal π rad, and so $\cos \theta$ to be -1 . (See below.)

Newton's law deduced from Gauss's law

In fig. 20.5 we want to find the force F acting on a point test mass m_2 placed a distance r from a point mass m_1 . For convenience the Gaussian surface selected is spherical.

Gauss's law is

$$\psi_g = -4\pi G \Sigma m$$

in which

$$\begin{aligned} \psi_g &= \Sigma g \cos \theta \cdot \Delta A \\ &= (g \cos \pi)(4\pi r^2) \\ &= -4\pi r^2 g \end{aligned}$$

and

$$\Sigma m = m_1$$

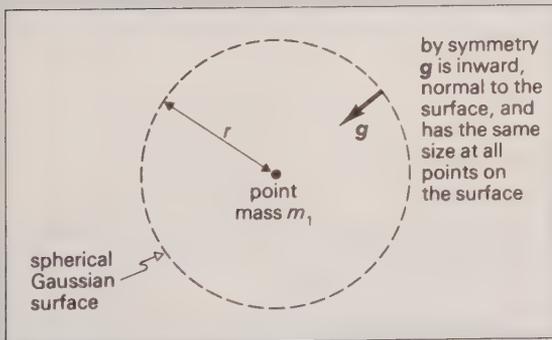


Fig. 20.5. Deduction of Newton's law from Gauss's law.

Substituting

$$\begin{aligned} -4\pi r^2 g &= -4\pi G m_1 \\ g &= \frac{G m_1}{r^2} \end{aligned}$$

As, by definition of g ,

$$F = m_2 g$$

this gives

$$F = G \frac{m_1 m_2}{r^2}$$

which is Newton's law of gravitation.

► 20.5 THE FIELDS OF SPHERICAL SHELLS AND SPHERES

The two results of this section follow closely from Gauss's law. They may also be proved by the integration of Newton's law applied to point masses.

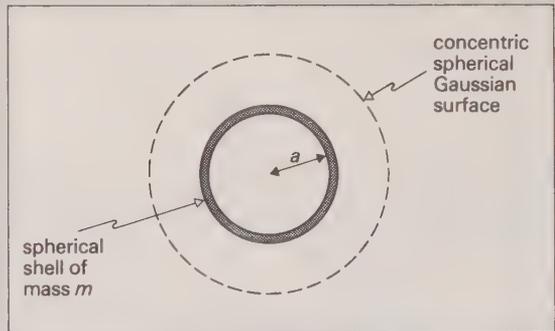


Fig. 20.6. A Gaussian surface of symmetry appropriate to a spherical shell.

The field established by a spherical shell

The thin uniform spherical shell of fig. 20.6 has spherical symmetry. This means that the field it produces

(a) has the same size at each point on the Gaussian surface (the surface being chosen to ensure this), and

(b) cuts the Gaussian surface everywhere at right angles.

We can follow the argument of section 20.4 to the point where

$$g = \frac{Gm}{r^2} \quad (\text{for } r > a)$$

The gravitational field established outside itself by a thin uniform spherical shell is the same as it would be if the mass of the shell were concentrated at its centre.

When $r < a$ our Gaussian surface would appear inside the shell. The argument would then run:

$$\begin{aligned}\Sigma m &= 0 \\ \therefore \psi_g &= 0 \quad (\text{Gauss's law})\end{aligned}$$

The symmetry of the situation would enable us to conclude that

The gravitational field established inside itself by a thin uniform spherical shell is zero everywhere.

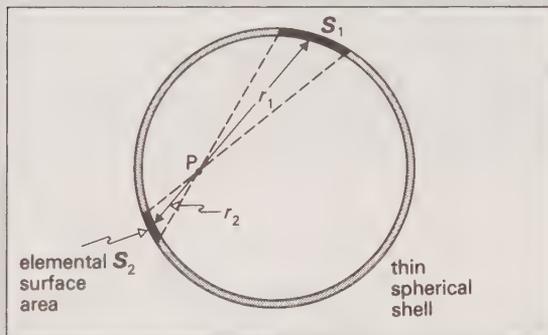


Fig. 20.7. A different viewpoint to investigate the field within a thin spherical shell.

Fig. 20.7 shows this result to be in direct agreement with *Newton's* statement of the law of gravitation, when we remember that geometry requires that

$$\frac{S_1}{r_1^2} = \frac{S_2}{r_2^2}$$

Note that there is *no shielding effect* in gravitation as there is in electrostatics and in magnetism. A gravitational field, established by some second body, *can* exist within a uniform spherical shell.

The fields established by spherical bodies

The conclusions reached for thin spherical shells can be extended simply not only to uniform spheres, but also to other spherical bodies whose density at a given volume element depends only upon the distance of that element from the centre of the sphere (and not upon direction). This is a condition which is satisfied, to a good approximation, by most planets and stars.

The gravitational field established outside itself by a spherically symmetric body is the same as it would be if the mass of the body were concentrated at its centre.

For points inside such a body there is no contribution from that part of its mass which lies at a greater distance from the centre than the point considered, and we apply the result above using the remaining mass.

20.6 THE EARTH'S GRAVITATIONAL FIELD

The general principles established in this section apply to any massive body which possesses the necessary spherical symmetry.

(A) Relation between G and g_0

The sizes of bodies on the Earth's surface are so much less than the Earth's radius R that we may treat them as particles: we will use *Newton's* law.

Consider a body of mass m on the surface of the Earth of mass M . We can equate the value of the gravitational force calculated from *Newton's* law to that found from the definition of g :

$$G \frac{mM}{R^2} = mg_0$$

On the assumption that we cannot distinguish inertial mass from gravitational mass, we have

$$g_0 = \frac{GM}{R^2}$$

A knowledge of G or M enables the other to be found: that is why the measurement of G is known colloquially as 'weighing the Earth'. It also enables us to find a value for the Earth's mean density.

(B) The field outside the Earth ($r > R$)

(a) *Large distances.* The argument in (A) above leads to

$$g = \frac{GM}{r^2}$$

where g is the field strength a distance r from the centre.

► (b) *Small distances.* When $r > R$ we may apply

$$g = \frac{GM}{r^2}$$

$$\begin{aligned}\therefore \ln g &= \ln(GM) - \ln(r^2) \\ &= \ln(GM) - 2 \ln r\end{aligned}$$

Differentiating

$$\frac{\delta g}{g} = -2 \frac{\delta r}{r}$$

If we move a *small* distance h ($\ll R$) away from the Earth's surface

$$\delta g \approx -2 \left(\frac{h}{R} \right) g_0$$

or

$$g \approx g_0 \left(1 - \frac{2h}{R} \right)$$

(C) The field inside the Earth ($r < R$)

Although the Earth has approximately spherical symmetry, its density is non-uniform, varying from about $2.8 \times 10^3 \text{ kg m}^{-3}$ in the crust to $9.7 \times 10^3 \text{ kg m}^{-3}$ at the surface of the core. The average density is $5.5 \times 10^3 \text{ kg m}^{-3}$. This increasing density causes the value of g to continue to *increase* with depth for some distance below the surface.

Suppose, for discussion, that the Earth has uniform density ρ . At a distance r from the centre ($r < R$)

$$g = G \left(\frac{4}{3} \pi r^3 \rho \right) / r^2$$

$$= \left(\frac{4}{3} G \pi \rho \right) r$$

Thus $g \propto r$. At a depth d below the surface

$$g = g_0 (1 - d/R)$$

The results of this paragraph are summarized in fig. 20.8.

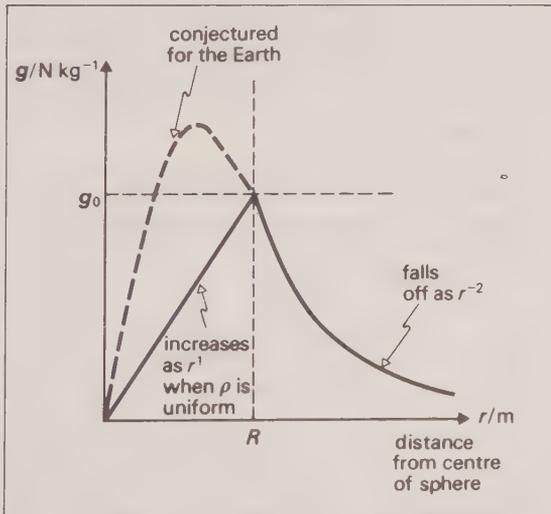


Fig. 20.8. Variation of g established by bodies of spherical symmetry. (To be compared with fig. 45.8.)

20.7 GRAVITATIONAL POTENTIAL

The detailed discussion of electric potential (chapter 45) applies broadly to gravitational potential. A concise summary only of some of the essential ideas is given here.

(a) The **gravitational potential energy** of a system is the work done by an external agent in assembling a particular configuration of masses. It depends upon the sizes of the masses and their relative separations. It is *always negative*, since gravitational forces are attractive.

(b) Suppose an external agent does work W in bringing a test mass m_0 from infinity to a particular point in a gravitational field. (W will of course be negative.)

We define the **gravitational potential** U at that point by the equation

$$U = \frac{W}{m_0}$$

$$U \left[\frac{\text{J}}{\text{kg}} \right] = \frac{W[\text{J}]}{m_0 [\text{kg}]}$$

(c) **Potential differences** are calculated from

$$U_{ab} = U_b - U_a = W_{ab}/m_0$$

(d) The **theoretical zero of potential** is that of a point at infinity, but it is sometimes convenient to treat the surface of the Earth as the *practical zero*. This is valid when (as nearly always) we are dealing with *differences* of potential.

(e) Since the gravitational field is *conservative* (p. 46) gravitational potentials are unique—there is only one possible value for the potential at a given point. This is one reason why gravitational potential is such an important and useful concept.

(f) An **equipotential surface** is one on which the potential is the same at all points—no work is done if a test mass is moved from one point to another on the same surface.

(g) Field strength and potential lend themselves to alternative (but equivalent) ways of describing the same field. We can

- (i) give values of g or U at various points,
- (ii) calculate the forces on, or potential energies of, given masses at a point, and
- (iii) draw field lines or equipotential surfaces.

(h) In a *uniform* field we can relate potential changes and field by

$$\Delta U = g \cdot \Delta x$$

where Δx is a displacement parallel to the field lines. The sign of the potential change can be inserted by inspection.

(i) For a non-uniform field

$$g = -\frac{dU}{dx}$$

The gravitational field strength is numerically equal to the potential gradient (a vector quantity). The negative sign is necessary because the vector g points in the direction along which U decreases.

► 20.8 THE POTENTIAL DUE TO A SPHERICAL MASS

Refer to fig. 20.9. We want to find the work W' done by an external agent that brings a test mass m_0 from infinity to a point distance r from a spherical mass M .

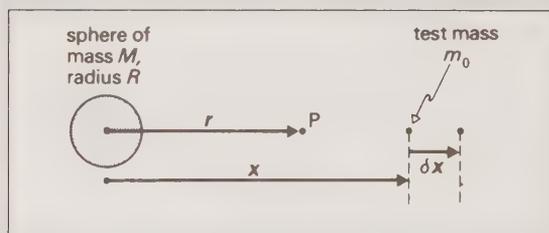


Fig. 20.9. Calculation of the gravitational potential established by a sphere.

The work δW done by the external agent in causing the displacement δx away from the sphere is

$$\begin{aligned}\delta W &= F \cdot \delta x \\ &= \left(G \frac{Mm_0}{x^2} \right) \delta x\end{aligned}$$

To move from P to infinity would require work

$$\begin{aligned}W &= \int_r^\infty G \frac{Mm_0}{x^2} \cdot dx \\ &= \frac{GMm_0}{r}\end{aligned}$$

The work W' done by an external agent in bringing the test mass *from* infinity to P is negative, since the agent would exert a restraining force.

Using $W' = -W$, and the definition $U = W'/m_0$, we have

$$U_P = -\frac{GM}{r}$$

The gravitational potential established inside itself by a spherical shell of radius R has the same value at every point, and is equal to the value at the surface, $-GM/R$.

Application to escape speed

At the surface of a sphere

$$U = -GM/R$$

The p.e. E_p of a mass m at the surface is given by

$$E_p = Um = -GMm/R$$

The p.e. of the mass an infinite distance away is zero. To escape completely from the sphere the mass must gain p.e. GMm/R . If the mass leaves the sphere at speed v_e and just escapes

$$\begin{aligned}\frac{1}{2}mv_e^2 &= \frac{GMm}{R} \\ v_e &= \sqrt{(2GM/R)}\end{aligned}$$

This value of v_e is referred to as the **escape speed**. Since the mass may be moving in any direction above the horizontal the size of v_e has an important bearing on the composition of a planet's atmosphere.

20.9 ORBITAL MOTION

(A) Verification of Kepler's third law

Consider a light body of mass m moving at angular speed ω in a circular orbit of radius r around a massive planet of mass M .

$$\text{Using } F = ma$$

$$\frac{GMm}{r^2} = m\omega^2 r$$

$$\text{So } GM = \omega^2 r^3 = 4\pi^2 \cdot \frac{r^3}{T^2} \quad \left(T = \frac{2\pi}{\omega} \right)$$

Thus the ratio T^2/r^3 is the same for all circular orbits of such light bodies around the parent planet, since G is a constant.

Kepler's third law applies universally, and is illustrated (fig. 20.10) by

- (a) the Sun and its planets (since in fact it applies for all elliptical orbits),
- (b) Jupiter and its moons,
- (c) the Earth and its system of satellites and the Moon.

Kepler's laws were used by *Leverrier* and *Adams* to predict the position of a hitherto unknown planet that was thought to be responsible for the perturbations in the orbit of Uranus. Their work led to the discovery of Neptune in 1846 and to the calculation of its orbital period (165 years).

(B) Satellites

For a satellite in a circular orbit moving at speed v_0

$$\frac{GmM}{r^2} = \frac{mv_0^2}{r}$$

$$v_0^2 = \frac{GM}{r}$$

For a given radius of orbit, there is a particular orbital speed, and vice versa.

- (a) When $r = R$, the radius of the planet,

$$v_0 = \sqrt{\frac{GM}{R}}$$

Thus

$$v_e = \sqrt{2} v_0$$

- (b) If the satellite is given a different speed, its orbit could be (fig. 20.11)

- (i) a *hyperbola* at high speeds,
- (ii) a *parabola* at the escape speed, or
- (iii) an *ellipse* at lower speeds. (A circle is possible: it is an ellipse of zero eccentricity.)

Orders of magnitude for the Earth

$$v_0 = 7.9 \text{ km s}^{-1} \quad (\text{when } r = R_E)$$

$$v_e = 11.2 \text{ km s}^{-1}$$

A space vehicle with a distant destination, such as another planet, may be put into a circular *parking orbit* round the Earth with speed v_0 . The firing of the final rocket then accelerates it to speed v_e in the required direction.

If a satellite has a nearly circular path, but encounters air resistance, there will be a conversion of mechanical energy to internal energy. This reduction of total energy causes a spiral path in which the gravitational attraction of the Earth has a small component in the direction of motion. The

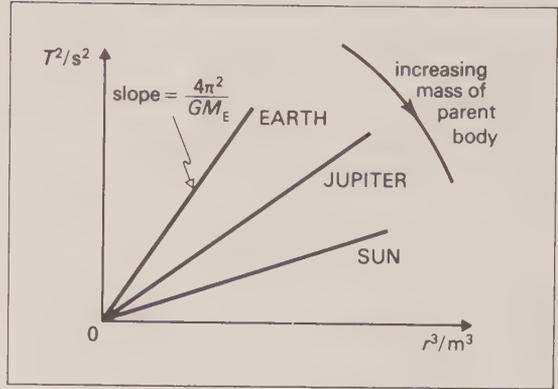


Fig. 20.10. Graphs for three planetary systems which illustrate Kepler's third law. (Not to scale.)

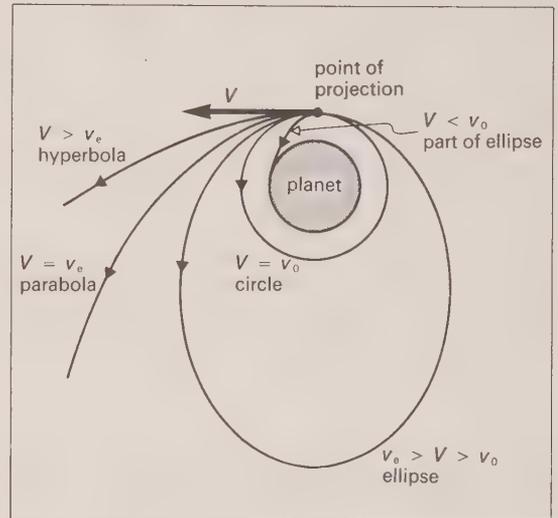


Fig. 20.11. Possible paths for a satellite projected close to the surface of a planet at speed V .

result is an increase in the kinetic energy (see fig. 20.12 (overleaf)) as the satellite spirals down to the Earth's surface.

► (C) Energy and satellite motion

Imagine a satellite of mass m a distance r from the centre of the parent planet of mass M ($M \gg m$).

Since
$$\frac{GmM}{r^2} = m \left(\frac{v^2}{r} \right) \quad (F = ma)$$

the kinetic energy $E_k = \frac{1}{2}mv^2$

$$= \frac{GmM}{2r}$$

which shows that E_k is always positive.

$$\text{But the p.e. } E_p = -\frac{GmM}{r}$$

which shows that E_p is always negative.

The total energy

$$\begin{aligned} E_{\text{tot}} &= E_k + E_p \\ &= -\frac{GmM}{2r} \end{aligned}$$

Our choice of the zero of p.e. makes E_{tot} negative. As r decreases, so E_p decreases, E_k increases (but less rapidly) and so E_{tot} decreases.

These ideas are illustrated by fig. 20.12.

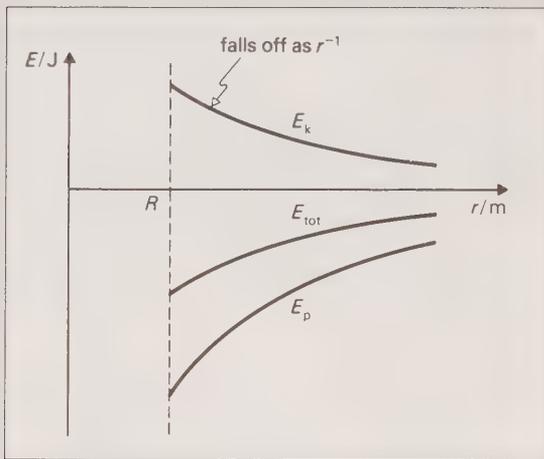


Fig. 20.12. The energies of satellite motion.

(D) Geosynchronous satellites

A **geosynchronous** or **geostationary satellite** rotates about the Earth's axis in the same sense and with the same angular speed as the Earth. The orbital speed v_s and radius r_s of the satellite's orbit are such that it appears to remain above the same point on the equator, thereby making it of vital importance for *telecommunications*. Using the result in (A) above,

$$GM_E = \omega^2 r_s^3$$

where $\omega = \left(\frac{2\pi}{24 \times 3600}\right) \text{ rad s}^{-1}$, it can be shown that $r_s = 42 \times 10^3 \text{ km}$ and $v_s = 3.1 \text{ km s}^{-1}$.

20.10 TUNNELS AND TIDES

► (A) Tunnels through the Earth

Fig. 20.8 shows that a body moving along a tunnel through the Earth (along either a diameter or a chord) would describe s.h.m. It can be shown that the time period would be 84 minutes, the same as that of a satellite in a circular orbit close to the surface of the Earth. (This assumes the Earth's density is uniform.)

(B) The tides

The two tides per day are caused by the *unequal attractions* of the Moon (and Sun) for masses at different sides of the Earth, by the rotation of the Earth about its axis, and by the rotation of the Earth and Moon about their common centre of mass (which lies within the Earth) (fig. 20.13).

The Sun causes two tides per day, and the Moon two tides every 25 hours. These tides exhibit *beats* (p. 114); **spring** tides occur when they are in phase, and **neap** tides when they are out of phase.

A high tide is experienced *after* the Moon has passed overhead.

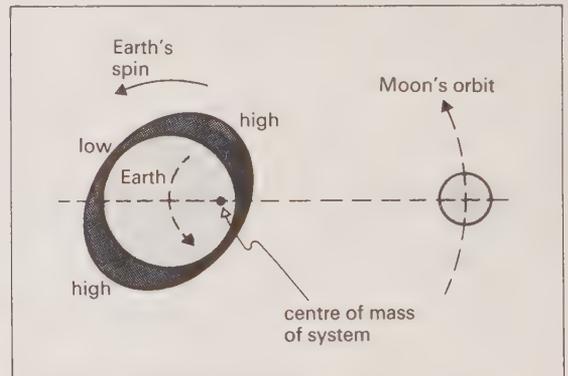


Fig. 20.13. Explanation of formation of tides on the Earth (not to scale).

20.11 MEASUREMENT OF G

Several methods have been used.

(a) *Large-scale methods*, using natural masses, include:

(i) *Bouguer's method* (1740), using a mountain as an attracting mass, and

(ii) in principle, *Airy's method* (1854). This would involve measuring g_0 and g at the bottom of a mine. The experiment is complicated by the fact that the Earth does not have a uniform density.

(b) *Laboratory methods include:*

- (i) *Cavendish's torsion balance method (1798) and its improvement by Boys (1895),*
- (ii) *Poynting's chemical balance method (1893), and*
- (iii) *Heyl's vibrating mass method (1930).*

Cavendish's determination of G

The whole apparatus was enclosed to reduce the effects of air currents. Suppose the suspended system is deflected by an angle θ [rad] by the torque of the gravitational forces. At equilibrium

(gravitational torque) = (elastic restoring torque)

$$\frac{GmM}{r^2}l = c\theta \quad (\text{p. 148})$$

The *suspension constant* c [N m rad⁻¹] was calculated from $T = 2\pi\sqrt{l/c}$ by a separate torsional oscillation experiment (p. 64).

The experiment is clumsy and needs many corrections.

Boys improved *Cavendish's* apparatus by the introduction of fine vitreous silica (quartz) suspension fibres. These have:

- (a) great tensile strength, but
- (b) a very small radius r , and a low shear modulus (therefore small c , since $c \propto r^4$).

They enabled him to reduce the size of the apparatus *without diminishing the sensitivity*. For *Cavendish* l was ~ 2 m, but for *Boys* l was ~ 20 mm.

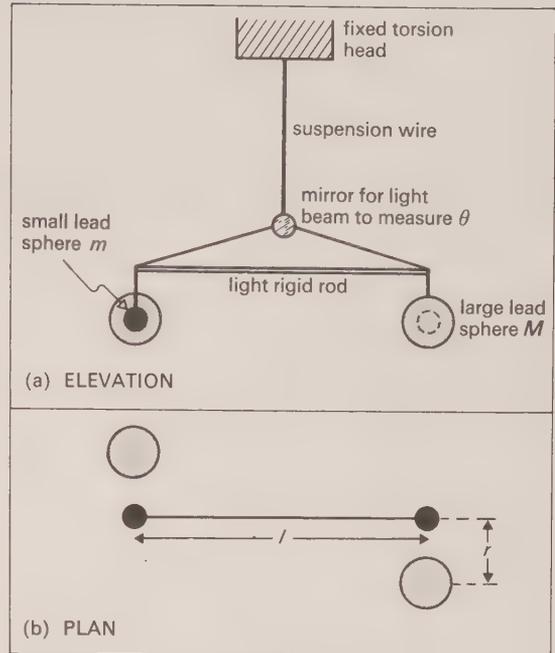


Fig. 20.14. *Cavendish's torsion balance.*

No experiment has yet detected a substance which plays the same screening role in gravitation as that played by conductors in electrostatics, and by ferromagnetic materials in electromagnetism.

V THERMAL PROPERTIES OF MATTER

21	TEMPERATURE	180
22	EXPANSION OF SOLIDS AND LIQUIDS	187
23	HEAT CAPACITY	189
24	IDEAL GASES: KINETIC THEORY	196
25	IDEAL GASES: THERMAL BEHAVIOUR	204
26	CHANGE, DISORDER AND ENTROPY	214
27	CHANGE OF PHASE	225
28	REAL GASES	231
29	THERMAL CONDUCTION	236
30	THERMAL RADIATION	243

21 TEMPERATURE

21.1	THE ZEROETH LAW OF THERMODYNAMICS	180
21.2	DEFINING A TEMPERATURE SCALE	180
21.3	REALIZATION OF A USEFUL TEMPERATURE SCALE	181
21.4	THE CELSIUS TEMPERATURE	182

21.5	AGREEMENT BETWEEN TEMPERATURES MEASURED BY DIFFERENT THERMOMETERS	184
21.6	THERMOMETERS IN PRACTICE	184
21.7	THERMOMETERS COMPARED	186

21.1 THE ZEROETH LAW OF THERMODYNAMICS

Temperature (like mass, length and time) is a *fundamental quantity*, and as such, cannot be defined in terms of simpler quantities (as velocity can be, for example). We can therefore make an *arbitrary choice* as to the *unit* in which to measure temperature, but the procedure is involved.

Suppose that bodies A and B in fig. 21.1(a) are left for a long period of time, so that they reach **thermal equilibrium** (defined on p. 204). The property that becomes common to the two bodies A and B is **temperature**.

The zeroth law of thermodynamics states that if two bodies A and B are separately in thermal equilibrium with a third body C, then A and B are in thermal equilibrium with each other.

C could be a *thermometer*: if C reads the same in both containers, then A and B (of fig. 21.1(b) and (c)) are at the same temperature.

A **thermometer** is a device that measures temperature (using the procedure described in this chapter).

Temperature

(a) Temperature is that *scalar quantity*, sometimes called degree of hotness, that is common to two systems in thermodynamic equilibrium.

(b) Temperature is described by a number (on an arbitrarily chosen scale) such that when two bodies are put into contact, heat flows from the body of higher temperature to that of lower temperature.

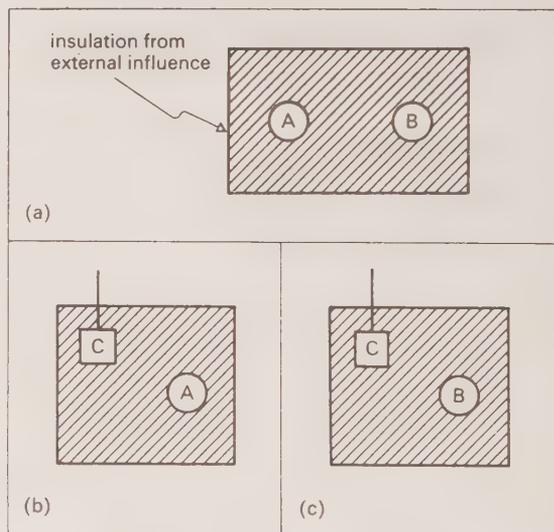


Fig. 21.1. Thermal equilibrium and the zeroth law of thermodynamics.

21.2 DEFINING A TEMPERATURE SCALE

To describe a given degree of hotness by a value (number and unit) which has meaning to different people (i.e. which is *communicable*) we adopt the following procedure:

- (1) We choose a given body of a given substance.
- (2) We select a **thermometric property** of that body—some property whose value varies *continuously* with degree of hotness.

(3) We select two **fixed points** and measure the values of the thermometric property at the lower and upper points, and at the degree of hotness which we wish to measure. Let the measured values of X be X_l , X_u and X_m respectively.

(4) We now allot a value t_m to the degree of hotness, where

$$t_m = \frac{X_m - X_l}{X_u - X_l} \cdot N + t_l$$

in which N is the number of divisions into which we choose to divide the interval between the upper and lower fixed points, and t_l is the temperature of the lower fixed point. The interval is called the **fundamental interval**.

The equation defines the temperature t_m on this particular scale. Note that we have postulated a linear relationship between the thermometric property and the temperature measured on that scale (fig. 21.2). This is justified because the procedure defines temperature.

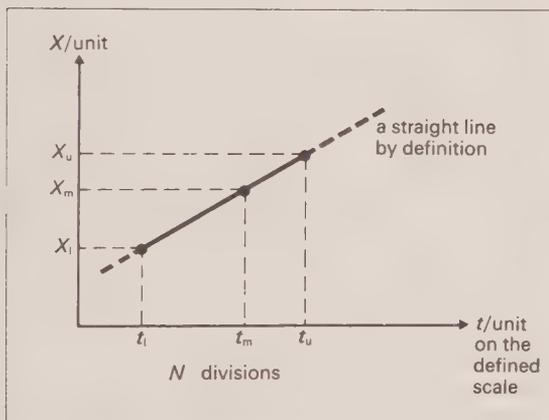


Fig. 21.2. Temperature t defined for some unspecified thermometric property X .

Thermometric properties

The following measurable physical quantities are some of those which prove to be suitable thermometric properties:

- the length of a liquid column in a glass capillary tube,
- the electrical resistance of a platinum wire wound into a coil,
- the pressure of a gas whose volume is kept constant,
- the s.v.p. of a liquid,
- the e.m.f. of a thermocouple,

- the quality ('colour') of electromagnetic radiation emitted by a very hot source,
- the speed of sound through a gas.

Notes

(i) We are entitled to extrapolate our measured values of t_m outside the fixed points, provided we remember that the scale has real meaning only over the range in which the thermometric property is measurable.

(ii) Temperature scales defined according to this procedure do *not* necessarily give the same number for the same degree of hotness (see p. 184 for an example). The numbers will coincide at the **calibration points**.

21.3 REALIZATION OF A USEFUL TEMPERATURE SCALE

It is highly inconvenient to have an indefinite number of arbitrary temperature scales, on which a given degree of hotness is expressed by different values. In practice we agree to use a scale at which we arrive in the following way:

(A) The Kelvin scale

There exists an absolute **thermodynamic scale** proposed by *Kelvin*, which is based on the theoretical efficiency of a perfectly reversible heat engine (26.13). Unfortunately, although this is not dependent on the properties of any particular substance, it is *theoretical* and cannot be put directly to use.

(B) The ideal-gas scale

For this scale the working substance is a fixed mass of ideal gas, whose volume is to be kept at a fixed value. The thermometric property is the pressure p of this gas. This scale is made to coincide with the thermodynamic scale by appropriate choice of fixed points.

(a) The **lower fixed point** is **absolute zero**, to which a value 0 K is allotted. This is the temperature at which the energy of molecular motion becomes a minimum, as a molecule possesses only **zero-point energy**—its *thermal* energy is zero.

(b) The **upper fixed point** is the temperature of the **triple point** of pure air-free water, to which (for historical reasons) a value of 273.16 K is allotted. (This is that unique temperature at which water exists in the solid, liquid and gaseous phases in equilibrium.)

The choice of these two fixed points defines our unit of temperature, the **kelvin** (K) (p. 9).

For our defining equation

$$t_m = \frac{X_m - X_l}{X_u - X_l} \cdot N + t_l$$

$$\begin{aligned} \text{We write } t_m &= T & X_m &= p_T \\ t_l &= 0 \text{ K} & X_u &= p_{tr} \\ N &= 273.16 \text{ K} & X_l &= p_0 = 0 \text{ Pa} \end{aligned}$$

We now define the **ideal-gas temperature** T (equivalent to the thermodynamic temperature) by

$$T = \frac{p_T}{p_{tr}} \times 273.16 \text{ K}$$

This defining equation is consistent with our definition of an ideal gas (on a macroscopic scale) as one which satisfies the equation $pV = \text{constant}$, at constant temperature, for all values of p . It follows that

$$pV \propto T$$

which is illustrated in fig. 21.3.

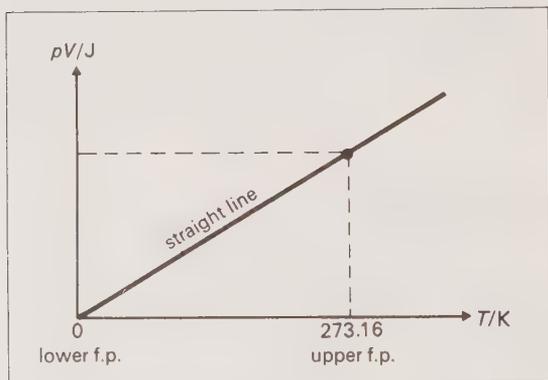


Fig. 21.3. The absolute ideal-gas scale.

(C) Real-gas scales

Although the ideal gas does not exist, it is found in practice that different **real-gas scales** give particularly close agreement over a wide range of temperatures when the gas pressure is low (i.e. when the molecules are well separated). When we extrapolate to zero pressure, the behaviour of a real gas becomes ideal, and the agreement is exact (p. 197).

Real-gas temperature is calculated from

$$T = T_{tr} \lim_{p \rightarrow 0} \frac{(pV)_T}{(pV)_{tr}}$$

in which $T_{tr} = 273.16 \text{ K}$.

In this way a *real constant-volume gas thermometer* is used as a **standard** against which other thermometers can be calibrated. The thermometric substance is a 'permanent' gas (such as hydrogen, helium or nitrogen), whose pressure is the thermometric property. The values that we measure in this way are corrected by a small fraction of a kelvin to obtain the temperature that would be recorded by an ideal-gas thermometer, and so to the thermodynamic scale.

(D) The International Practical Temperature Scale (I.P.T.S. – 1968)

In practice the real-gas thermometer is inconvenient, and so to reduce experimental difficulties there is general agreement to adopt the following procedure.

(a) Using the standard thermometer, temperatures have been assigned to a number of reproducible equilibrium states—these are called the **defining fixed points**, and are available for *calibration* purposes.

(b) A particular instrument is specified for measuring temperatures within a particular range, and the temperature is calculated from a given formula. The diagram on the right shows the general scheme.

(c) Extrapolation is necessary beyond 1338 K:

(i) Optical pyrometers (p. 249) give values based on the *Planck* law of radiation.

(ii) Spectroscopic methods are used at very high temperatures, but the number expressing T depends upon the method.

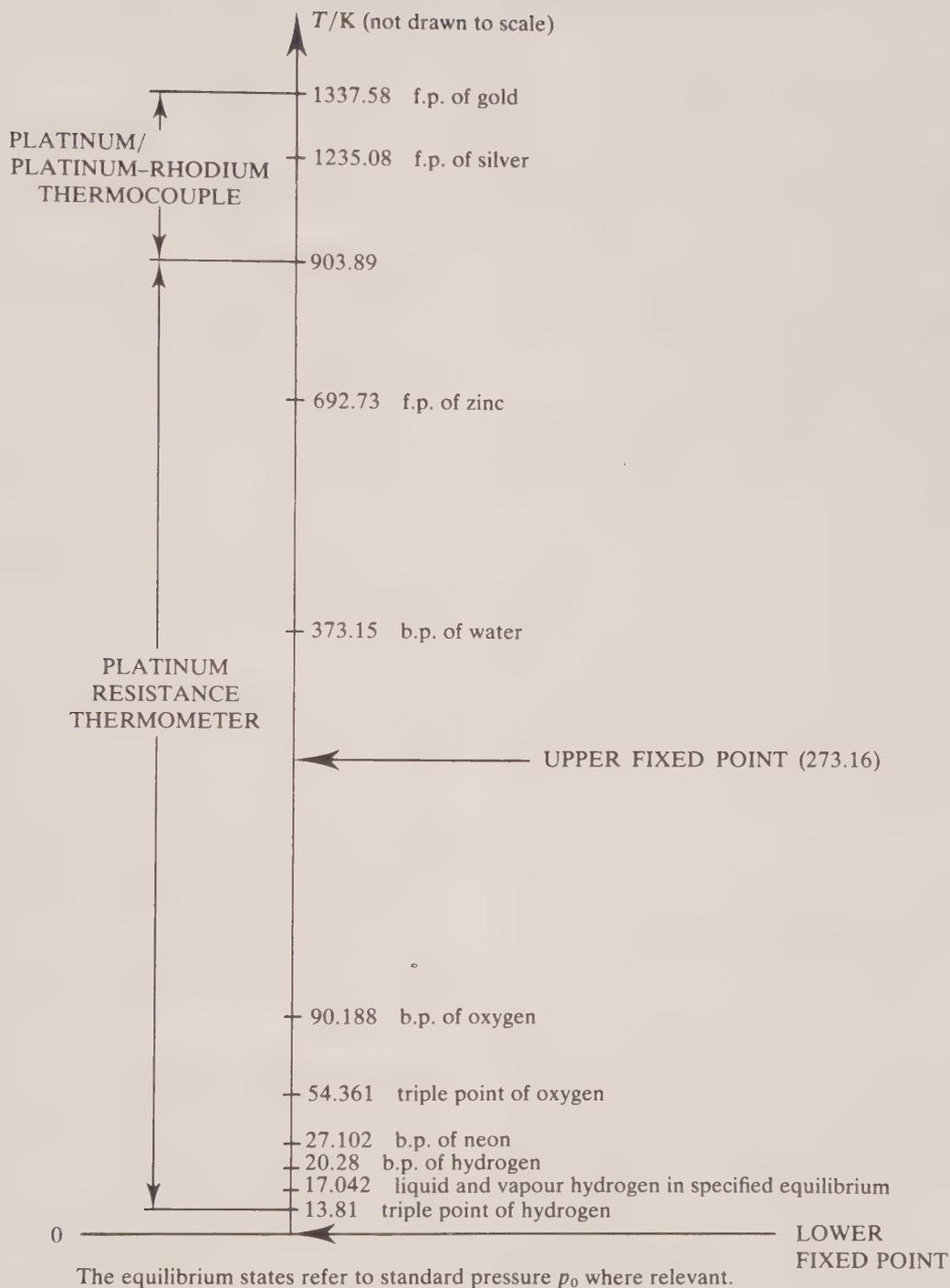
(d) At very low temperatures the speed of sound through helium gas and the vapour pressure of ^4He and ^3He are used as thermometric properties.

21.4 THE CELSIUS TEMPERATURE

Experiment shows

(a) that the **ice point** $T_{ice} = 273.15 \text{ K}$. This is the thermodynamic temperature at which air-saturated water is in equilibrium with ice at standard pressure (101 325 Pa).

(b) that the **steam point** $T_{st} = 373.15 \text{ K}$. This is the temperature at which water vapour is in equilibrium with pure boiling water at standard pressure.



There exists a Celsius thermodynamic temperature interval θ_C , or Celsius temperature, defined by

$$\theta_C / ^\circ\text{C} = T / \text{K} - 273.15$$

Notes on the Celsius scale

(a) The unit on this scale is the **degree Celsius** ($^\circ\text{C}$), which represents an identical temperature interval to the kelvin (K).

(b) At the ice point $\theta_C = 0^\circ\text{C}$, and at the steam point $\theta_C = 100^\circ\text{C}$.

Many mercury thermometers in everyday use are calibrated at these temperatures, and are conveniently marked in degrees Celsius, rather than in kelvins. They have a fundamental interval of 100 degrees Celsius.

(c) Celsius temperatures are called customary or **common temperatures** (to distinguish them from the *thermodynamic* temperatures). [Sometimes they are incorrectly called 'centigrade temperatures'.]

(d) Although the degree Celsius is not an SI unit, in this book we adopt the following policy:

(i) All temperature intervals are in *kelvins* (K).

(ii) Temperature *values* are in degrees Celsius ($^\circ\text{C}$) where they give information students are likely to obtain themselves with a mercury thermometer (but see (iii)). We use the symbol θ .

(iii) Temperature *values* are given in kelvins in tables and in situations where thermodynamic temperatures *must* be used (as for the treatment of gases and in radiation theory). We use the symbol T .

21.5 AGREEMENT BETWEEN TEMPERATURES MEASURED BY DIFFERENT THERMOMETERS

The temperature recorded by one thermometer will *not* (if calculated from the equation defining temperature t on p. 181) *necessarily* agree with a second thermometer, except at the calibration points, where they *must* agree. One would obtain the same temperature from all different thermometers only if all thermometric properties varied with degree of hotness *in the same way*, which they do not.

Example. In fig. 21.4(b) the water vapour pressure thermometer was calibrated at the ice and steam points. When water vapour exerts a pressure of 92 mm Hg, then in

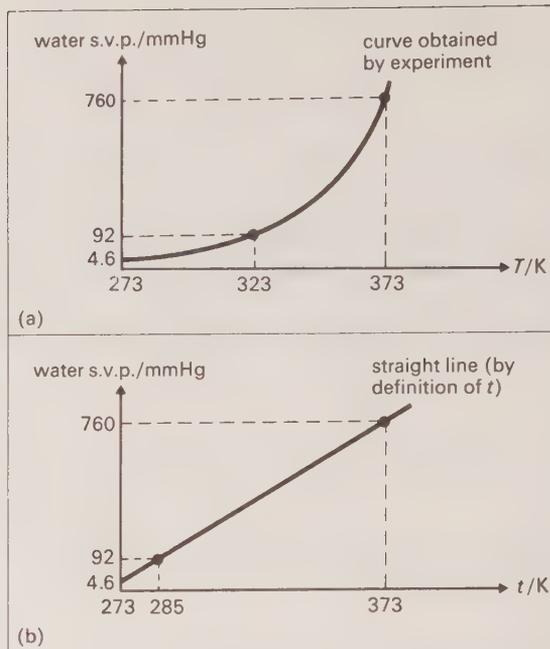


Fig. 21.4. In (a) T is measured on the thermodynamic scale, whereas in (b) t is the temperature value given by a water vapour pressure thermometer. (Not to scale.)

(a) we see that the thermodynamic temperature is 323 K, but in

(b) the water vapour pressure scale temperature is close to 285 K.

285 K is *not wrong*, but the fact that the water vapour pressure scale was used must be quoted explicitly if the value is to be meaningful.

In school laboratories thermometric substances and properties are chosen in such a way that for normal temperature ranges they represent a given degree of hotness by nearly equal numbers, and approximate to the thermodynamic scale.

21.6 THERMOMETERS IN PRACTICE

These descriptions should be read together with the table of section 21.7 (page 186).

(A) The constant-volume gas thermometer

In fig. 21.5 two adjustments are required:

(1) The height of the reservoir is adjusted until the mercury just touches the lower ivory pointer. (This ensures constant volume.)

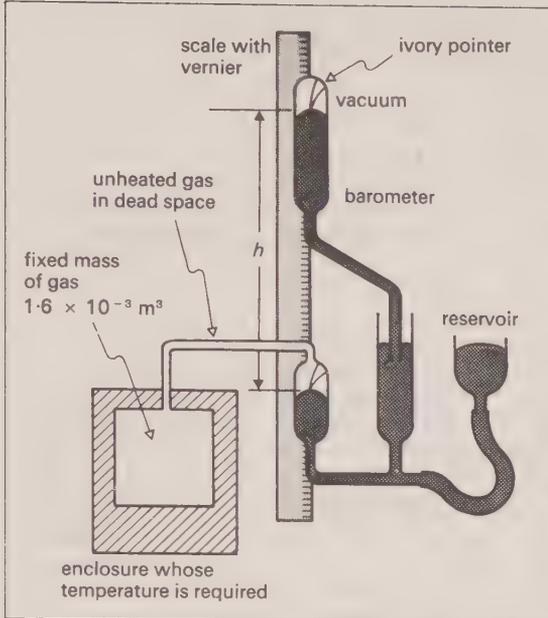


Fig. 21.5. An accurate constant-volume gas thermometer (Harker-Chappuis).

(2) The tube containing the upper ivory pointer is adjusted until the pointer just touches the barometer mercury.

The thermometric property is the gas pressure p , given by $p = h\rho g$.

Corrections are made for

- (i) thermal expansion of bulb,
- (ii) the gas in the dead space not being at the required temperature,
- (iii) capillary effects at glass-mercury contact,
- (iv) the fact that a real gas is used—corrections are made for non-ideal behaviour.

(B) The platinum resistance thermometer

In fig. 21.6 AB is a uniform resistance wire of resistance per unit length r .

Procedure

- (1) Select P accurately equal to Q .
- (2) Position the jockey J at an appropriate

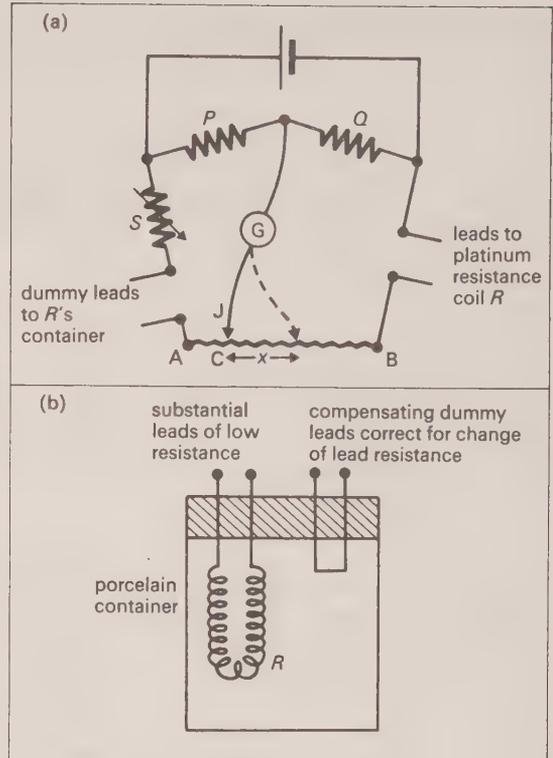


Fig. 21.6. The platinum resistance thermometer: (a) the bridge circuit, and (b) the platinum coil.

point C on AB, and alter S so as to balance the bridge. Suppose $AC = l_1$ and $BC = l_2$.

(3) Put the porcelain container into the enclosure, the temperature of which is required: suppose J has to be moved a distance x to the right to retain balance.

Then because $P = Q$

$$S + (l_1 + x)r = R_T + (l_2 - x)r$$

where R_T is the new coil resistance.

If one assumes a *parabolic* relation between R_T and temperature measured on the I.P.T.S., then calibration at *three* fixed points enables that temperature to be calculated.

(C) The thermocouple thermometer

The *Seebeck effect* is discussed on p. 391.

The resulting thermoelectric e.m.f. can be measured accurately by the potentiometer method described on p. 397.

21.7 THERMOMETERS COMPARED

Name of thermometer	Thermometric property	Maximum temperature range/K	Advantages	Disadvantages	Other comments
(a) Mercury-in-glass	length of mercury column in capillary tube	234 to 630	(i) portable (ii) direct reading	(i) relatively small (ii) not very accurate	(i) the upper range can be extended beyond the normal b.p. (ii) calibrated for accurate work by reference to (b)
(b) Constant-volume gas	pressure of a fixed mass of low-density gas maintained at constant volume	3 to 1750	(i) very wide range (ii) very accurate (iii) very sensitive	(i) cumbersome (the bulb may have a capacity of about 10^{-3} m^3) (ii) very inconvenient and slow to use	(i) the <i>standard</i> by which others are ultimately calibrated (ii) He, H ₂ and N ₂ are used for different temperature ranges (iii) can be corrected to the <i>ideal-gas</i> scale
(c) Platinum resistance	electrical resistance of a platinum coil	13.8 to 1400	(i) wide range (ii) best for small steady differences of temperature (iii) the most accurate thermometer in the range 13.8 K to 904 K	not suitable for varying temperatures because of (i) large heat capacity, and (ii) use of materials of low thermal conductivity in sheath	using an agreed formula, it is used to define the temperature from 13.8 K to 904 K on the international practical temperature scale (I.P.T.S.)
(d) Thermo-couple	e.m.f. of two wires of dissimilar metals	25 to 1750	(i) wide range (ii) very small, so useful for both local and rapidly varying temperatures (iii) the most accurate thermometer in the range 904 K to 1338 K	not as accurate as (c) in the range 13.8 K to 904 K	(i) using an agreed formula, it is used to define the temperature from 904 K to 1338 K on the I.P.T.S. (ii) different pairs of metals are chosen for different temperature ranges
(e) Radiation pyrometer (see p. 249)	quality of electromagnetic radiations emitted by hot body	beyond about 1250	the only instrument capable of measuring temperatures above about 1750 K	(i) cumbersome (ii) not direct reading	(i) defines I.P.T.S. temperatures above 1338 K (ii) readings above 1750 K are calculated from <i>Planck's</i> radiation laws

In considering their relative merits, we should remember that

(a) a thermometer absorbs heat from the substance whose temperature it is measuring,

(b) a thermometer records its own temperature,

(c) there may be a time lag before thermal equilibrium is reached between a thermometer and its surroundings.

22 EXPANSION OF SOLIDS AND LIQUIDS

22.1 MOLECULAR BEHAVIOUR	187
22.2 LINEAR EXPANSION	187

22.3 VOLUME EXPANSION	188
-----------------------	-----

22.1 MOLECULAR BEHAVIOUR

As the common temperature θ increases, so does the energy and amplitude of the molecular vibration. The asymmetry of the curve of fig. 22.1 (departure from true parabolic shape) indicates

- (a) that the molecule does not describe s.h.m., and
- (b) that r increases as θ increases.

This microscopic behaviour is magnified on a macroscopic scale, and appears as an increase in separation of *any two points* in an isotropic homogeneous solid.

Substances that shrink with temperature increase do not have a E_p-r curve of this shape.

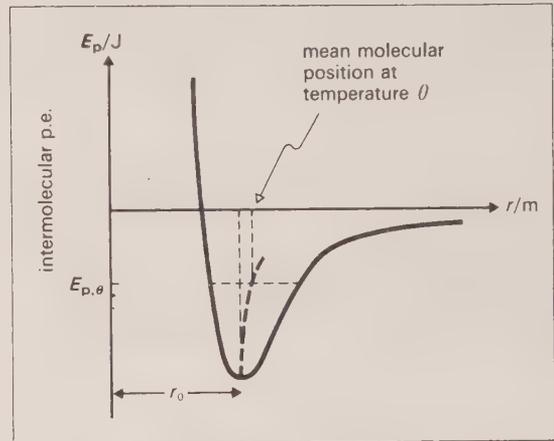


Fig. 22.1. Variation of molecular p.e. E_p with nuclear separation r .

22.2 LINEAR EXPANSION

For a change of common temperature $\Delta\theta$, experiment shows (fig. 22.2) that

- (a) $\frac{\Delta l}{l_0} \propto \theta$ approximately, and
- (b) $\frac{\Delta l}{l_0 \Delta\theta}$ depends on the substance under test,

where Δl is the change of length resulting from $\Delta\theta$.

The **linear expansivity** α of a substance is defined by the equation

$$\alpha = \frac{\Delta l}{l_0 \Delta\theta}$$

$$\alpha \left[\frac{1}{K} \right] = \frac{\Delta l [m]}{l_0 [m] \Delta\theta [K]}$$

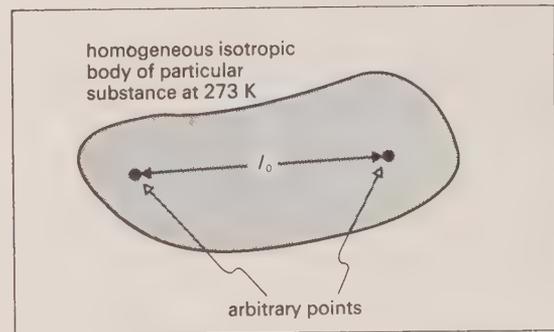


Fig. 22.2. Thermal expansion.

It follows that

$$l_\theta = l_0(1 + \alpha\theta)$$

where θ is the body's common temperature.

Although α is defined relative to l_0 measured at 273 K, it varies only a little with temperature, and we can usually justify the approximation

$$l_2 = l_1 \{1 + \alpha(\theta_2 - \theta_1)\}$$

when $(\theta_2 - \theta_1)$ is of the order of 100 K. (But note from fig. 22.1 that $\alpha \rightarrow 0$ at low temperatures, since the energy of vibration is then small.)

α for a metal is about 10^{-5} K^{-1} . To measure α we use

$$\alpha = \frac{\Delta l}{l_0 \Delta \theta}$$

l_0 and $\Delta \theta$ are easily measured. Δl is very small, so we either

(a) magnify it by a device such as an optical lever, or

(b) use a very sensitive measuring instrument (as in the *comparator* method).

The change ΔA of a body's area can be calculated from

$$\Delta A = A_0 \beta \Delta \theta$$

where β , the material's **superficial expansivity**, is approximately 2α .

22.3 VOLUME EXPANSION

The **cubic expansivity** γ of a substance is defined by

$$\gamma = \frac{\Delta V}{V_0 \Delta \theta}$$

$$\gamma \left[\frac{1}{\text{K}} \right] = \frac{\Delta V [\text{m}^3]}{V_0 [\text{m}^3] \Delta \theta [\text{K}]}$$

By considering the linear expansion of the sides of a cube, it is easily shown that $\gamma \approx 3\alpha$ (since $\alpha^2 \ll \alpha$).

For a typical metal α and $\gamma \sim 10^{-5} \text{ K}^{-1}$, and so $\alpha^2 \sim 10^{-10} \text{ K}^{-2}$.

Comments on volume expansion

(a) The equation $V_\theta = V_0(1 + \gamma\theta)$ is useful.

(b) Hollow bodies of a given material (such as a relative density bottle) expand without distortion as though they were solid. Their internal and external dimensions change in the same proportion.

(c) The term *linear expansivity* has no meaning for a liquid or gas, since a fluid takes the shape of its container.

(d) When the temperature of a liquid changes, the internal volume of its container will also change, and one must distinguish between the **real and apparent cubic expansivity**.

Variation with temperature of density

For a fixed mass m of a substance

$$m = V_1 \rho_1 = V_2 \rho_2$$

$$\therefore \frac{\rho_1}{\rho_2} = \frac{V_2}{V_1} = 1 + \gamma(\theta_2 - \theta_1)$$

Pressure changes in liquids are given by

$$\Delta p = \Delta h \rho g \quad (\text{p. 72})$$

A comparison of pressures enables densities to be compared and so γ can be evaluated. This is the basis of the *balancing column* method.

(Remember that water is anomalous in that *its density increases* when it is heated from its melting point 273 K to 277 K and then decreases on further heating.)

23 HEAT CAPACITY

23.1 HEAT AND WORK	189	23.5 (B) CONTINUOUS-FLOW METHOD	192
23.2 DEFINITIONS	189	23.6 (C) A COOLING METHOD	192
23.3 GENERAL NOTES ON CALORIMETRY	190	23.7 THE EFFECT OF PRESSURE ON SPECIFIC HEAT CAPACITIES	194
<i>Methods of measuring c</i>			
23.4 (A) ELECTRICAL METHODS	191	23.8 MOLAR HEAT CAPACITY OF ELEMENTS	194

23.1 HEAT AND WORK

Work W is energy that is transmitted from one system to another when each exerts on the other macroscopic forces whose points of application are moved through finite distances.

When two parts of a body are at different temperatures the molecules in different regions will have different average vibrational energies, and this will enable the more energetic molecules to do work (but on a microscopic scale) on the less energetic.

Heat Q is that energy that moves from one point to another because there exists between them a temperature difference.

Conduction and (less obviously) radiation are both methods of energy transfer that involve work being done on a microscopic scale. Work and heat thus both represent different forms of energy in transit, and both are measured in joules.

Once energy has been absorbed by a system we have no way of telling whether it was transferred as work or heat. Within the system it is referred to simply as *internal energy*.

The **internal energy** U of a system is the sum total of the kinetic and mutual potential energies of the particles that make up the system.

The relationship between W , Q and U is discussed on p. 207 under the first law of thermodynamics.

23.2 DEFINITIONS

Suppose heat ΔQ added to a body causes an increase $\Delta\theta$ in its common temperature θ .

The **heat capacity** C of the body is defined by

$$C = \frac{\Delta Q}{\Delta\theta}$$

$$C \left[\frac{\text{J}}{\text{K}} \right] = \frac{\Delta Q[\text{J}]}{\Delta\theta[\text{K}]}$$

The value of C may depend on the temperature θ in the vicinity of which $\Delta\theta$ is measured: strictly we measure only an *average* value.

(The word **specific** used in front of a physical quantity should always mean 'per unit mass' (e.g. kg^{-1}). An example of this use follows.)

Specific heat capacity is the heat capacity of unit mass of a *substance* making up a body. Suppose we add heat ΔQ to a mass m of a substance, and that this causes an increase $\Delta\theta$ in its common temperature.

The **specific heat capacity** c of the substance is defined by the equation

$$c = \frac{\Delta Q}{m\Delta\theta}$$

$$c \left[\frac{\text{J}}{\text{kg K}} \right] = \frac{\Delta Q[\text{J}]}{m[\text{kg}]\Delta\theta[\text{K}]}$$

Notes

(a) The value of c may depend on the temperature θ in the vicinity of which $\Delta\theta$ is measured. If $\Delta\theta = (\theta_2 - \theta_1)$, then we measure only the *average* value over that range. As $\Delta\theta \rightarrow 0$, then we have measured c at the temperature $(\theta_1 + \theta_2)/2$.

(b) The equation

$$\Delta Q = mc \Delta\theta$$

is a useful relationship that follows from the definition. When the temperature is close to 300 K, or above, we usually assume that c remains constant over the range $\Delta\theta$.

(c) $C = \Delta Q/\Delta\theta = mc$ is a useful relationship for calculating the heat capacity C .

(d) For gases the **molar heat capacity** C_m (i.e. heat capacity *per mole*) is a more useful quantity than *specific* heat capacity. It is defined on p. 207.

(e) The dimensions of c are $[L^2T^{-2}\Theta^{-1}]$. Those of C_m are $[ML^2T^{-2}\Theta^{-1}N^{-1}]$.

(f) Some orders of magnitude for c and C_m are given in the table opposite.

Orders of magnitude

Substance	Temperature of measurement	$c/\text{J kg}^{-1} \text{K}^{-1}$	$C_m/\text{J mol}^{-1} \text{K}^{-1}$
helium (at constant pressure)	10–3000 K	5.2×10^3	21
copper	23 K	1.4×10^1	0.9
copper	293 K	3.9×10^2	24
lead	293–373 K	1.3×10^2	26
water	288 K	4.2×10^3	75

23.3 GENERAL NOTES ON CALORIMETRY

Summary of available methods

Method	Can be used to find (specific) heat capacity of		
	Solid	Liquid	Gas
(a) electrical	yes	yes	$C_{V,m}$
(b) continuous-flow (usually electrical)	no	yes	$C_{p,m}$
(c) cooling	no	yes	no
(d) change of phase	yes	yes	$C_{V,m}$
(e) mixtures	yes	yes	$C_{p,m}$

Methods (a), (b) and (c) are described in this chapter.

Minimizing the effects of heat loss

Several techniques are available, which include

(a) Polishing to reduce radiation loss; an outer container to reduce convection and conduction loss, and insulating supports to reduce conduction.

(b) Arranging (by preliminary experiment) for the initial temperature to be as much below room temperature as the final one will be above.

(c) Surrounding the calorimeter with an electrically heated jacket at calorimeter temperature. A thermocouple can be used to operate the heater circuit switch.

(d) A simple cooling correction:

Record θ throughout the experiment, and plot a $\theta-t$ graph (fig. 23.1).

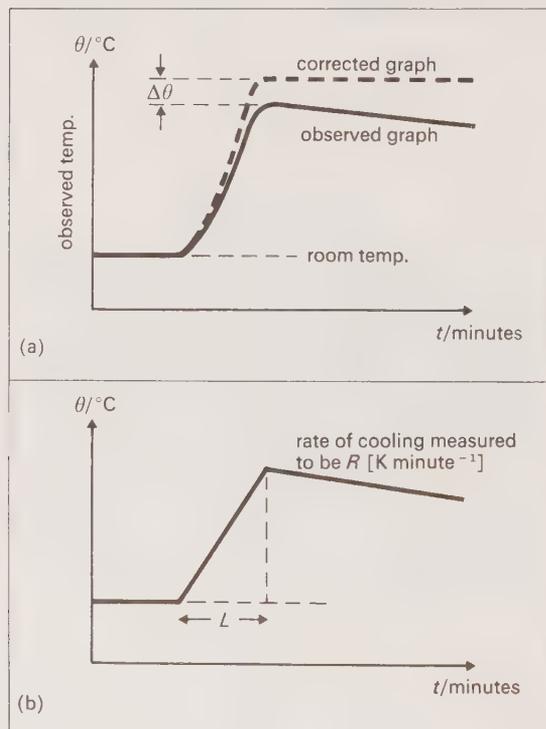


Fig. 23.1. Graphs for a simple cooling correction; (a) a real $\theta-t$ curve, and (b) an idealized simplification.

Suppose the rate of cooling at the maximum observed temperature is R . Then the average rate of cooling over the heating period $L \approx R/2$.

Thus the *correction* to the observed maximum is given by

$$\Delta\theta[\text{K}] = \frac{1}{2}R \left[\frac{\text{K}}{\text{min}} \right] L[\text{min}]$$

For this correction we assume that the rate of loss of heat is proportional to the excess temperature. This procedure is justified for heat loss by

- (a) conduction (see p. 236),
- (b) convection, provided either
 - (i) the convection is forced, or
 - (ii) the temperature excess is small (p. 193), and
- (c) radiation, provided the temperature excess is small (p. 248).

Methods of measuring c

23.4 (A) ELECTRICAL METHODS

Introduction

The electrical energy converted to thermal energy in the resistor is given by

$$W[\text{J}] = V[\text{V}]I[\text{A}]t[\text{s}]$$

W can be found since V , I and t can all be measured both accurately and conveniently.

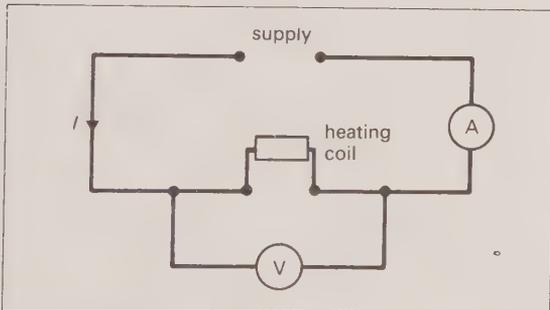


Fig. 23.2. The basic circuit.

Nernst's method for solids

Can be used for solids that are good conductors (fig. 23.3). Special features of the method are:

- (a) the solid needs no calorimeter,
- (b) the heating coil can be used as a resistance thermometer.

In the equation

$$\Delta Q = mc \Delta\theta = Vit$$

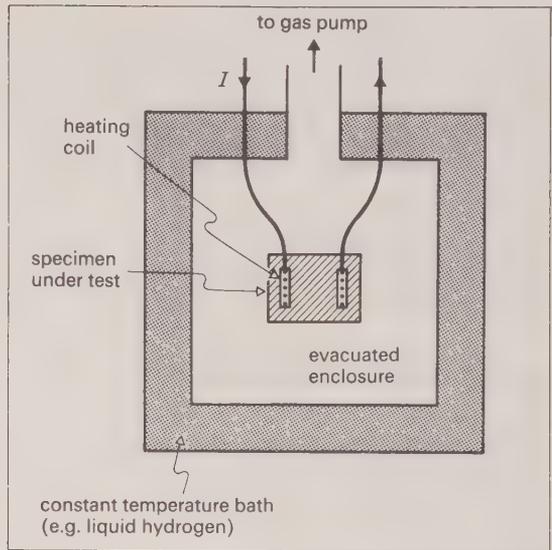


Fig. 23.3. The Nernst calorimeter.

we measure m , $\Delta\theta$, V , I and t for a small range $\Delta\theta$ (which can be as small as 10^{-3} K) near a particular θ .

Hydrogen (a good gaseous conductor) can then be admitted to the evacuated enclosure to change the temperature, the enclosure again evacuated, and the process repeated.

A vacuum flask method for liquids

(1) Fill the flask (fig. 23.4) to a known depth with liquid (such as water) of known specific heat capacity c_1 .

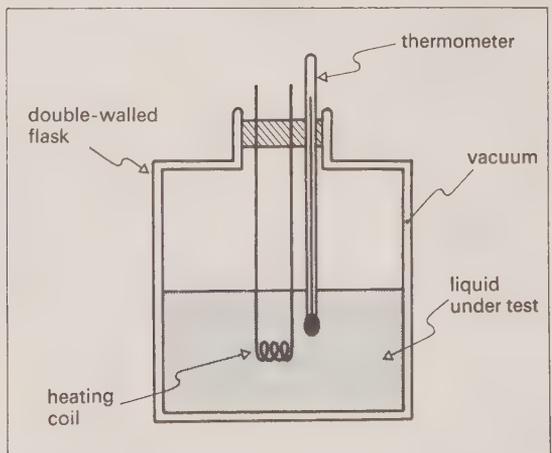


Fig. 23.4. Using a vacuum flask to measure the specific heat capacity of a liquid.

$$V_1 I_1 t = m_1 c_1 \Delta\theta + Q$$

where Q is the sum of the energy taken by the flask, and that lost by radiation and conduction.

(2) Fill to the same depth with the liquid whose specific heat capacity c_2 is to be found.

Adjust V and I so as to obtain the same $\Delta\theta$ in the same t .

$$V_2 I_2 t = m_2 c_2 \Delta\theta + Q$$

Since Q is the same as in (1), it can be eliminated and c_2 evaluated.

23.5 (B) CONTINUOUS-FLOW METHOD

Procedure

(1) Pass the liquid through the calorimeter of fig. 23.5 until steady conditions hold.

Measure V_1 , I_1 , θ_1 , θ_2 , and the rate of liquid flow m_1 [kg] in time t [s].

Then

$$V_1 I_1 t = m_1 c (\theta_2 - \theta_1) + Q$$

where c [J kg⁻¹ K⁻¹] is the liquid specific heat capacity, and Q is the energy loss by radiation in time t under these conditions.

(2) Repeat, using adjusted values V_2 , I_2 and m_2 so that θ_1 and θ_2 are unaltered.

Then for the same t

$$V_2 I_2 t = m_2 c (\theta_2 - \theta_1) + Q$$

where Q , the energy loss by radiation, is unchanged under identical conditions for θ_1 and θ_2 .

(3) Subtracting

$$(V_1 I_1 - V_2 I_2) t = (m_1 - m_2) c (\theta_2 - \theta_1)$$

from which c can be calculated.

Notes

(a) The values of $(V_1 I_1 - V_2 I_2)$ and $(m_1 - m_2)$ should be as large as possible to reduce the percentage error, though V_2 , I_2 and m_2 should not be small.

(b) The method is convenient for finding the variation of the specific heat capacity of a liquid with temperature. (See also p. 190.)

Advantages

(a) Radiation loss is quantitatively controlled.

(b) Convection losses are eliminated.

(c) Conduction losses are negligible.

(d) Under steady temperature conditions no thermal energy is absorbed by the calorimeter.

(e) The temperatures to be recorded are static.

(f) $(\theta_2 - \theta_1)$ can be measured accurately with a differential resistance thermometer.

(g) V and I are accurate readings made by potentiometer methods.

Experimental detail

(a) A large quantity of liquid at constant temperature must be available.

(b) A constant-head apparatus is essential to ensure steady flow and steady temperature conditions.

23.6 (C) A COOLING METHOD

(Cooling methods are not in themselves important, but are mentioned here as an example of how to tackle problems on cooling.)

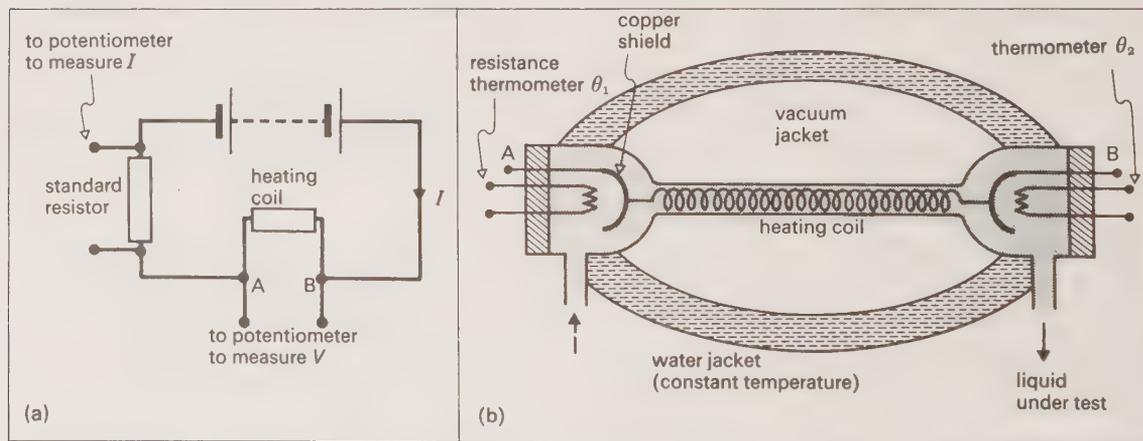


Fig. 23.5. Callendar and Barnes's continuous-flow method: (a) the circuit, and (b) details of the calorimeter.

Some cooling laws

The five-fourths power law

For cooling by natural convection (in a still atmosphere)

$$(\text{rate of heat loss}) \propto (\theta - \theta_s)^{5/4}$$

for a body at temperature θ in surroundings at temperature θ_s .

For small values of $(\theta - \theta_s)$, the five-fourths law approximates to a linear law, and for cooling corrections we usually assume the linear relationship (p. 191).

Newton's Law of cooling

For cooling under conditions of forced convection (i.e. in a steady draught)

$$(\text{rate of heat loss}) \propto (\theta - \theta_s)$$

for all methods by which the body loses heat.

This is an empirical law which holds for quite large temperature differences.

► The use of Newton's Law

Newton's Law states

$$\frac{dQ}{dt} \propto -(\theta - \theta_s)$$

where the minus sign indicates that the body is cooling ($d\theta/dt$ is negative).

$$\text{Since } \frac{dQ}{dt} = mc \frac{d\theta}{dt} = C \frac{d\theta}{dt}$$

we can write

$$C \frac{d\theta}{dt} \propto -(\theta - \theta_s)$$

$$\text{or } \frac{d\theta}{(\theta - \theta_s)} = -k dt$$

The value of k (which is usually taken to be constant) takes account of

- (a) the heat capacity C ,
- (b) the body's surface area,
- (c) the nature of its surface (e.g. its roughness), and
- (d) the shape of the surface (convex, flat or concave, etc.) and its orientation (vertical, horizontal, etc.).

Suppose a body cools from a temperature excess $(\theta - \theta_s)$ of $\Delta\theta_0$ to one of $\Delta\theta$ in a time interval Δt . Integrating we have

$$\ln \left(\frac{\Delta\theta}{\Delta\theta_0} \right) = -k \Delta t$$

$$\text{or } \Delta\theta = \Delta\theta_0 e^{-k\Delta t}$$

This result can be used for calculating changes of temperature under conditions to which *Newton's Law* applies.

Comparison of liquid specific heat capacities

Principle

If two bodies are to outward appearance identical (same size, shape and surface finish), then under the same cooling conditions (same $(\theta - \theta_s)$ and degree of convection) they will lose heat at the same rate.

This does *not* depend on any law controlling the rate of cooling.

Method

- (1) Take equal volumes of two liquids.
- (2) Place them successively in the same calorimeter.
- (3) Heat them to a conveniently high temperature.
- (4) Plot a $\theta-t$ curve for their cooling under identical conditions (fig. 23.6).

Because $Q = mc\theta$,

$$\frac{dQ}{dt} = mc \frac{d\theta}{dt} = C \frac{d\theta}{dt}$$

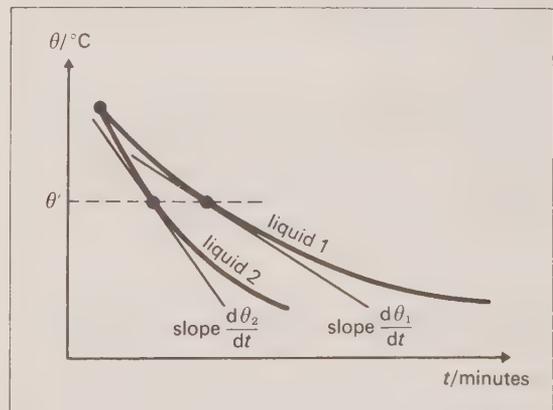


Fig. 23.6. Cooling curves being used to compare liquid specific heat capacities.

At the selected common temperature θ' shown on the graph, the rates of loss of heat of both liquids in their calorimeters are the same. (The nature of the liquids inside the calorimeter cannot affect dQ/dt , provided they are stirred.)

$$(C_c + m_1c_1)\frac{d\theta_1}{dt} = (C_c + m_2c_2)\frac{d\theta_2}{dt}$$

where C_c is the heat capacity of the calorimeter. If c_1 is known, c_2 can be found.

Note that the liquid of smaller c cools the more rapidly.

23.7 THE EFFECT OF PRESSURE ON SPECIFIC HEAT CAPACITIES

The first law of thermodynamics (p. 207) can be written

$$\left(\begin{array}{c} \text{heat given} \\ \text{to body} \end{array}\right) = \left(\begin{array}{c} \text{increase} \\ \text{of} \\ \text{internal} \\ \text{energy} \end{array}\right) + \left(\begin{array}{c} \text{work done} \\ \text{by body} \end{array}\right)$$

The change of internal energy controls the observed change of temperature $\Delta\theta$. When we add a given amount of heat ΔQ , the observed value of $\Delta\theta$ depends upon how much work the body does in pushing against the force exerted by its surroundings.

For convenience we distinguish these two principal specific heat capacities:

(a) The specific heat capacity at **constant volume** c_v is the value it takes if the volume change is zero (so that the work done by the body is zero).

(b) The specific heat capacity at **constant pressure** c_p is its value when the work done by the body (if it expands) is in opposition to a nearly constant force.

Solids and liquids have very large bulk moduli, which means that we cannot control their volumes. In practice, therefore, we only work with c_p . Comparison with theory, which gives c_v , requires the difference ($c_p - c_v$) to be taken into account.

The effect of the volume change for gases is discussed on p. 208.

23.8 MOLAR HEAT CAPACITY OF ELEMENTS

The **molar heat capacity** C_m of an element (sometimes called its atomic heat) is the heat capacity of one mole of its atoms. If we put $M_m =$ molar mass (not relative molecular mass), then we calculate C_m from

$$C_m = M_m c$$

$$C_m \left[\frac{\text{J}}{\text{mol K}} \right] = M_m \left[\frac{\text{kg}}{\text{mol}} \right] c \left[\frac{\text{J}}{\text{kg K}} \right]$$

Dulong and Petit (1819) found empirically that for many solid elements C_m is about $25 \text{ J mol}^{-1} \text{ K}^{-1}$ at ordinary temperatures.

Hard elements such as diamond, are important exceptions to this rule, having molar heat capacities that are lower at room temperature, but which approach $25 \text{ J mol}^{-1} \text{ K}^{-1}$ at higher temperatures (fig. 23.7).

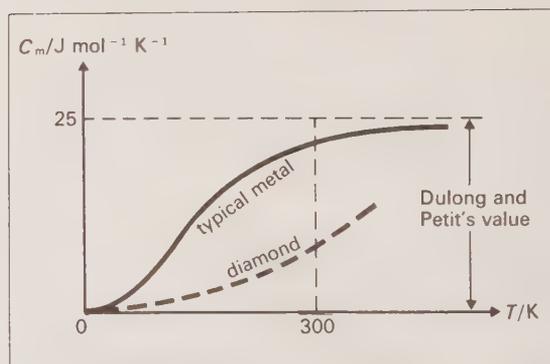


Fig. 23.7. Variation with temperature T of molar heat capacity C_m .

If we ignore the free electrons in a metal, the high-temperature values can be explained by classical kinetic theory (below). The shape of the curve at low temperatures needs quantum theory treatment.

► Kinetic theory prediction of molar heat capacity

Each ion in a metallic lattice has 6 degrees of freedom (p. 210), made up of two degrees for each of the possible vibrations along three mutually perpendicular directions.

The average energy for each ion is $6(\frac{1}{2}kT)$

according to the *principle of equipartition of energy* (p. 211).

The internal energy U for an amount of substance μ [mol] is given by

$$\begin{aligned} U &= \mu N_A 3kT \\ &= 3\mu RT \quad \text{since } R = N_A k \end{aligned}$$

At constant volume the molar heat capacity $C_{V,m}$ is given by

$$C_{V,m} = \frac{\Delta Q}{\mu \Delta T} = \frac{\Delta U}{\mu \Delta T} = 3R$$

Since $R = 8.31 \text{ J mol}^{-1} \text{ K}^{-1}$, we predict that the value of $C_{V,m}$ should be $24.9 \text{ J mol}^{-1} \text{ K}^{-1}$.

24 IDEAL GASES: KINETIC THEORY

24.1	INTRODUCTION	196
24.2	THE UNIVERSAL MOLAR GAS CONSTANT R	197
24.3	ASSUMPTIONS OF THE KINETIC THEORY	198
24.4	THE SPEEDS OF GAS MOLECULES	198
24.5	THE PRESSURE EXERTED BY AN IDEAL GAS	199
24.6	DEDUCTIONS FROM $p = \frac{1}{3} \rho \overline{c^2}$	199

▶ 24.7	MEAN FREE PATH λ	201
▶ 24.8	TRANSPORT PHENOMENA: MOLECULAR DIFFUSION	201
▶ 24.9	THERMAL CONDUCTION	202
▶ 24.10	VISCOSITY	203
▶ 24.11	CORRELATION OF TRANSPORT PHENOMENA	203

24.1 INTRODUCTION

Definitions

The **Avogadro constant** N_A is numerically equal to the number of atoms in exactly 12×10^{-3} kg of the nuclide carbon-12.

By experiment

$$N_A = 6.022\,045 \times 10^{23} \text{ mol}^{-1}$$

N_A has the dimensions of [amount of substance] $^{-1}$.

The **relative molecular mass** M_r (commonly miscalled molecular weight) is defined by

$$M_r = \frac{\text{mass of a molecule}}{\text{mass of } {}^{12}\text{C atom}} \times 12$$

Because M_r is a *dimensionless* quantity it has no unit.

The **mole** was defined on page 9. The *mass* of a mole is called the **molar mass** M_m , and is measured in kg mol^{-1} (see the following table).

Take great care not to confuse the *relative molecular mass* M_r and the *molar mass* M_m : they have different dimensions, and in the SI their numerical values differ by a factor of 10^{-3} .

Thus
$$M_m = \frac{M_r}{1000} \text{ kg mol}^{-1}$$

Examples

Substance	M_r	$M_m/\text{kg mol}^{-1}$
hydrogen	2	2×10^{-3}
helium	4	4×10^{-3}
carbon-12	12 (exactly)	12×10^{-3}
nitrogen	28	28×10^{-3}
oxygen	32	32×10^{-3}

Symbols

Note the following carefully. Complete familiarity with them now will prevent confusion later.

- (a) For a *sample of gas* in a particular enclosure
 N = total number of gas molecules
 n = number of molecules in unit volume (the **number density** of molecules), which has dimensions $[\text{L}^{-3}]$
 $(N_A = \text{the Avogadro constant})$
 $^*\mu$ = the amount of gas (which will be expressed in moles)
 M_g = the mass of the gas sample.
- (b) For a *particular molecular species*
 m = mass of a molecule
 M_r = relative molecular mass
 M_m = molar mass.

* The recommended symbol for an amount of substance is n ; our use of n for number density forces us to choose a different symbol.

These relationships follow:

$$(i) \mu [\text{mol}] = \frac{N[\text{number of molecules}]}{N_A[\text{number of molecules mol}^{-1}]}$$

$$\mu = \frac{N}{N_A}$$

$$(ii) \mu [\text{mol}] = \frac{M_g[\text{kg}]}{M_m[\text{kg mol}^{-1}]} = \frac{Nm}{M_m}$$

$$\mu = \frac{M_g}{M_m} = \frac{Nm}{M_m}$$

24.2 THE UNIVERSAL MOLAR GAS CONSTANT R

Macroscopic observations

Early experiments on a fixed amount or a fixed mass of gas showed that many common gases conformed fairly well to the behaviour described by the so-called *gas laws*.

(a) Boyle's Law

If T is constant, $pV = \text{constant}$.

(b) Charles's Law

If p is constant, $V/T = \text{constant}$.

(c) Pressure Law

If V is constant, $p/T = \text{constant}$

The temperature T was found by adding 273 K to the *Celsius* reading of a mercury thermometer. This empirical behaviour is summarized by the **gas equation**

$$\frac{p_1 V_1}{T_1} = \frac{p_2 V_2}{T_2}$$

A universal constant

Consider a particular gas, and take samples containing different amounts of gas μ [mol]. Suppose we vary the pressure p , the volume V and the thermodynamic temperature T (p. 181) of these samples. Plot the graph shown in fig. 24.1. We find

- (a) a series of curves, *all* of which cross the $pV/\mu T$ axis at the same point when extrapolated,
- (b) that other gases subjected to the same treatment give curves that cross the $pV/\mu T$ axis at the same point.

We deduce that the quantity $pV/\mu T$ is the same for all gases as the pressure tends to zero.

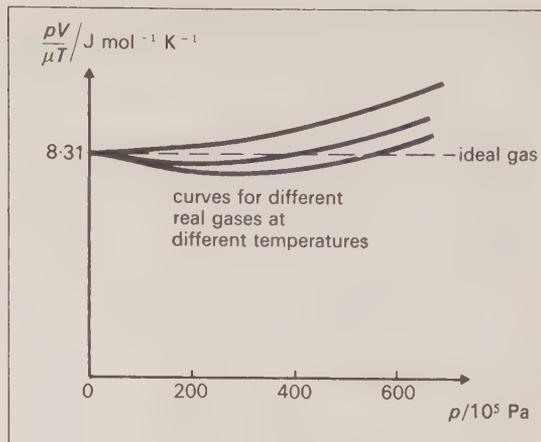


Fig. 24.1. Origin of the universal molar gas constant R .

The ideal gas

On a macroscopic scale we define an ideal gas to be one which obeys *Boyle's law* exactly, i.e.

$$pV = \text{constant}$$

for a fixed μ and T , and for all values of p .

We define *ideal-gas temperature* by

$$pV \propto T \quad (\text{p. 182})$$

By combining these two definitions we arrive at the **equation of state** for an ideal gas

$$pV = \mu RT$$

$$p \left[\frac{\text{N}}{\text{m}^2} \right] V [\text{m}^3] = \mu [\text{mol}] R \left[\frac{\text{J}}{\text{mol K}} \right] T [\text{K}]$$

The incorporation of μ into the equation has enabled us to write

$$pV/\mu T = R = 8.31 \text{ J mol}^{-1} \text{ K}^{-1}$$

where R is a *universal* constant, called the **universal molar ideal gas constant**.

The status of the gas laws

It is now seen that an ideal gas will automatically obey the gas laws, since for such a gas they are no more than definitions.

Real gases which do obey the laws are showing near-ideal behaviour.

Other forms of the equation of state

We can write the equation in the following forms:

(a) For 1 mole of gas,

$$pV = (1 \text{ mol})RT$$

or

$$pV_m = RT$$

where V_m is the **molar volume** and has the unit $\text{m}^3 \text{mol}^{-1}$.

(b) For a mass M_g of gas

$$pV = \left(\frac{M_g}{M_m} \right) RT = \left(\frac{Nm}{M_m} \right) RT$$

(c) For N molecules

$$pV = \frac{N}{N_A} RT$$

(A form using the *Boltzmann constant* k is given on p. 211.)

The molar volume V_m

Consider 1 mole of any ideal gas at s.t.p.

$$V = \frac{\mu RT}{p}$$

$$V_m = \frac{V}{\mu} = \frac{\left(8.31 \frac{\text{J}}{\text{mol K}} \right) \times (273 \text{ K})}{\left(1.01 \times 10^5 \frac{\text{N}}{\text{m}^2} \right)}$$

$$= 22.4 \times 10^{-2} \text{ m}^3 \text{ mol}^{-1}$$

One mole of any ideal gas at s.t.p. has a volume of $22.4 \times 10^{-3} \text{ m}^3$. This information is equivalent to being told that $R = 8.31 \text{ J mol}^{-1} \text{ K}^{-1}$.

(A volume of $1 \times 10^{-3} \text{ m}^3$ used to be called a litre.)

24.3 ASSUMPTIONS OF THE KINETIC THEORY

General aim

(a) To make some assumptions about the nature and properties of molecules, and

(b) To apply the laws of mechanics to gas molecules, in an attempt to relate their *microscopic* properties (speeds, masses and number per unit volume) to the *macroscopic* behaviour of an ideal gas.

Assumptions of the kinetic theory

(a) A gas consists of particles called *molecules*.

(b) The molecules are in constant *random motion*, but because we are dealing with perhaps 10^{23} molecules there are as many molecules travelling

in one direction as any other, and so the centre of mass of the gas is at rest.

(c) The *range of intermolecular forces* is small compared to the average separation of the gas molecules, so

(i) the intermolecular forces (both attractive and repulsive) are negligible except during a collision,

(ii) the volume of the gas molecules is negligible compared to the volume *occupied by the gas*,

(iii) the duration of a collision is negligible compared to the time spent in free motion.

(d) A molecule moves with uniform velocity between collisions—we ignore gravitational effects.

(e) On average we can consider the collisions of the molecules with one another and with the walls to be *perfectly elastic* (the effects of inelastic and superelastic collisions cancel out).

(f) We can apply *Newtonian mechanics* to molecular collisions.

These assumptions define, on a *microscopic* scale, what we mean by an **ideal gas**. (See also p. 197.)

24.4 THE SPEEDS OF GAS MOLECULES

The graph of fig. 24.2 is called the **Maxwellian distribution**. Its shape can be calculated using statistical mechanics.

Suppose a gas has N molecules. We define

(a) \hat{c} , the **most probable speed**, as that possessed by the greatest fraction of molecules.

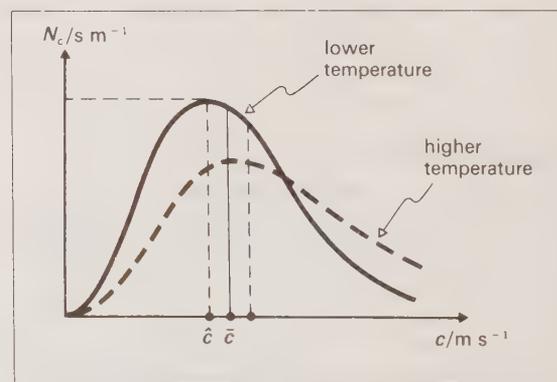


Fig. 24.2. The Maxwellian distribution. N is defined so that $N_c \delta c$ is the number of molecules with speeds between c and $(c + \delta c)$ at a particular temperature.

(b) \bar{c} , the average or mean speed, by

$$\bar{c} = \frac{1}{N}(c_1 + c_2 + \dots + c_N)$$

(c) $\overline{c^2}$, the mean square speed, by

$$\overline{c^2} = \frac{1}{N}(c_1^2 + c_2^2 + \dots + c_N^2)$$

It can be shown that, for a large value of N

(i) $\bar{c} \sim 0.8 \sqrt{\overline{c^2}}$, and

(ii) $\bar{c} \sim 0.9 \sqrt{\overline{c^2}}$.

24.5 THE PRESSURE EXERTED BY AN IDEAL GAS

(The bold A at the side of a statement indicates the use of an assumption.)

Suppose there are N molecules in a rectangular box of dimensions $a \times b \times l$ (fig. 24.3).

Suppose that the molecule shown has a velocity c of components c_x, c_y and c_z . When the molecule hits the shaded face, the change of momentum is

$$-mc_x - (mc_x) = -2mc_x$$

to the right. We assume an elastic collision. **A**

The impulse imparted to the shaded face is

$$+2mc_x \quad (\text{to the right})$$

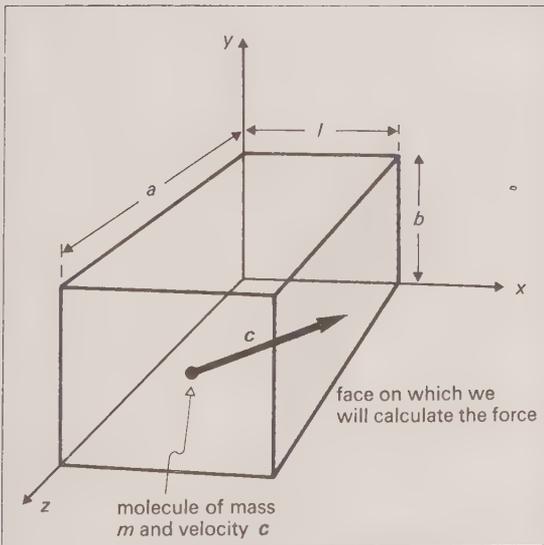


Fig. 24.3. Calculation of ideal gas pressure.

The time interval before the same molecule makes a second collision at the face is

$$\frac{2l}{c_x} \quad (\text{if molecular diameter} \ll l) \quad \mathbf{A}$$

Hence the number of collisions on the shaded face of this molecule in unit time is $c_x/2l$.

Force exerted by this molecule on the face is

$$\begin{aligned} & \left(\begin{array}{l} \text{rate at which molecule transfers} \\ \text{momentum to the face—Newton II} \end{array} \right) \\ &= 2mc_x \frac{c_x}{2l} = \frac{mc_x^2}{l} \end{aligned}$$

Pressure at shaded face is

$$p = \frac{\text{total force exerted by } N \text{ molecules}}{\text{area of shaded face}} = \frac{(m/l)[(c_x^2)_1 + (c_x^2)_2 + \dots + (c_x^2)_N]}{ab}$$

Substitute $n = N/abl$, and gas density $\rho = nm$, so

$$p = \rho \overline{c_x^2}$$

For any molecule $c^2 = c_x^2 + c_y^2 + c_z^2$.

Because (i) N is large, and (ii) the molecules have random motion, **AA**

$$\overline{c_x^2} = \overline{c_y^2} = \overline{c_z^2} = \frac{1}{3} \overline{c^2}$$

This implies that they do not move parallel to any axis preferentially, or indeed parallel to any of these three axes.

We conclude

$$p = \frac{1}{3} \rho \overline{c^2}$$

Notes

(a) The result is valid for all shapes of container.

(b) According to *Pascal's principle*, the pressure on the other faces will be the same if we neglect the weight of the gas—assumption (d). **A**

(c) Assumption (c) implies that we can neglect intermolecular collisions. Any such collisions cannot alter the net rate of transfer of momentum in any direction, and so the result is valid. **A**

24.6 DEDUCTIONS FROM $p = \frac{1}{3} \rho \overline{c^2}$

(A) Kinetic interpretation of temperature

From $p = \frac{1}{3} \rho \overline{c^2}$, it follows that

$$\begin{aligned} pV &= \frac{1}{3} Nmc\overline{c^2} \\ &= \frac{2}{3} N \left(\frac{1}{2} mc\overline{c^2} \right) \end{aligned}$$

Compare

(a) $pV \propto \frac{1}{2}mc^2$, the consequence of our kinetic theory assumptions, with

(b) $pV \propto T$, our definition of temperature on the ideal-gas scale (p. 181), which corresponds to the thermodynamic scale.

We conclude that

The thermodynamic temperature of an ideal gas is proportional to the mean translational k.e. of its molecules (and to the total translational kinetic energy).

This interpretation is consistent with the ideal-gas equation, and so (p. 197) with the gas laws.

The idea is developed further on page 211.

Because $\frac{1}{2}mc^2 \propto T$

(1) for a particular gas (m constant),

$$\overline{c^2} \propto T, \text{ and}$$

(2) for different gases at the same T

$$\overline{c^2} \propto \frac{1}{m}$$

(B) Avogadro's law

Consider two ideal gases

$$p_1V_1 = \frac{1}{3}N_1m_1\overline{c_1^2}$$

$$p_2V_2 = \frac{1}{3}N_2m_2\overline{c_2^2}$$

If they have the same pressure, volume and temperature, then

$$p_1 = p_2$$

$$V_1 = V_2$$

$$\frac{1}{2}m_1\overline{c_1^2} = \frac{1}{2}m_2\overline{c_2^2}$$

which shows

$$N_1 = N_2$$

Equal volumes of all ideal gases under the same conditions of temperature and pressure contain the same number of molecules.

(C) Gaseous diffusion and Graham's law

Since

$$pV = \frac{1}{3}M_g\overline{c^2}$$

and

$$pV = (M_g/M_m)RT$$

it follows that

$$\overline{c^2} = 3RT/M_m$$

or

$$c_{r.m.s.} = \sqrt{\frac{3RT}{M_m}} = \sqrt{\frac{3p}{\rho}}$$

Gaseous diffusion is treated as a transport phenomenon in 24.8, and as a random process in 26.5. For a simple treatment we will assume that the number of particles passing through a given small hole in unit time $\propto \bar{c} \propto c_{r.m.s.}$. Then examination of the boxed line above shows

The number of molecules escaping in unit time

(a) *at a fixed temperature is inversely proportional to the square root of the relative molecular mass, and*

(b) *at a fixed pressure is inversely proportional to the square root of the density.*

To calculate the rate at which the mass escapes we would need to take into account the mass of the molecules.

(D) Dalton's law of partial pressures

Consider a mixture of ideal gases in volume V .

Then

$$p_1V = \frac{1}{3}N_1m_1\overline{c_1^2}$$

and

$$p_2V = \frac{1}{3}N_2m_2\overline{c_2^2}$$

When the gases have acquired the same temperature

$$m_1\overline{c_1^2} = m_2\overline{c_2^2} = m\overline{c^2} \text{ (say)}$$

Adding

$$(p_1 + p_2 + \dots)V = \frac{1}{3}m\overline{c^2}(N_1 + N_2 + \dots)$$

Hence total pressure $p = \Sigma p_i$.

When two or more gases which do not react chemically are present in the same container, the total pressure is the sum of the partial pressures (the pressure which each gas would exert if isolated in the container).

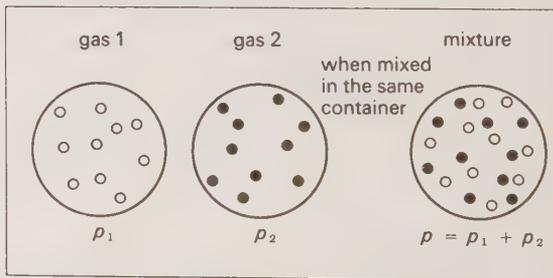


Fig. 24.4. Symbolic representation of Dalton's Law.

► 24.7 MEAN FREE PATH λ

The mean free path λ of a gas molecule is the average distance it travels between collisions. An approximate value for λ can be found from the following simplified argument.

Let r_0 (fig. 24.5) be the effective molecular diameter.

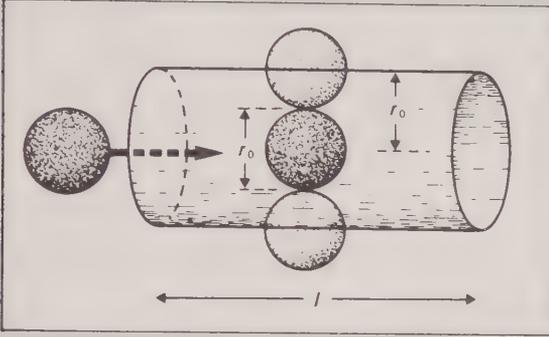


Fig. 24.5. Calculation of mean free path λ .

Then if the molecule travels a distance l in time Δt the volume swept out will be $\pi r_0^2 l$. We assume that the molecule will collide with any other molecule within this volume. The volume contains $\pi r_0^2 l n$ molecules (where $n[\text{m}^{-3}]$ is their number density), and this is the number of collisions during the interval Δt .

$$\text{Mean free path } \lambda = \frac{\text{distance travelled}}{\text{number of collisions}}$$

$$= \frac{l}{\pi r_0^2 l n} = \frac{1}{\pi r_0^2 n}$$

$$\lambda[\text{m}] = \frac{1}{\pi r_0^2 [\text{m}^2] n [\text{m}^{-3}]}$$

Orders of magnitude

Gas	n/m^{-3}	λ/m	Pressure p/Pa
electron gas in copper	10^{29}	10^{-8}	—
atmosphere			
— sea level	10^{25}	10^{-7}	10^5
— at 100 km	10^{17}	1	10^{-1}
— at 300 km	10^{14}	10^4	10^{-4}

► 24.8 TRANSPORT PHENOMENA: MOLECULAR DIFFUSION

Introduction

The continuous random motion of gas particles gives rise, by the statistical processes described in chapter 26, to the net macroscopic transport of

- (a) matter—molecular diffusion
- (b) energy—thermal conduction
- (c) momentum—viscosity.

These transfers, together with the net movement of electric charge (*electrical conduction*) constitute a group of processes with common features called the **transport phenomena**.

In this chapter only gaseous diffusion is analysed in detail, but mention is made of gaseous conduction and viscosity, since their study involves similar techniques.

Molecular diffusion

When the number density n of molecules varies from place to place, diffusion (the net transfer of molecules) occurs in the direction in which n decreases, i.e. from a region of high to one of low concentration, thereby equalizing the distribution of diffusing molecules. Diffusion is a statistical process caused by random motion (see 26.5). We will consider molecules of one type only (**self-diffusion**).

Fick's law

Suppose that gas molecules are diffusing in the x -direction. Let N molecules cross an area A placed normal to that direction in time Δt . Then the **particle current density** j_N is given by

$$j_N = \frac{N}{A \Delta t}$$

$$j_N \left[\frac{1}{\text{m}^2 \text{ s}} \right] = \frac{N}{A [\text{m}^2] \Delta t [\text{s}]}$$

Experiment shows that, so long as statistical results are applicable, the particle current density is proportional to the rate at which the concentration changes in the x -direction.

$$j_N \propto \left(\frac{dn}{dx} \right)$$

$$j_N = -D \left(\frac{dn}{dx} \right)$$

$$j_N \left[\frac{1}{\text{m}^2 \text{ s}} \right] = -D \left[\frac{\text{m}^2}{\text{s}} \right] \frac{dn}{dx} \left[\frac{1}{\text{m}^4} \right]$$

Notes

- (a) The boxed equation is called **Fick's law**.
 (b) The constant of proportionality D is called the **diffusion coefficient**. It has the unit $\text{m}^2 \text{s}^{-1}$.
 (c) The negative sign indicates that the molecules move in the direction in which n decreases.

Calculation of the diffusion coefficient

In fig. 24.6 molecules are diffusing in the positive x -direction. This means that dn/dx is negative, but we do not anticipate this by writing $-(dn/dx)$.

(a) Since a molecule travels an average distance λ between collisions, the concentrations at planes R and P are of the form

$$n + \left(\frac{dn}{dx} \right) \Delta x = n \pm \left(\frac{dn}{dx} \right) \lambda,$$

as shown in the diagram. (For plane R $\Delta x = +\lambda$, and for plane P $\Delta x = -\lambda$. Both are measured relative to Q.)

(b) If the molecular motion is *random*, at any instant we may visualize one sixth of the molecules to be travelling in the positive x -direction, and one sixth in the negative. Thus in time Δt the number of molecules that cross an area A in the central plane Q is

$$(i) \frac{1}{6} \left(n - \lambda \frac{dn}{dx} \right) A \bar{c} \Delta t, \text{ travelling upwards, and}$$

$$(ii) \frac{1}{6} \left(n + \lambda \frac{dn}{dx} \right) A \bar{c} \Delta t, \text{ travelling downwards.}$$

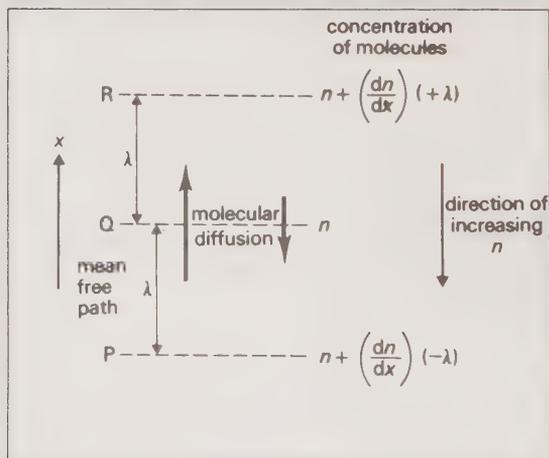


Fig. 24.6. Calculation of the diffusion coefficient. P, Q and R are planes drawn normal to the x -direction.

(c) The *net* upward movement

$$N = -\frac{1}{3} \lambda \left(\frac{dn}{dx} \right) A \bar{c} \Delta t$$

and so
$$j_N = \frac{N}{A \Delta t} = -\frac{1}{3} \lambda \bar{c} \left(\frac{dn}{dx} \right)$$

(d) Comparison with the definition of D yields

$$D = \frac{1}{3} \lambda \bar{c}$$

Correlation with experiment

At fixed pressure

$$\lambda \propto 1/\rho \propto T$$

and

$$\bar{c} \propto \sqrt{T}$$

since

$$\frac{1}{2} m \bar{c}^2 = \frac{2}{3} kT$$

We predict that, at fixed pressure,

$$D \propto T^{3/2}$$

Experiment yields a slightly higher power.

► 24.9 THERMAL CONDUCTION

Thermal conduction occurs when the temperature of a substance varies from place to place, the transport of energy being in that direction in which the temperature decreases. Following our treatment of diffusion we define an **energy current density** j_Q by

$$j_Q = \frac{Q}{A \Delta t}$$

leading to the experimental result

$$j_Q = -k \frac{dT}{dx}$$

This equation states **Fourier's law**, and defines for us the **thermal conductivity** k . The equation is more familiar in the form

$$\frac{dQ}{dt} = -\lambda A \frac{d\theta}{dx} \quad (\text{p. 237})$$

in which λ is used instead of k . (In this chapter λ represents the molecular mean free path.)

The kinetic theory calculation of k for an ideal gas is similar to that for D , with the temperature T (and by implication molecular energy) taking the rôle of the concentration n . The energy transfer by molecules migrating at equal rates in opposite directions is related to the mass transfer (and hence the density ρ), and to the specific heat capacity at constant volume c_V . The result is

$$k = \frac{1}{3} \bar{c} \rho c_V \lambda$$

Now the product $\lambda \rho$ is independent of pressure, as are \bar{c} and c_V . We predict k to be independent of pressure, and experiment agrees over a wide range of pressures. This is excellent support for the kinetic theory.

► 24.10 VISCOSITY

Viscosity occurs when the drift or convection velocity v within a fluid varies from place to place, the transport of momentum p being in that direction in which the velocity decreases (fig. 19.2). Following our earlier treatments, we define a **momentum current density** j_p by

$$j_p = \frac{p}{A \Delta t}$$

leading to the result

$$j_p = -\eta \frac{dv}{dx}$$

This equation states the **law of viscous flow**, and defines for us the **coefficient of viscosity** η . The equation is more familiar in the form

$$F = -\eta A \frac{dv}{dx} \quad (\text{p. 163})$$

with the negative sign being omitted in a non-vector approach. For *Newtonian* fluids (which include gases) η is found by experiment to be a constant at a fixed temperature.

The kinetic theory calculation of η for an ideal gas is similar to that for D , with the systematic drift velocity v (and by implication the drift momentum p) taking the rôle of the concentration n . The result is

$$\eta = \frac{1}{3} \rho \bar{c} \lambda$$

Notes

(a) Since the product $\lambda \rho$ and \bar{c} are independent of pressure, we predict that η will be too. *Maxwell* confirmed experimentally that this is true over a wide range of pressures. This is further excellent support for the kinetic theory.

(b) The equation suggests

$$\eta \propto \bar{c} \propto \sqrt{T},$$

since $\frac{1}{2} m \bar{c}^2 = \frac{3}{2} kT$, and $\lambda \rho$ is independent of temperature. Experiment indicates that η increases with temperature, but that the power exceeds $\frac{1}{2}$.

► 24.11 CORRELATION OF TRANSPORT PHENOMENA

The table summarizes the discussion so far.

Name of process	Entity or property being transported	Definition of coefficient	Kinetic theory prediction of coefficient
diffusion	molecules N	$j_N = -D \frac{dn}{dx}$	$D = \frac{1}{3} \lambda \bar{c}$
thermal conduction	energy Q	$j_Q = -k \frac{dT}{dx}$	$k = \frac{1}{3} c_V \rho \lambda$
viscosity	momentum p	$j_p = -\eta \frac{dv}{dx}$	$\eta = \frac{1}{3} \rho \bar{c} \lambda$

(a) If we combine the results for k and η , we predict

$$k/\eta c_V = 1$$

Experiments using a wide range of gases give values between 1.4 and 2.5, an order-of-magnitude agreement which supports the validity of our elementary theory.

(b) The simple argument of 24.7 showed

$$\lambda = 1/\pi r_0^2 n$$

This means that a measurement of D , k and/or η for a gas enables independent estimates to be made of the molecular diameter r_0 . There is good correlation between the results of the three different methods.

25 IDEAL GASES: THERMAL BEHAVIOUR

25.1	IMPORTANT TERMS	204
25.2	EXTERNAL WORK AND THE INDICATOR DIAGRAM	205
25.3	METHODS OF CHANGING THE STATE OF A GAS	205
25.4	THE FIRST LAW OF THERMODYNAMICS	205
25.5	THE MOLAR HEAT CAPACITIES OF GASES	207
25.6	MAYER'S EQUATION (1842) $C_{p,m} - C_{v,m} = R$	208
25.7	INTRODUCTION TO ISOTHERMAL AND ADIABATIC PROCESSES	208

25.8	ISOTHERMAL PROCESSES	208
▶ 25.9	ADIABATIC PROCESSES	209
▶ 25.10	THE SECOND LAW OF THERMODYNAMICS	210
▶ 25.11	DEGREES OF FREEDOM	210
▶ 25.12	EQUIPARTITION OF ENERGY	211
▶ 25.13	THE KINETIC THEORY PREDICTION OF GAS MOLAR HEAT CAPACITIES	211
▶ 25.14	THE MEASUREMENT OF γ	212
▶ 25.15	VARIATION OF $C_{v,m}$ WITH TEMPERATURE	213

25.1 IMPORTANT TERMS

A **dissipative process** is one in which an energy conversion (e.g. by the performance of work) usually shows itself as a temperature increase.

Examples

- (a) Work done against friction.
- (b) The action of an electrical resistor.

A **non-dissipative process** is one in which the energy converted can be recovered by reversing the direction of the displacement that caused the energy conversion.

Examples

- (a) The very slow compression of a gas.
- (b) The charging of an electrical capacitor, if we neglect *Joule* heating.

Thermodynamic equilibrium

A system is in thermodynamic equilibrium if

- (a) it is in chemical equilibrium, and
- (b) there are no temperature and pressure gradients (which might cause its *state* (q.v.) to change as time passes).

The **state of a substance** is specified by quoting its pressure p and temperature T . For p and T to

have unique meaning, the substance must be in *thermodynamic equilibrium* (q.v.). The **equation of state** is the mathematical relationship between p , V and T , which are the *macroscopic* properties of the system. (See p. 197.) (Be careful to distinguish between this use of *state*, and the word *phase*.)

Reversibility

A process is reversible if the state of a system is changed in such a way that the system is at all times effectively in thermodynamic equilibrium.

The implications are

- (a) that the process would take an infinite time (be **quasi-static**),
- (b) that at every stage it would be possible to make the direction of the process reverse by an infinitesimal opposite change in the external conditions,
- (c) that the departure from equilibrium is negligible.

(A reversible change does *not* mean that a *cycle* is necessarily considered—we do not imply that the state of the substance will be restored to its original value.)

25.2 EXTERNAL WORK AND THE INDICATOR DIAGRAM

External work

To calculate the external work done by any substance which expands *reversibly*:

work done by substance = (force) \times (displacement)

$$\begin{aligned}\delta W &= (pA) \delta x \\ &= p \delta V\end{aligned}$$

For a reversible expansion which changes the volume from V_1 to V_2

$$W = \int_{V_1}^{V_2} p \, dV$$

$$\delta W[\text{J}] = p \left[\frac{\text{N}}{\text{m}^2} \right] \delta V[\text{m}^3]$$

Notes

(a) For a *non-reversible* expansion the value of p is not properly defined.

(b) The work done by the substance is positive when its pressure–force and the displacement δx are in the same direction.

(c) To evaluate the integral we must be able to write p in terms of V , as we can when

- (i) p is constant, so that $W = p(V_2 - V_1)$,
- (ii) the change is isothermal (p. 208),
- (iii) the change is adiabatic (p. 209).

The indicator diagram

A graph of p against V is called an **indicator diagram**. The work done during a reversible change depends on the shape of the p – V curve (fig. 25.2). It does *not* have a unique value for a given V_1 and V_2 .

A **cycle** is any process during which the state of a substance (particularly a gas) undergoes changes which ultimately restore its original value. It is represented by a *closed loop* on the indicator diagram.

Example of a cycle in which a gas does work

In the cycle of fig. 25.3 (overleaf) the state of the gas has returned to its original value (p_1, V_1, T_1) and yet the gas has done work. We explain this by saying that energy has been transferred not just by doing work, but also by *thermal energy transfer* (heat).

During the cycle there must have been given to the gas an amount of heat equal to the work done by the gas.

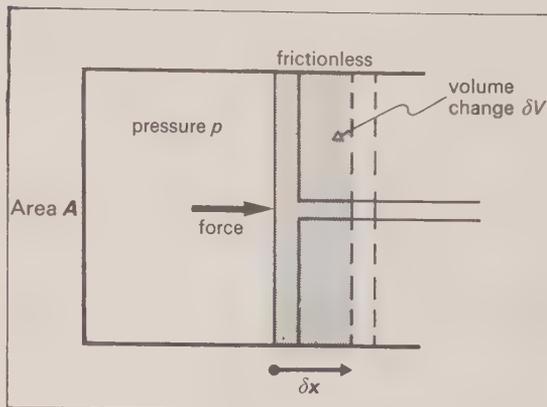


Fig. 25.1. Calculation of external work.

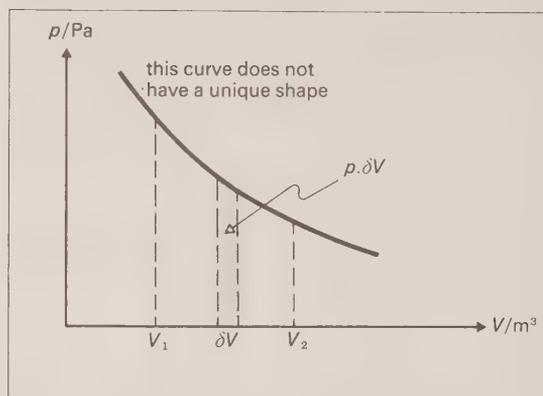


Fig. 25.2. To show on an indicator diagram the work done during a reversible expansion.

25.3 METHODS OF CHANGING THE STATE OF A GAS

The detailed shape of the p – V curve depends on the way in which the state of the gas is changed. Study fig. 25.4 carefully, which shows four *special cases*. Note the behaviour of T during each change.

The *isothermal* and *adiabatic* processes are discussed on p. 208. Fig. 25.4(c) shows a pair of rectangular hyperbolas (*Boyle's Law* curves).

25.4 THE FIRST LAW OF THERMODYNAMICS

Suppose we supply heat ΔQ to a gas, and that this produces two effects:

- (1) it raises the *internal energy* of the gas (below) from U to $(U + \Delta U)$, and
- (2) it enables the gas to do *external work* ΔW .

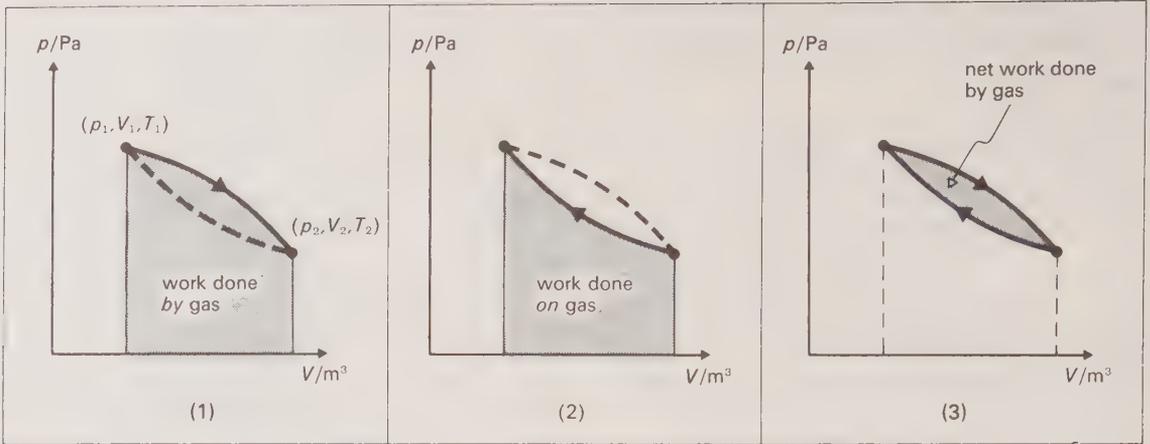


Fig. 25.3. A cycle represented on an indicator diagram.

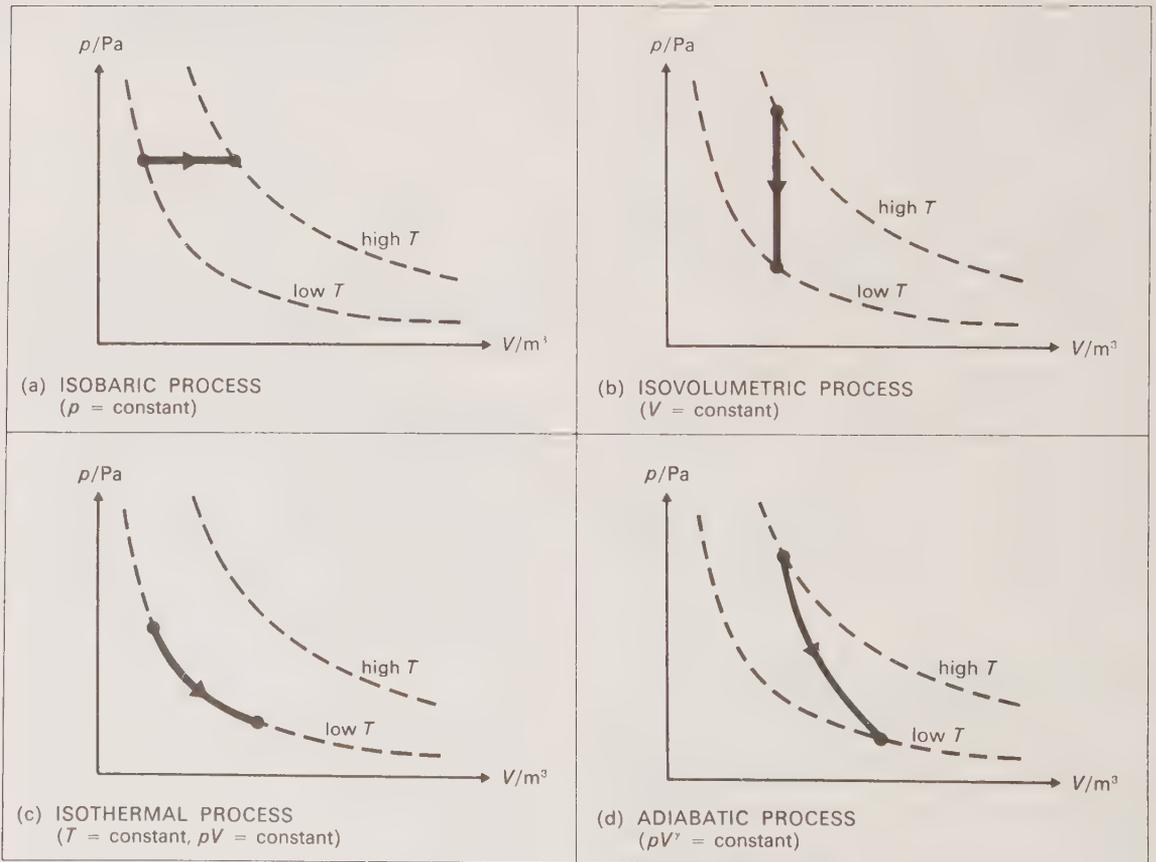


Fig. 25.4. Some ways of changing the state of a gas. Each process is conducted reversibly.

The first law of thermodynamics expresses the principle of conservation of energy for this situation by

$$\Delta Q = \Delta U + \Delta W$$

The law recognizes two different processes by which we can transfer energy

- (i) by doing work on a macroscopic scale—ordered energy, or
- (ii) by doing work on a microscopic scale—heat, i.e. the transfer of disordered energy.

What happens when heat is given to a gas

Suppose we give heat to a gas in a situation where a volume change is allowed to occur. The energy may be used

- (a) to do **external work** in pushing against the forces exerted by the environment ($\int p dV$ if the change is reversible), and/or
- (b) to increase the **internal energy** U of the system.

For example it may increase

- (i) the translational k.e. of the gas molecules,
- (ii) the rotational energy of polyatomic molecules (p. 210),
- (iii) the vibrational energy of polyatomic molecules (p. 211),
- (iv) the intermolecular potential energy if work has to be done to increase the molecular separation. (This would be zero for an ideal gas.)

These five terms could be reclassified as

- (1) those that depend *only on temperature*—the translational, rotational and vibrational energies, and
- (2) those that depend *on volume*—the external work and intermolecular p.e.

25.5 THE MOLAR HEAT CAPACITIES OF GASES

Suppose we add heat ΔQ to an amount of gas μ [mol] under specified conditions (see box), and that this causes a temperature increase ΔT .

Then we define the **principal molar heat capacities** $C_{V,m}$ and $C_{p,m}$ by the equations

$$C_{V,m} = \frac{\Delta Q}{\mu \Delta T} \quad C_{p,m} = \frac{\Delta Q}{\mu \Delta T}$$

constant volume constant pressure

$$C_m \left[\frac{\text{J}}{\text{mol K}} \right] = \frac{\Delta Q [\text{J}]}{\mu [\text{mol}] \Delta T [\text{K}]}$$

A gas has an infinite number of molar heat capacities because it can be made to do a variable amount of external work. The value of C_m may vary with temperature (p. 213).

The *internal energy* of an ideal gas depends only on temperature. This means that for any process which changes T to $(T + \Delta T)$, the internal energy will always change from U to $(U + \Delta U)$, and that the *same* ΔU will always produce the same ΔT .

For the process a–b of fig. 25.5

$$\Delta Q = \Delta U + \Delta W$$

becomes

$$\mu C_{V,m} \Delta T = \Delta U + 0$$

since $\Delta V = 0$ and no external work is done.

The relation

$$\Delta U = \mu C_{V,m} \Delta T$$

holds for all methods of changing U for an ideal gas, even if the change is not isovolumetric.

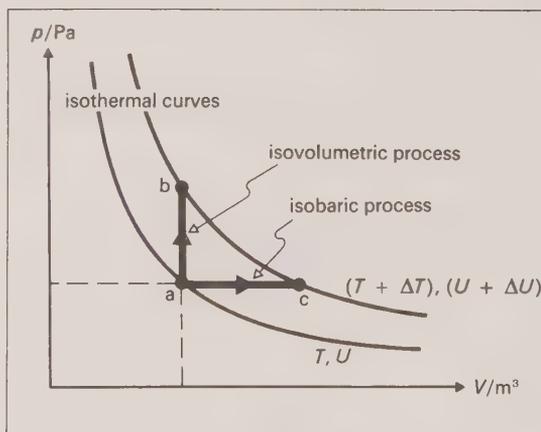


Fig. 25.5. An ideal gas experiencing the same change of internal energy ΔU while undergoing isobaric (a–c) and isovolumetric (a–b) processes.

Measurement of gas molar heat capacities

(a) $C_{V,m}$ can be measured by *Joly's* differential steam calorimeter.

(b) $C_{p,m}$ can be measured by

(i) the method of mixtures (*Regnault*), or

(ii) the continuous-flow calorimeter (*Scheel and Heuse*).

25.6 MAYER'S EQUATION (1842)

$$C_{p,m} - C_{V,m} = R$$

Refer to fig. 25.5. Suppose we take an amount μ [mol] of an ideal gas through the isobaric process a-c reversibly.

Then in $\Delta Q = \Delta U + \Delta W$, we have

(a) $\Delta Q = \mu C_{p,m} \Delta T$ (by definition of $C_{p,m}$)

(b) $\Delta U = \mu C_{V,m} \Delta T$ (previous section), and

(c) $\Delta W = p \Delta V$ (because the pressure is constant and the change reversible).

For an ideal gas

$$pV = \mu RT$$

and if only V and T are varied (as they are here),

$$p \Delta V = \mu R \Delta T$$

Then

$$\Delta Q = \Delta U + \Delta W$$

becomes $\mu C_{p,m} \Delta T = \mu C_{V,m} \Delta T + \mu R \Delta T$

or

$$C_{p,m} = C_{V,m} + R$$

$$C_{p,m} - C_{V,m} = R$$

Mayer's equation

$$pV = \text{constant}$$

This is true for *any ideal gas* (one for which there is no intermolecular p.e.).

25.7 INTRODUCTION TO ISOTHERMAL AND ADIABATIC PROCESSES

It is convenient to relate the first law of thermodynamics

$$\Delta Q = \Delta U + \Delta W$$

to the four methods of changing the state of a gas shown in fig. 25.4.

(a) In the **isobaric** reversible process

$$\Delta W = p \Delta V$$

(b) In the **isovolumetric** process

$$\Delta V = 0 \quad \Delta W = 0$$

and this led to $\Delta U = \mu C_{V,m} \Delta T$

(c) In an **isothermal** process

$$\Delta T = 0 \quad \Delta U = 0$$

and this leads to $\Delta Q = \Delta W$. There is no change of internal energy for an ideal gas.

(d) We define an **adiabatic** process to be one for which $\Delta Q = 0$, so

$$0 = \Delta U + \Delta W \quad \therefore -\Delta U = \Delta W$$

If a gas is made to do work, there is a decrease of internal energy which becomes apparent as a temperature decrease.

Examples of processes which are approximately adiabatic are those

(i) in insulated containers,

(ii) which take place very rapidly,

(iii) experienced by a very large mass of a substance.

25.8 ISOTHERMAL PROCESSES**(A) Equation**

$$pV = \mu RT$$

is the equation of state for an ideal gas, and

$$\Delta T = 0$$

describes an isothermal process.

So

is the equation which relates p and V during such a process.

(B) Work done during a reversible process

The process of fig. 25.1 must be very slow (quasi-static) to ensure isothermal reversible conditions.

If this is so, the work done by the gas

$$\begin{aligned} W &= \int_{V_1}^{V_2} p \, dV \\ &= \int_{V_1}^{V_2} \frac{\mu RT}{V} \, dV \\ &= \mu RT \ln \left(\frac{V_2}{V_1} \right) \end{aligned}$$

$$W[J] = \mu[\text{mol}]R \left[\frac{J}{\text{mol K}} \right] T[\text{K}] \ln \left(\frac{V_2}{V_1} \right)$$

The (natural) logarithm has no unit.

► 25.9 ADIABATIC PROCESSES

These processes can be

(a) *non-reversible*, as in the *Joule* and the *Joule–Kelvin* experiments (p. 233), or

(b) *approximately reversible*, when there is only a small departure from the equilibrium situation. For example

(i) the *Clément and Désormes* method for measuring γ (p. 212), and

(ii) the alternating pressure changes associated with the passage of a sound wave (p. 328).

(A) Equation

Consider a *reversible* adiabatic change experienced by an amount μ [mol] of an *ideal* gas.

In $\Delta Q = \Delta U + \Delta W$, we have

- (a) $\Delta Q = 0$ (adiabatic change)
 (b) $\Delta U = \mu C_{V,m} \Delta T$ (section 25.5) since the gas is ideal, and
 (c) $\Delta W = \int p \, dV$ since the change is reversible.

For an infinitesimal process

$$-\Delta U = \Delta W$$

becomes $-\mu C_{V,m} \delta T = p \delta V$ (1)

Since the gas is ideal

$$\begin{aligned} pV &= \mu RT \\ p \delta V + V \delta p &= \mu R \delta T \\ \delta T &= \frac{1}{\mu R} (p \delta V + V \delta p) \end{aligned} \quad (2)$$

Eliminating δT from (1) and (2)

$$\begin{aligned} -C_{V,m}(p \delta V + V \delta p) &= R p \delta V \\ -C_{V,m} V \delta p &= (R + C_{V,m}) p \delta V \end{aligned}$$

Using $C_{p,m} - C_{V,m} = R$ we have

$$-V \delta p = \left(\frac{C_{p,m}}{C_{V,m}} \right) p \delta V$$

and putting $\frac{C_{p,m}}{C_{V,m}} = \gamma$

$$\begin{aligned} -\frac{\delta p}{p} &= \gamma \frac{\delta V}{V} \\ \therefore -\int \frac{dp}{p} &= \gamma \int \frac{dV}{V} \end{aligned}$$

If we can assume $C_{p,m}$ and $C_{V,m}$ (and hence γ) do not vary with T for this process,

then $-\ln p = \gamma \ln V + \text{constant}$

Hence

$$pV^\gamma = \text{constant}$$

and this equation relates p and V for a *reversible adiabatic* process for an *ideal* gas for which γ remains constant.

(B) Change of temperature

Fig. 25.6 shows a reversible adiabatic expansion during which a fall of temperature occurs ($T_1 > T_2$).

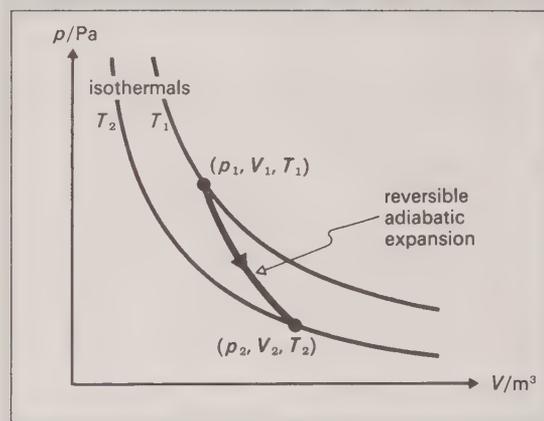


Fig. 25.6. To calculate the temperature change accompanying a reversible adiabatic expansion.

From $pV = \text{constant}$ and $pV^\gamma = \text{constant}$ it follows that $\gamma = \frac{\text{slope of adiabatic}}{\text{slope of isothermal}}$

at a particular point.

For an ideal gas

$$p_1 V_1 = \mu R T_1 \quad \text{and} \quad p_2 V_2 = \mu R T_2$$

$$\text{so} \quad \frac{T_1}{T_2} = \frac{p_1 V_1}{p_2 V_2} \quad (1)$$

For a reversible adiabatic change

$$p_1 V_1^\gamma = p_2 V_2^\gamma$$

$$\text{so} \quad \left(\frac{p_1}{p_2} \right) = \left(\frac{V_2}{V_1} \right)^\gamma \quad (2)$$

$$\text{and} \quad \left(\frac{V_1}{V_2} \right) = \left(\frac{p_1}{p_2} \right)^{-1/\gamma} \quad (3)$$

From (1) and (2)

$$\frac{T_1}{T_2} = \left(\frac{V_2}{V_1}\right)^{\gamma-1} \quad \text{or} \quad \boxed{TV^{\gamma-1} = \text{constant}}$$

From (1) and (3)

$$\frac{T_1}{T_2} = \left(\frac{p_1}{p_2}\right)^{(\gamma-1)/\gamma} \quad \boxed{T^\gamma p^{1-\gamma} = \text{constant}}$$

subject to the same conditions as section 25.9(A).

(C) Work done during a reversible process

Suppose the process is described validly by

$$pV^\gamma = \text{constant}, k$$

Then the work done

$$\begin{aligned} W &= \int_{V_1}^{V_2} p \, dV \\ &= k \int_{V_1}^{V_2} \frac{dV}{V^\gamma} \\ &= k \left[\frac{V^{1-\gamma}}{1-\gamma} \right]_{V_1}^{V_2} \\ &= \frac{k}{1-\gamma} \cdot (V_2^{1-\gamma} - V_1^{1-\gamma}) \\ &= \frac{1}{\gamma-1} \cdot (p_1 V_1 - p_2 V_2) \end{aligned}$$

since $p_1 V_1^\gamma = p_2 V_2^\gamma = k$.

For an expansion

$$T_1 > T_2 \quad \text{and} \quad p_1 V_1 > p_2 V_2$$

and since $\gamma > 1$ always, $W > 0$: the gas always does work.

Notes

(a) A gas does more work (for a given volume increase) in a *reversible* expansion than in a *non-reversible* expansion. For example, when an ideal gas expands irreversibly into a vacuum it does no work.

(b) A gas does more work in a reversible *isothermal* expansion than in a reversible *adiabatic* expansion.

► 25.10 THE SECOND LAW OF THERMODYNAMICS

The *first* law of thermodynamics (p. 207) would not be violated if either

(i) a heat engine were to convert internal energy entirely into mechanical energy, or

(ii) heat were to be transferred from a cold body to a hot one, without any loss of energy.

Nevertheless practical experience tells us that neither of the above processes is ever achieved, and we express this observation formally in the **second law of thermodynamics**. There are several different statements of the law, each of which conveys the same information. For example

(a) **Clausius** (1850) stated it by

No heat pump, reversible or irreversible, can transfer internal energy from a low temperature reservoir to a high temperature reservoir without work being done on it by some external agent.

(b) **Kelvin** (1851)

No heat engine, reversible or irreversible, operating in a cycle, can take in heat from its surroundings, and totally convert it into work.

Thus the second law determines the direction of heat flow, and implies that mechanical energy (work) cannot be produced by extracting heat from a single reservoir without returning heat to some other reservoir at a lower temperature. The law is discussed further in chapter 26.

► 25.11 DEGREES OF FREEDOM

Each method by which a system can absorb energy is called a **degree of freedom**.

Examples

(a) **Translational energy**

A gas molecule of mass m has translational k.e.

$$\frac{1}{2}mc^2 = \frac{1}{2}mc_x^2 + \frac{1}{2}mc_y^2 + \frac{1}{2}mc_z^2$$

Its translational motion gives it three degrees of freedom.

(b) **Rotational energy**

A polyatomic gas molecule has rotational k.e. $\frac{1}{2}I\omega^2$ made up of terms like $\frac{1}{2}I_x\omega_x^2$.

Each term of this type represents a degree of freedom.

Since a gas molecule has three possible axes about which it *may* rotate, it may have up to three such degrees of freedom.

(c) **Vibrational energy**

The energy of vibration in fig. 25.7

$$E = E_k + E_p$$

which may be written

$$E = \frac{1}{2}\mu v^2 + \frac{1}{2}kx^2$$

where μ is called the **reduced mass**.

This vibration has two degrees of freedom.

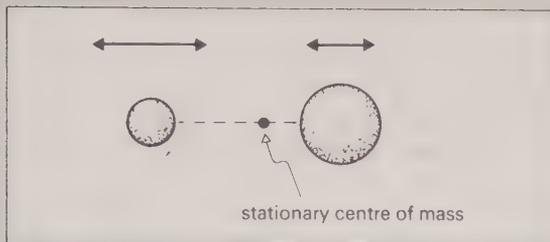


Fig. 25.7. Vibration of a diatomic molecule.

Each degree of freedom has energy written in the form

$$(\text{positive constant})(\text{variable})^2$$

Some examples are given in section 25.13.

► 25.12 EQUIPARTITION OF ENERGY

Analysis by statistical mechanics gives us the **Maxwell–Boltzmann principle**:

When we are dealing with a large number of particles in thermal equilibrium to which we can apply Newtonian mechanics, the energy associated with each degree of freedom has the same average value, and this average value depends only on the temperature.

For an example, see the *Brownian motion* (p. 135).

Application to a monatomic molecule

$$\text{For any ideal gas } pV = \frac{2}{3} N \left(\frac{1}{2} mc^2 \right) \quad (\text{p. 199})$$

$$\text{and } pV = \mu RT = \frac{N}{N_A} RT \quad (\text{p. 197})$$

$$\text{from which } \frac{2}{3} N \left(\frac{1}{2} mc^2 \right) = \frac{N}{N_A} RT$$

so

$$\frac{1}{2} mc^2 = \frac{3}{2} kT$$

where

$$k = \frac{R}{N_A}$$

k is called the **Boltzmann constant**, and has the value

$$k = 1.380 \times 10^{-23} \text{ J K}^{-1}$$

It represents the gas constant *per molecule* for an ideal gas and has dimensions $[\text{ML}^2\text{T}^{-2}\Theta^{-1}]$.

The equation of state can be written

$$pV = NkT$$

for N molecules of an ideal gas.

The equation

$$\frac{1}{2} mc^2 = \frac{3}{2} kT$$

tells us that energy $\frac{3}{2} kT$ is associated with the translational motion, which has three degrees of freedom. We deduce that for this special case, each degree of freedom has an average energy $\frac{1}{2} kT$.

More generally, at temperature T

$$\left(\begin{array}{l} \text{energy of a} \\ \text{molecule} \end{array} \right) = \left(\begin{array}{l} \text{number of} \\ \text{degrees of} \\ \text{freedom} \end{array} \right) \frac{1}{2} kT$$

► 25.13 THE KINETIC THEORY PREDICTION OF GAS MOLAR HEAT CAPACITIES

(A) Monatomic gases

Average energy per molecule $= \frac{3}{2} kT$.

Energy U for an amount of gas μ [mol] is given by

$$U = \frac{3}{2} \mu N_A kT$$

$$\therefore U = \frac{3}{2} \mu RT$$

Then

$$C_{V,m} = \frac{\Delta Q}{\mu \Delta T} = \frac{\Delta U}{\mu \Delta T}$$

since $\Delta Q = \Delta U$ when V is fixed.

$$\therefore C_{V,m} = \frac{1}{\mu} \frac{dU}{dT} = \frac{3}{2} R = 12.48 \text{ J mol}^{-1} \text{ K}^{-1}$$

and $C_{p,m} = C_{V,m} + R = 20.80 \text{ J mol}^{-1} \text{ K}^{-1}$.

$$\gamma = \frac{C_{p,m}}{C_{V,m}} = \frac{5}{3} = 1.67$$

for an ideal gas.

(B) Diatomic gases

There are 5 degrees of freedom, 3 translational and 2 rotational. No rotation occurs about the dumb-bell axis, and no additional vibration is excited near room temperature for strongly-bound molecules such as O_2 and N_2 (see p. 213).

Hence average energy per molecule = $\frac{5}{2}kT$ leading to

$$C_{V,m} = \frac{5}{2}R = 20.8 \text{ J mol}^{-1} \text{ K}^{-1}$$

and $C_{p,m} = \frac{7}{2}R = 29.1 \text{ J mol}^{-1} \text{ K}^{-1}$

$$\gamma = \frac{7}{5} = 1.40$$

for an ideal gas.

(C) Polyatomic gases

These have 6 degrees of freedom, 3 translational and 3 rotational (if there is no vibration). This gives

$$C_{V,m} = 3R, \quad C_{p,m} = 4R \quad \text{and} \quad \gamma = \frac{4}{3} = 1.33$$

For n degrees of freedom

$$\gamma = \frac{n+2}{n}$$

Compare these predictions with the experimental values for real gases near room temperature (below).

Gas	$C_{p,m}/\text{J mol}^{-1} \text{ K}^{-1}$	$C_{v,m}/\text{J mol}^{-1} \text{ K}^{-1}$	γ
He	20.8	12.5	1.67
Ar	20.8	12.5	1.67
H ₂	28.8	20.4	1.41
O ₂	29.4	21.0	1.40

The behaviour of real polyatomic gases is very complicated but some do have 6 degrees of freedom.

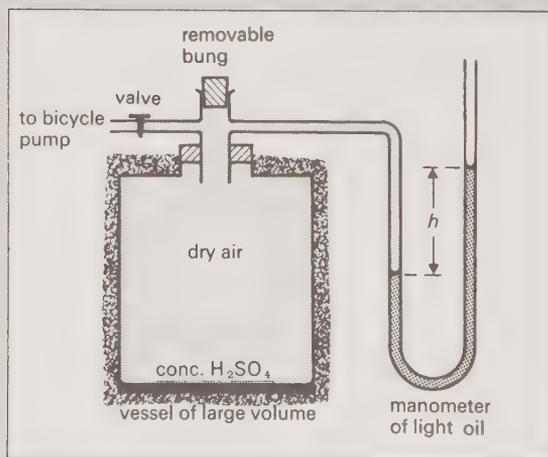


Fig. 25.8. The apparatus of Clément and Désormes.

► 25.14 THE MEASUREMENT OF γ

(A) Using the speed of sound

The pressure changes associated with acoustic wave propagation at audible frequencies are effectively adiabatic and reversible.

$$c = \sqrt{\left(\frac{\text{bulk modulus}}{\text{density}}\right)}$$

$$= \sqrt{\frac{\gamma p}{\rho}} \quad (\text{see p. 328})$$

The speeds of sound in different gases can be compared by the dust-tube method (p. 332), from which γ can be deduced.

(B) Clément and Désormes's method

Procedure

- (1) Pump in air, and allow it to reach room temperature. Record manometer reading h_1 .
- (2) Remove bung for about 1 second. The gas expands roughly adiabatically and reversibly.
- (3) Allow air to regain room temperature, and record the new reading h_2 ($< h_1$).

Calculation

Consider the fixed mass of air that does not escape. Suppose its volume increases from V_1 to V_2 .

When the tap is opened

$$p_1 V_1^\gamma = p V_2^\gamma \quad (1)$$

where $p_1 > p$ (p = atmospheric pressure).

After warming again to room temperature

$$p_1 V_1 = p_2 V_2 \quad (2)$$

Boyle's Law applied at room temperature.

$$\text{From (1)} \quad \left(\frac{V_1}{V_2}\right)^\gamma = \frac{p}{p_1}$$

$$\text{and (2)} \quad \left(\frac{V_1}{V_2}\right)^\gamma = \left(\frac{p_2}{p_1}\right)^\gamma$$

$$\text{Equating} \quad \left(\frac{p_2}{p_1}\right)^\gamma = \frac{p}{p_1}$$

$$\text{so} \quad \gamma = \frac{\log p - \log p_1}{\log p_2 - \log p_1}$$

in which

$$p_1 = p + h_1 \rho g$$

$$p_2 = p + h_2 \rho g$$

For $h_1 \ll p$, it can be shown that

$$\gamma \approx \frac{h_1}{h_1 - h_2}$$

Note. The bung size should be adjusted to optimum diameter

(a) to prevent excessive air oscillation (which requires it to be small), and

(b) to ensure a rapid and adiabatic expansion (for which it should be large).

► 25.15 VARIATION OF $C_{V,m}$ WITH TEMPERATURE

On p. 211 we showed why

$$C_{V,m} = \frac{n}{2} \cdot R$$

where n is the number of degrees of freedom of a molecule. For small (molecular) dimensions, quantum theory rather than *Newtonian* mechanics should be applied, and so the principle of equipartition of energy does not hold.

Fig. 25.9 demonstrates that the energies of rotation and vibration are *quantized*. A particular molecule may not vibrate or rotate until a sufficiently high temperature has been reached. It will then do so (fully) with its energy $\frac{1}{2}kT$ for each degree of freedom.

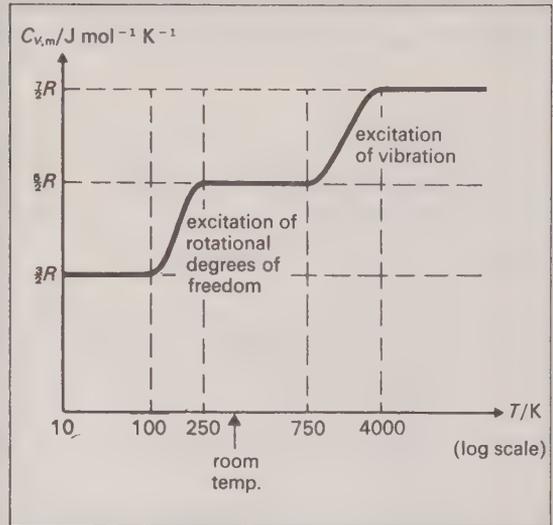


Fig. 25.9. Variation with temperature of $C_{V,m}$ for diatomic hydrogen.

Of course not every molecule will start vibrating or rotating at the same temperature. For example, above about 4000 K hydrogen molecules cease to behave as rigid dumb-bells, but show 7 degrees of freedom—3 translational, and 2 each rotational and vibrational. Below this temperature any vibrational energy will be that of the zero-point energy.

26 CHANCE, DISORDER AND ENTROPY

▶ 26.1	EXAMPLES OF IRREVERSIBLE PROCESSES	214
▶ 26.2	SOME USEFUL PROBABILITY CONCEPTS	215
▶ 26.3	STATISTICAL DESCRIPTION OF A PHYSICAL SYSTEM	215
▶ 26.4	PROBABILITY AND DISORDER	215
▶ 26.5	GASEOUS EXPANSION INTO A VACUUM	216
▶ 26.6	SOME STATISTICAL PROCESSES IN GASES	216
▶ 26.7	THERMAL CONDUCTION	217

▶ 26.8	STATISTICAL INTERPRETATION OF TEMPERATURE	217
▶ 26.9	ENTROPY AND MACROSCOPIC IDEAS	218
▶ 26.10	ENTROPY AND MICROSCOPIC IDEAS	219
▶ 26.11	ENTROPY AND THE SECOND LAW	219
▶ 26.12	THE CARNOT CYCLE	221
▶ 26.13	THE THERMODYNAMIC TEMPERATURE SCALE	222
26.14	REAL HEAT ENGINES	223
▶ 26.15	DEGRADATION OF ENERGY	223

A proper understanding of the concepts mentioned in this chapter can be achieved only after a very full explanation. The function of this book is to provide a convenient summary of the more important ideas involved. Detailed reading can be found in, for example,

(a) *The Second Law*, by Bent (Oxford University Press), 1965

(b) *PSSC Advanced Physics Supplement* (Heath), 1966

(c) *Patterns in Physics*, by Bolton (McGraw-Hill), 2nd edition, 1986

(d) *Heat Engines*, by Sandfort (Heinemann), 1962

(e) *Engines, Energy and Entropy*, by Fenn (Freeman), 1982

(f) *The Second Law*, by Atkins (Freeman), 1984.

▶ 26.1 EXAMPLES OF IRREVERSIBLE PROCESSES

The following illustrate processes that seem to happen in only one direction of time. The reverse

is never observed to occur, and so they are said to be **irreversible**. (See also chapter 25.)

(a) *Inelastic collisions*. The reduction of ordered (mechanical) k.e. by the gain of random molecular vibrational (internal) energy leads to an increase in temperature.

(b) *Heat flow*. The example of internal energy, through the agency of heat, by two bodies at different initial temperatures placed in thermal contact leads to temperature equalization.

(c) *Gaseous expansion* into a vacuum. This is discussed in more detail in 26.5.

(d) The breaking of an egg.

(e) The growth of a plant.

(f) The mixing of fluids of different colours.

Although the reverse of any of these processes would not violate any of the known conservation laws, in practice they are never observed.

Such a reverse process would require a high degree of coincidence in the random motion of a number of particles. A real macroscopic physical system may contain 10^{23} atoms or more, a number so large that the chance of a reverse process occurring is virtually nil.

► 26.2 SOME USEFUL PROBABILITY CONCEPTS

The **chance** or **probability** P of a particular event happening is defined by

$$P = \frac{\text{number of favourable observations}}{\text{total number of observations}}$$

where the total number of observations or trials is very large, and where experience suggests that the event is equally likely to happen in each trial.

(a) *Mutually exclusive events.* Let the probability that event X will happen be P_X , and the probability that Y will happen be P_Y .

Suppose that if event X happens, then event Y cannot happen. (X and Y are then said to be **mutually exclusive**.)

Then the probability that X or Y will happen is

$$P_X + P_Y$$

(b) *Sum of probabilities.* In any particular trial (such as the tossing of a coin) at least one of a set of mutually exclusive events must happen. This means that the sum of their probabilities is 1.

For example if events X and Y are the *only* two possibilities, then

$$P_X + P_Y = 1$$

(c) *Simultaneous events.* In any particular trial the probability that two *independent* events will *both* happen is the *product* of the probabilities of the two single events. For example the chance that both X and Y will happen is

$$(P_X)(P_Y)$$

► 26.3 STATISTICAL DESCRIPTION OF A PHYSICAL SYSTEM

State

(a) *Macroscopic.* A description of a physical system in terms of directly observable and measurable properties, such as temperature, pressure and volume.

(b) *Microscopic.* A description which requires a detailed knowledge of the behaviour (e.g. position and energy) of every particle of the system.

Distribution

Suppose that a given system has a total number of N particles, of which

- n_1 particles have energy E_1 ,
- n_2 particles have energy E_2 , etc.

The allowed values of E_1, E_2 etc. will be determined by the nature of the system. We will assume that N remains constant, where

$$N = n_1 + n_2 + n_3 + \dots$$

The total energy of the system is given by

$$U = n_1E_1 + n_2E_2 + n_3E_3 + \dots$$

and will be constant if the system is isolated.

The particular numbers n_1, n_2 etc. at a given instant constitute a **distribution**, or a **partition**.

Statistical equilibrium

(a) For a given macroscopic state of the system there is one distribution which is more probable than any other. When that distribution exists the system is in **statistical equilibrium**.

(b) There will be statistical fluctuations from that equilibrium which give rise to departures from the most probable distribution.

(c) The average energy of the particles is given by

$$\bar{E} = \frac{U}{N} = \frac{n_1E_1 + n_2E_2 + n_3E_3 + \dots}{n_1 + n_2 + n_3 + \dots}$$

When N is large \bar{E} has a value which determines a well-defined temperature. *Statistical equilibrium* and *thermal equilibrium* are equivalent terms.

(d) When an isolated system is not in equilibrium, the probability of its existing state is less than that of the equilibrium (most probable) state. Chance interactions between its component particles will cause the system to evolve until it reaches the most probable state: the system will have attained statistical equilibrium by moving to a distribution of *greater probability*.

The aims of statistical mechanics

Given the composition of an isolated system, statistical mechanics attempts

(a) to calculate the most probable distribution corresponding to statistical equilibrium, and hence

(b) to evaluate the macroscopic (measurable) properties of the system when in thermal equilibrium.

► 26.4 PROBABILITY AND DISORDER

A simple coin experiment

Suppose two coins are spun simultaneously. The outcome could be

HH, HT, TH or TT

There are four equally likely possible outcomes, of which two are not distinguished. The most probable outcome is thus a head and a tail, since this combination can be realized in twice as many ways as the other possibilities.

The experiment illustrates a general conclusion:

The more ways there are for realizing an event, the more likely is it to occur.

Disorder

Order and disorder are associated with the simplicity of description of a state. Two heads or two tails imply order, whereas a head and a tail imply relative disorder. The particles of a crystal are in a state of order, whereas the molecules of a gas are in a state of disorder (random energies and positions).

There are many fewer ways of arranging particles in a state of order than there are of arranging the same particles in a state of disorder. This means that, other things being equal, the state of maximum disorder is the most probable, since there are many more ways in which it can be achieved.

When large numbers are involved the most probable becomes virtually certain.

► 26.5 GASEOUS EXPANSION INTO A VACUUM

Qualitative

The molecules of fig. 26.1 are in constant random motion, and *chance* alone can move a molecule from A to B or B to A. A particular molecule has the same chance of moving either way.

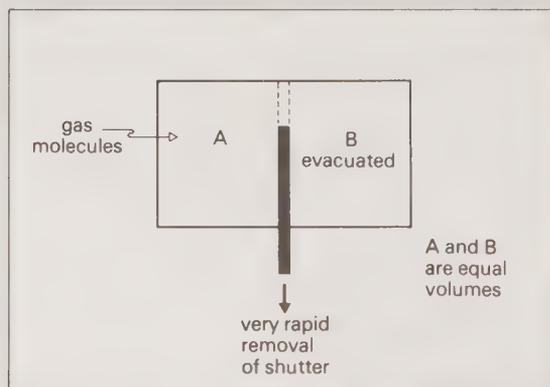


Fig. 26.1. Gaseous expansion into a vacuum.

(a) Since all the molecules start in A, the rate of motion from A to B is initially the greater.

(b) When equilibrium is reached the number of molecules per unit volume in A equals that in B.

(c) Random motion is then equally likely to cause a molecule in A to move to B, as to cause one in B to move to A.

(d) The chance that random motion alone will reverse the sequence of events leading to the establishing of thermal equilibrium is completely negligible.

In general, the distribution tends to evolve (over a time period characteristic of the system) from a less probable one to a more probable one.

Quantitative

Question: What is the probability that all the molecules will, at some later time, be again found in A?

We assume that, after some suitable small time interval, a molecule is equally likely to be found in A as in B.

(a) *One molecule.* The probability that it is in A is $\frac{1}{2}$.

(b) *Two molecules.*

(i) The probability that one is in A is $\frac{1}{2}$.

(ii) The probability that the other is (independently) in A is also $\frac{1}{2}$.

(iii) The probability that both are in A together is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$.

(c) *N molecules.* The probability that all will be found in A simultaneously is

$$\frac{1}{2} \times \frac{1}{2} \times \dots = \left(\frac{1}{2}\right)^N$$

For one mole $N = (1 \text{ mol}) \times N_A = 6 \times 10^{23}$ molecules. $\left(\frac{1}{2}\right)^{6 \times 10^{23}}$ is completely negligible: *chance* alone dictates that the process will not be observed.

► 26.6 SOME STATISTICAL PROCESSES IN GASES

The smaller the number of particles in a system the greater the chance of finding significant departures from the macroscopic averages associated with statistical equilibrium.

If N particles are involved, the variation is typically of the order of \sqrt{N} . The proportional fluctuation \sqrt{N}/N becomes smaller as N increases, being negligible when $N \approx 10^{23}$.

(a) *Brownian motion.* The size of a pollen grain or a smoke particle is sufficiently small for the bombardment by fluid particles on one side to be significantly different from that on the other. (In

$F = pA$, the pressure p is a statistical concept which is valid only for large numbers of molecules.)

(b) *The blue sky.* If electromagnetic radiation is scattered by an array of oscillators which are continuous and uniformly spaced, then the scattered radiation is coherent and causes no net transfer of energy perpendicular to the original direction of travel.

Consider a box of side l . When l approaches the wavelength of blue light, the number N of air molecules it contains is small enough for the proportional fluctuation \sqrt{N}/N to become significant. The accompanying density fluctuation causes scattering which is responsible (in part) for the blue colour of the sky.

(c) *Transport phenomena* in gases (section 24.8) are the consequence of random statistical processes.

► 26.7 THERMAL CONDUCTION

Distributing quanta between two atoms

Fig. 26.2 shows the number of ways in which different numbers of quanta can be distributed between two atoms.

Notes

(a) The more quanta there are to be shared between the two atoms, then the greater the number of possible ways of sharing them.

(b) The more ways there are for distributing the quanta amongst the atoms, the more likely it is that that arrangement will happen.

Probability and heat flow

In fig. 26.3 we imagine two solids A and B about to be put into thermal contact. A is initially at the higher temperature.

(a) Suppose that there are W_A ways of arranging quanta amongst A's particles, and W_B ways for B. Then the total number of ways of arranging these quanta is $W_A \times W_B$.

(b) The two bodies are now put into thermal contact, and reach thermal equilibrium. To achieve this, quanta move from A to B: as a consequence W_A falls to w_A , and W_B rises to w_B . The new total number of ways of arranging quanta is $w_A \times w_B$.

(c) Calculation shows that if $T_A > T_B$,

$$(w_A \times w_B) > (W_A \times W_B)$$

There are more ways of arranging the quanta between the two bodies if some move from A to B than if they move from B to A (or stay as they are).

number of quanta	distribution of quanta between atoms			number of ways
0	○ ○			1
1	⊙ ○	○ ⊙		2
2	⊙ ⊙	⊙ ○	○ ⊙	3
3	⊙ ⊙	⊙ ⊙	⊙ ⊙	4

Fig. 26.2. Distribution of quanta between atoms.

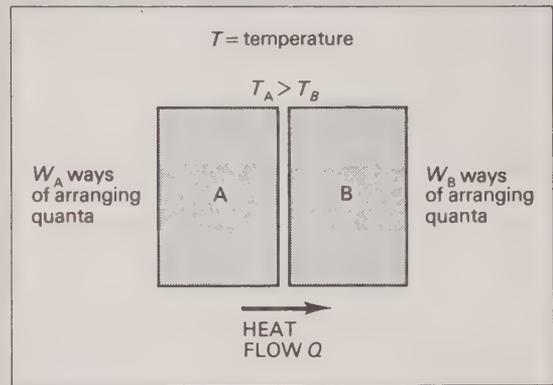


Fig. 26.3. Heat flow explained by probability.

The decrease in the number of ways for body A is more than compensated by the increase for B.

(d) The more ways in which an arrangement can be made, the more likely is it to happen. We deduce that quanta are more likely to flow from A to B than from B to A. (The numbers involved ensure that the move is virtually certain.)

(e) The physical process of *heat flow* is identified as the random spreading of quanta of energy seeking more ways in which they can be arranged.

► 26.8 STATISTICAL INTERPRETATION OF TEMPERATURE

The situation of section 26.7 leads to the following argument.

(a) When two objects with the same average number of quanta per atom are put into thermal contact, there may be more ways of arranging those quanta if some move from one body to the other than there would be without such movement.

(b) We deduce that temperature is not determined on a microscopic scale by the average number of quanta per atom. For the two bodies to stay at the same temperature, the effect of adding one quantum of energy to each must be to change the number of ways of arranging the quanta by the same factor.

(c) The addition of one quantum to a body at low temperature has a greater effect on the number of ways than the addition of one quantum has on the number for a body at a high temperature. That is, the lower the temperature, the larger the factor by which the number of ways changes.

(d) *Quantitative development.* At temperature T let there be W ways of arranging N quanta in a body, and W' ways for $(N + 1)$ quanta. Then

$$\left(\begin{array}{c} \text{factor by which the number} \\ \text{of ways is changed} \end{array} \right) = \left(\frac{W'}{W} \right)$$

The numerical value of T can be made to range from 0 to ∞ , as does temperature defined on the thermodynamic or *Kelvin* scale, by writing

$$\ln(W'/W) \propto 1/T$$

Our statistical temperature T defined in this way can be made identical to thermodynamic temperature by appropriate choice of the constant of proportionality.

► 26.9 ENTROPY AND MACROSCOPIC IDEAS

Suppose that during an infinitesimal reversible process a system at temperature T absorbs an amount of heat δQ_{rev} .

The change in entropy δS of the system is defined by the equation

$$\delta S = \frac{\delta Q_{\text{rev}}}{T}$$

$$\delta S \left[\frac{\text{J}}{\text{K}} \right] = \frac{\delta Q_{\text{rev}}[\text{J}]}{T[\text{K}]}$$

Thus for a reversible transformation from state 1 to state 2, the change in entropy

$$\Delta S = S_2 - S_1 = \int_1^2 \frac{dQ_{\text{rev}}}{T}$$

Notes

(a) ΔS depends only upon the initial and final states, and not upon the process followed.

(b) Since T is positive, ΔS is positive when heat is absorbed, and negative when it is rejected.

(c) For a reversible adiabatic transformation

$$\delta Q_{\text{rev}} = 0$$

so $\Delta S = 0$

so $S = \text{constant}$

(d) From the defining equation, for a reversible change

$$Q_{\text{rev}} = \int_1^2 T \cdot dS$$

(e) When a body is taken through a complete cycle it is returned to its original state (fig. 26.4).

$$\Delta S_{\text{cycle}} = 0$$

A transformation in which the entropy of the system does not change is called **isentropic**. (Thus a reversible adiabatic process is isentropic.)

$$\begin{aligned} Q_{\text{rev}} &= \oint T \cdot dS \\ &= \left(\begin{array}{c} \text{heat absorbed by system} \\ \text{during one cycle} \end{array} \right) \\ &= \left(\begin{array}{c} \text{work done by system} \\ \text{during that cycle} \end{array} \right) \end{aligned}$$

As an example the *Carnot cycle* is discussed in 26.12.

(f) At any temperature a single-phase substance will have unique values of U and S . These are macroscopic variables, and they can be used to describe a process in the same way as can p , V and T .

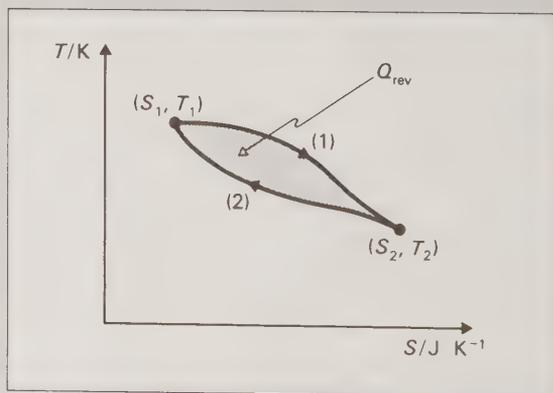


Fig. 26.4. The heat absorbed by a system undergoing a reversible cycle.

Some properties of entropy

Experiment indicates the following properties to be possessed by entropy.

(a) *Entropy tends towards a maximum* in an isolated system. (Isolation implies lack of ability to transfer mass or energy across the boundary of the system.)

(b) *Entropy is extensive*. It depends, for example, on both the volume and the number of particles within the system. (Thus the mass of a system is extensive, whereas the density is **intensive** — it is unchanged by an imaginary partition placed within the system.)

(c) *Entropy is additive*. Suppose a system has two parts of entropies S_1 and S_2 . Then the entropy of the whole is given by

$$S = S_1 + S_2$$

Calculation of entropy changes

(a) For any *reversible* process the change in entropy between states 1 and 2 is found from

$$\Delta S = S_2 - S_1 = \int_1^2 dQ_{\text{rev}}/T$$

The value of ΔS depends upon the initial and final states only.

(b) For an *irreversible* process we cannot evaluate $\int_1^2 dQ/T$ along the path concerned. Instead we try to find a *reversible* path which connects the same initial and final states, and evaluate $\int_1^2 dQ/T$ for that path.

Example: Free expansion

Consider the process of fig. 26.1, in which the volume changes *irreversibly* from V to $2V$, and suppose the gas to be ideal. Now an isothermal expansion, controlled so as to occur reversibly, would connect the same initial and final states. For either process

$$\begin{aligned} \Delta S &= S_2 - S_1 \\ &= \int_1^2 \frac{dQ_{\text{rev}}}{T} \\ &= \mu R \ln(2V/V) \\ &= \mu R \ln 2, \end{aligned}$$

which is *positive*. (We have used the fact that, since $T = \text{constant}$,

$$\Delta U = 0 \quad (\text{ideal gas})$$

giving $\Delta W = +\Delta Q$ (first law)

ΔW is calculated from $\int_1^2 p \cdot dV$.)

For this irreversible adiabatic process the entropy increases. This discussion should be compared with the probability argument in 26.5.

► 26.10 ENTROPY AND MICROSCOPIC IDEAS

Although we can measure entropy changes in practice from

$$\delta S = \frac{\delta Q_{\text{rev}}}{T}$$

for reversible processes, we want to know the fundamental significance of entropy changes in terms of particle behaviour, that is to relate entropy to statistical quantities.

Now $\ln W$, the logarithm of the number of ways in which quanta of energy can be arranged in a system, shows the three experimental properties of entropy listed in 26.9.

(a) In an isolated system W and $\ln W$ tend to a *maximum*.

(b) W and $\ln W$ depend upon the number of particles in the system, and in this sense are *extensive*.

(c) Whereas entropy is additive, numbers of ways are multiplicative, following the rule

$$W = W_A \times W_B$$

On the other hand $\ln W$ has the necessary *additive* characteristic, since

$$\ln W = \ln W_A + \ln W_B$$

We write

$$S \propto \ln W$$

or

$$S = k \ln W$$

$\ln W$ is dimensionless, and the constant k is in fact the *Boltzmann constant* (p. 211). $k = 1.38 \times 10^{-23} \text{ J K}^{-1}$, and ensures that the unit of S is consistent with the earlier definition.

In a completely ordered arrangement, such as an idealized crystalline solid of a pure substance at absolute zero, we would have

$$W = 1$$

and

$$S = 0$$

► 26.11 ENTROPY AND THE SECOND LAW

The zeroth law of thermodynamics (section 21.1) develops our concept of temperature. The first law deals with that of internal energy. The second law is bound up with the concept of entropy.

The second law

We have so far established that

(a) the more ways there are in which an event can be realized, the more likely it is to occur,

(b) the distribution which occurs in statistical equilibrium is that which is most probable, and

$$(c) \quad S = k \ln W$$

These statements tell us that the entropy of an isolated system in statistical equilibrium has the maximum value compatible with the physical conditions of the system. More formally:

The second law of thermodynamics states that the processes most likely to occur in an isolated system are those in which the entropy either increases or remains constant.

The usefulness of the law is that it tells us which processes in nature are likely to happen.

Notes

(a) For a reversible process $\Delta S = 0$: such a process can go equally well in either direction.

(b) For a system initially *not in equilibrium*

$$\Delta S > 0$$

Not until statistical equilibrium has been achieved will further changes yield

$$\Delta S = 0$$

(c) For an *irreversible* process $\Delta S > 0$, and these are the spontaneous processes which have a *natural direction* in nature. Transport processes are good examples. These include molecular diffusion and thermal conduction (below).

(d) The second law points to a *probable* course of events, not the only possible course. There may locally be *small-scale* exceptions to the law where decreases of entropy occur spontaneously.

(e) The graph of fig. 26.5 shows the typical approach to equilibrium of a characteristic one-way process.

A system with components

Suppose a system has two separate components which are put into thermal contact. Initially the total entropy

$$S = S_1 + S_2$$

$$\text{Finally} \quad S' = S_1' + S_2'$$

The net change of entropy

$$\begin{aligned} \Delta S &= S' - S \\ &= \Delta S_1 + \Delta S_2 \\ &\geq 0 \end{aligned}$$

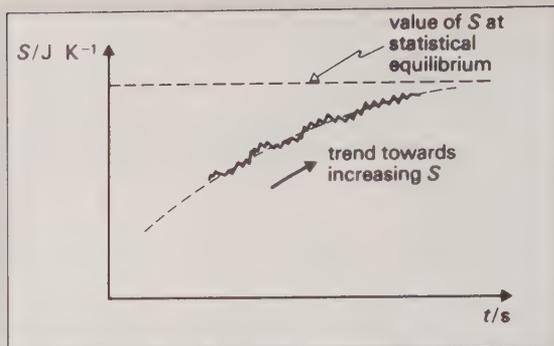


Fig. 26.5. Graph showing fluctuating value of the entropy of a system as it approaches statistical equilibrium.

If one component loses entropy (e.g. if ΔS_2 is negative), then the other component gains entropy (ΔS_1 is positive) in such a way that the net change of entropy for the whole system is either zero or positive.

The entropy changes in heat conduction

Refer to fig. 26.3. Suppose that within an adiabatic enclosure the two bodies eventually reach a common temperature T .

Principles

(a) The fact that $(T_A - T_B)$ is not infinitesimally small at all stages of the operation means that this process is *irreversible*.

(b) To calculate the energy change for the system we devise a *reversible* change which connects the same initial and final states, and apply

$$\Delta S = \int_1^2 dQ_{\text{rev}}/T$$

This can be done by means of an intermediate heat reservoir whose temperature can be adjusted to any required value.

Approximate calculation

Let T'_A and T'_B be the mean temperatures of the bodies when heat Q is rejected and absorbed, respectively.

$$\text{Then} \quad \Delta S_A \approx -\frac{Q}{T'_A} \quad \Delta S_B \approx +\frac{Q}{T'_B}$$

$$\Delta S = \Delta S_A + \Delta S_B$$

$$\approx -\frac{Q}{T'_A} + \frac{Q}{T'_B}$$

Since $T'_B < T'_A$, ΔS is *positive*, as is necessary in an *irreversible* process.

► 26.12 THE CARNOT CYCLE

The *Carnot cycle* is discussed here

- (a) as an illustration of a cyclic process,
- (b) to exemplify earlier work on entropy, and
- (c) because of its fundamental importance in the development of the thermodynamic scale of temperature.

The indicator diagram

Imagine a working substance, such as an ideal gas, contained in a cylinder by a piston. We subject it to the four processes shown in the indicator diagram of fig. 26.6.

- (1) The gas undergoes a slow reversible *isothermal expansion* by absorbing heat Q_1 at temperature T_1 . It does work.
- (2) The gas is allowed to *expand adiabatically* and reversibly. It does work, and its temperature falls to T_2 .
- (3) The gas is made to undergo a slow reversible *isothermal compression*, by having work done on it at temperature T_2 . It rejects heat Q_2 .
- (4) The gas is *compressed adiabatically* and reversibly. Work is done on it, and its temperature rises to T_1 .

The gas has been returned to its original state (p_1, V_1, T_1) , and so there is no change in its internal energy. This means that the work W done by the system during the cycle is equal to the net heat Q supplied to the system per cycle, where

$$Q = Q_1 - Q_2$$

The system is a **heat engine**.

Entropy changes during the Carnot cycle

Refer to the $T-S$ diagram of fig. 26.7, which represents the same *Carnot cycle*.

The table summarizes the entropy changes for each process.

Process	ΔW	ΔT	ΔQ	ΔS
(1) isothermal expansion	+ve	0	$Q_1(+ve)$	Q_1/T_1
(2) adiabatic expansion	+ve	-ve	0	0
(3) isothermal compression	-ve	0	$-Q_2(-ve)$	$-Q_2/T_2$
(4) adiabatic compression	-ve	+ve	0	0

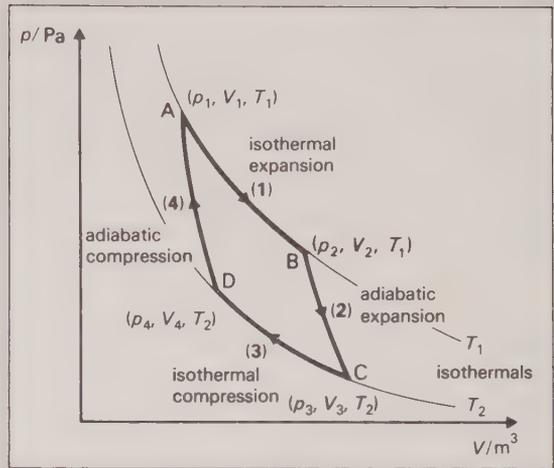


Fig. 26.6. The four steps of the Carnot cycle shown on an indicator diagram.

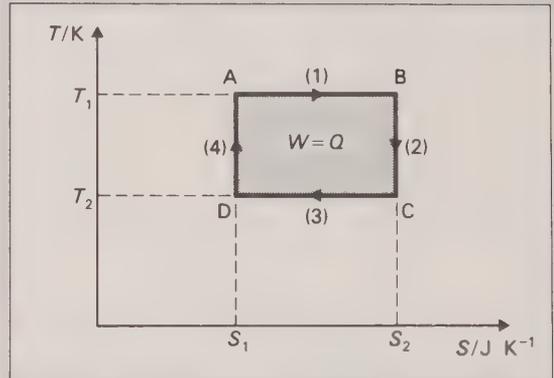


Fig. 26.7. The Carnot cycle represented on a $T-S$ diagram.

Since

$$\Delta S_{\text{cycle}} = 0$$

$$\frac{Q_1}{T_1} - \frac{Q_2}{T_2} = 0$$

$$\frac{Q_1}{T_1} = \frac{Q_2}{T_2}$$

The work done by the system per cycle is given by the shaded rectangle:

$$W = Q_1 - Q_2$$

$$= (T_1 - T_2)(S_2 - S_1)$$

Thermal efficiency

Although we have imagined the working substance to be an ideal gas, the result above would apply to any working substance.

In general, during a reversible isothermal process

$$\begin{aligned} S_2 - S_1 &= \int_1^2 \frac{dQ_{\text{rev}}}{T} \\ &= \frac{Q}{T} \quad (\text{since } T = \text{constant}) \end{aligned}$$

Thus for our *Carnot* cycle

$$Q_1 = T_1(S_2 - S_1)$$

The **efficiency** of a heat engine is found from

$$(\text{efficiency}) = \frac{\left(\begin{array}{c} \text{net work done by engine during} \\ \text{one cycle} \end{array} \right)}{\left(\begin{array}{c} \text{heat taken from the high-} \\ \text{temperature source during} \\ \text{that cycle} \end{array} \right)}$$

Here efficiency $E = W/Q_1$

$$= \frac{(T_1 - T_2)(S_2 - S_1)}{T_1(S_2 - S_1)}$$

$$E = \frac{T_1 - T_2}{T_1}$$

Notes

(a) An amount of heat Q_2 is delivered to the lower-temperature reservoir at a temperature T_2 . For 100% efficiency, i.e. for all the heat absorbed at T_1 to be converted into work, T_2 must be 0 K. As heat cannot be exhausted to a reservoir at absolute zero, a heat engine can never reach 100%.

(b) If the cycle is performed in reverse, heat Q_2 is removed from the lower-temperature reservoir, and the system constitutes a **refrigerator**. Work is done (or energy provided) by an external agent to move heat from the inside of the refrigerator to the surrounding room. In a **heat pump**, the lower-temperature reservoir could be a river or the ground, and the higher-temperature reservoir could be the inside of a house. A heat pump can be regarded as an air conditioner in reverse—outdoors is cooled and indoors is warmed. The **coefficient of performance** (or **figure of merit**) η is defined by the equations below:

(i) for a refrigerator

$$\eta_R = \frac{Q_2}{W} = \frac{T_2}{T_1 - T_2}$$

(ii) for a heat pump

$$\eta_P = \frac{Q_1}{W} = \frac{T_1}{T_1 - T_2}$$

Real refrigerators and heat pumps operate at much lower values of η than those which would be calculated from the above ideal equations.

(c) The last boxed equation summarizes *Carnot's* theorem: *The efficiency of a thermal engine operating in a Carnot cycle depends only upon the two operating temperatures, and is independent of the working substance.*

A thermal engine operating in a *Carnot* cycle is an *ideal engine*, and no engine can be more efficient than this.

► 26.13 THE THERMODYNAMIC TEMPERATURE SCALE

In the *Carnot* engine cycle, the temperatures in the equation $T_1/T_2 = Q_1/Q_2$ are ideal-gas temperatures which were defined in chapter 21 by the equation

$$T = \frac{p_T}{p_{\text{tr}}} \times 273.16 \text{ K} \quad (\text{constant-volume gas thermometer})$$

If the two temperatures θ_1 and θ_2 , between which a *Carnot* engine operates, are related by the equation $\theta_1/\theta_2 = Q_1/Q_2$, the **thermodynamic (Kelvin) temperature scale** is defined by the equation

$$\theta = \frac{Q_\theta}{Q_{\text{tr}}} \times 273.16 \text{ K}$$

On this scale, Q acts as the *thermometric property* and is independent of the properties of any particular working substance. The efficiency of a reversible engine is therefore given by

$$E = \frac{Q_1 - Q_2}{Q_1} = \frac{\theta_1 - \theta_2}{\theta_1}$$

By comparison with the expression for the efficiency of a *Carnot* engine using an ideal gas as the working substance,

$$E = \frac{Q_1 - Q_2}{Q_1} = \frac{T_1 - T_2}{T_1}$$

it is clear that $\theta = T$. This means that, if an ideal gas could be used in a constant-volume gas thermometer, the thermometer would give the thermodynamic temperature. Thus *ideal-gas scale temperatures are identical to thermodynamic scale temperatures.*

26.14 REAL HEAT ENGINES

Real heat engines are not thermodynamically reversible because of frictional forces, turbulence and loss of thermal energy. They are, therefore, less efficient than reversible engines. Their efficiency is further limited because the practical values of $(T_1 - T_2)$ tend to be small.

Example. Consider a steam engine taking steam from a boiler at 500 K and exhausting it at 320 K.

The maximum efficiency would be $(500 - 320)/500 = 0.36$.

The efficiency of the real engine would be far less than this due to its irreversibility.

Typical values for the efficiencies of real engines are:

Otto cycle (petrol engine)	0.25
Diesel cycle	0.40
Steam turbine cycle (working at high pressures)	0.65

(A) The Otto cycle

Fig. 26.8 shows the Otto cycle for a four-stroke petrol engine on an indicator diagram. The solid line shows the idealized cycle, and the dotted line is nearer to the real situation when the gas is exhausted from $e \rightarrow a$ and then replaced with new fuel-air mixture from $a \rightarrow b$. The cycle can be divided into the following stages.

(1) *The intake stroke, $a \rightarrow b$:* the combustible mixture is drawn into the cylinder at constant pressure.

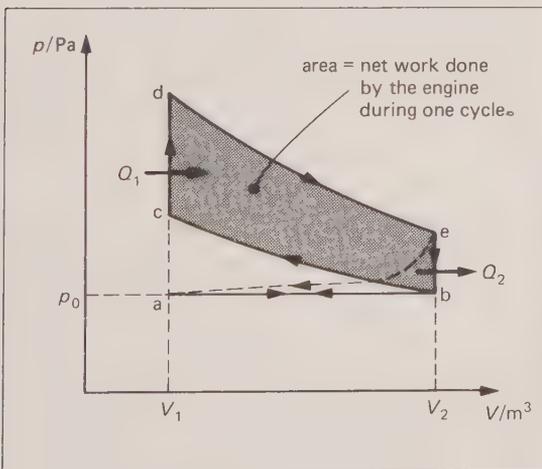


Fig. 26.8. The Otto cycle.

(2) *The compression stroke, $b \rightarrow c$:* the mixture is compressed by the piston adiabatically so that the temperature and pressure increase. The explosion results in further increase in temperature and pressure, $c \rightarrow d$, at constant volume.

(3) *The power stroke, $d \rightarrow e$:* the hot high-pressure gas expands adiabatically pushing the piston down.

(4) *The exhaust stroke, $e \rightarrow a$:* the piston pushes out the exhaust gases.

(B) The diesel cycle

Fig. 26.9 shows the diesel cycle on an indicator diagram. The constant-volume heat addition of the Otto cycle is replaced in the diesel cycle by constant-pressure heat addition, $c \rightarrow d$. Diesel engines can use higher compression ratios (V_2/V_1) and can therefore operate at higher peak pressures and higher efficiencies than Otto engines. If an Otto engine were operated at a high compression ratio, the combustible mixture would ignite before the top of the compression stroke. In the diesel engine, fuel is injected after compression and therefore precombustion is not possible.

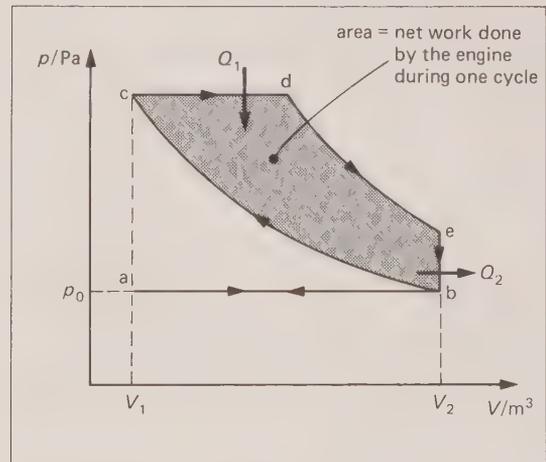


Fig. 26.9. The diesel cycle.

► 26.15 DEGRADATION OF ENERGY

Internal energy is unique amongst energy forms in that it may be converted to other forms only partially and temporarily.

In the Carnot cycle of 26.12 the engine returned an amount of heat Q_2 from the working substance

to the low-temperature reservoir at temperature T_2 . Even if no frictional or other losses are involved, i.e. if the practical design of the *machine is ideal*, only an amount $(Q_1 - Q_2)$ of the heat drawn from the high-temperature reservoir can be converted into useful work.

We can imagine a fossil fuel being burned to provide steam, the steam being made to do work, and the surplus energy being disposed of via a cooling tower. *Energy* is conserved (first law), but what is not conserved is the *availability* of some of that energy (second law). The efficiency of the engine is limited by the requirement that the entropy of an isolated system must increase or stay the same.

Such processes happen continually in nature, causing *Kelvin* to recognize the **principle of degradation of energy**. The first law controls the *quantity* of energy in a system, while the second law deals with its *quality*, which is degraded with time.

In summary

- (a) *irreversible processes continually lead to an increase in the disorder of energy,*
- (b) *the entropy of the universe increases inexorably towards a maximum, and*
- (c) *when that maximum is reached no further energy will be available for the performance of work.*

27 CHANGE OF PHASE

27.1 CHANGE OF PHASE AND LATENT HEAT L	225	27.4 BOILING	228
27.2 MEASUREMENT OF SPECIFIC LATENT HEAT l	226	27.5 MEASUREMENT OF S.V.P.	228
27.3 EVAPORATION	227	▶ 27.6 VARIATION OF VAPOUR PRESSURE WITH TEMPERATURE	229
		27.7 p - T GRAPHS AND THE TRIPLE POINT	229

27.1 CHANGE OF PHASE AND LATENT HEAT L

Molecular behaviour and energy changes

(A) Fusion

In solids molecules vibrate about fixed positions, and the pattern is *ordered*. In liquids molecules vibrate about variable positions, and the pattern is *disordered*.

To convert solid to liquid we provide energy (the **latent heat L**) which

(a) increases the p.e. on changing from the ordered to the disordered phase (while keeping roughly the same average molecular k.e., i.e. the same temperature), and

(b) does the external work (which is relatively small). If the substance contracts on melting, the external work is negative.

(B) Vaporization

The molecular separation in the gaseous phase is several times that in the liquid phase.

To convert liquid to vapour we provide energy (the latent heat L) which

(a) increases the p.e. of a molecule from about -0.9ϵ in the liquid phase (p. 128) to nearly zero in the gaseous phase (this energy enables work to be done against intermolecular forces), and

(b) does the external work, which is usually much larger than in (A).

(C) Sublimation

This is the process by which a solid is converted directly into a vapour.

In the first law of thermodynamics (p. 207),

$$L \equiv \Delta Q = \Delta U + \Delta W$$

we provide heat ΔQ , which enables the substance to do work ΔW , and which increases the internal energy by ΔU (but without a significant change of average molecular k.e.).

During a phase change ΔU represents basically a change of p.e.

Definition of specific latent heat of transformation l

Suppose heat L added to a mass m changes its phase without causing a temperature change.

The **specific latent heat of transformation l** is defined by the equation

$$l = \frac{L}{m}$$

$$l \left[\frac{\text{J}}{\text{kg}} \right] = \frac{L[\text{J}]}{m[\text{kg}]}$$

l would be called the specific latent heat of *fusion*, *vaporization* or *sublimation*. $[l] = [\text{L}^2 \text{T}^{-2}]$.

The value of l depends on the pressure at which it is measured. It is usually quoted for *standard pressure* (as are normal melting and boiling points).

Orders of magnitude

Substance	$T_{m.p.}/K$	Fusion $l/J\text{ kg}^{-1}$	$T_{b.p.}/K$	Vaporization $l/J\text{ kg}^{-1}$
water	273	33×10^4	373	23×10^5
oxygen	55	1.4×10^4	90	2.1×10^5
copper	1356	18×10^4	2573	73×10^5
lead	600	2.5×10^4	1893	7.3×10^5

Summary of phase changes shown in fig. 27.1

Note that either (a) the molecules gain *k.e.* (and at the same time *p.e.*) which shows as a *temperature increase* for a particular phase, or

(b) there is a gain of *p.e.* (with hardly any change of *k.e.*), which shows as a *phase change* at constant temperature.

In (a) the slope of the graph is inversely proportional to the substance's specific heat capacity, and in

(b) the length of a horizontal portion is proportional to a specific latent heat of transformation.

At a particular temperature molecules of any ideal gas (whatever its chemical nature) would have the same average translational *k.e.* (p. 200).

27.2 MEASUREMENT OF SPECIFIC LATENT HEAT l

In $\Delta Q = ml$, we measure m and ΔQ .

(a) Fusion

- (i) Method of mixtures.
- (ii) Bunsen's ice calorimeter.

(b) Vaporization

- (i) Method of mixtures.
- (ii) Continuous-flow, as in Henning's method.

Henning's method

This adaptation includes the principle of the *self-jacketing vaporizer* (fig. 27.2).

Allow the apparatus to warm up throughout by prolonged continuous boiling, so that the vapour jacket is at the temperature of the boiling liquid, and a *steady state* is maintained.

Measure V_1 , I_1 and the mass m_1 [kg] of liquid that evaporates and condenses in a time t [s]. Then

$$V_1 I_1 t = m_1 l + Q$$

where Q is the (small) heat loss through the jacket.

Repeat the experiment with a significantly different power input $V_2 I_2$, and collect a mass m_2 in the *same time* t .

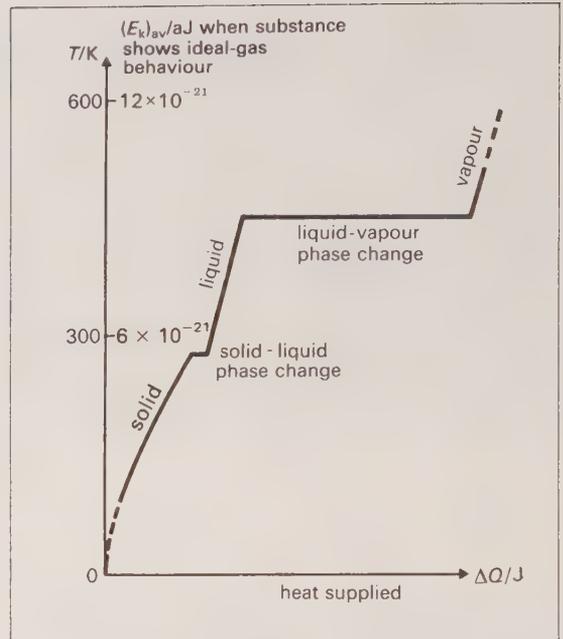


Fig. 27.1. Schematic representation of a fixed mass of substance gaining heat.

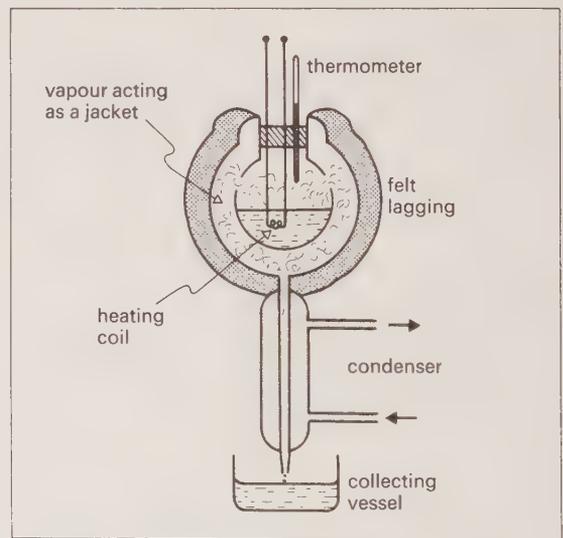


Fig. 27.2. The self-jacketing vaporizer.

$$V_2 I_2 t = m_2 l + Q$$

Q is the same because in an identical environment we have the same power loss, and it occurs for the same time. Subtracting to eliminate Q

$$l \left[\frac{\text{J}}{\text{kg}} \right] = \left(\frac{V_1 I_1 - V_2 I_2}{m_1 - m_2} \right) \cdot t$$

The method has the usual advantages of continuous-flow calorimetry (p. 192).

27.3 EVAPORATION

Molecular behaviour

Evaporation is the escape of molecules from the surface of a liquid. Upward-moving molecules near the surface may have enough energy

- (a) to escape completely, or
- (b) to escape to a position from which they are repelled back into the liquid by other molecules above the surface, or
- (c) to move only a little way before other liquid molecules pull them back.

The *rate* of evaporation for a given liquid can be increased by

- (1) increasing the surface *area*,
- (2) having faster (more energetic) molecules: this is achieved by a *temperature* increase,
- (3) having fewer molecules above the surface (by reducing the *pressure*), and/or
- (4) removing escaped molecules in a *draught*.

Evaporation occurs *at all temperatures*, but because of (2) is faster at high temperatures.

Vapour pressure

When the space above a liquid is closed, it quickly becomes **saturated** with vapour, and a *dynamic equilibrium* is established. The vapour molecules exert a **saturation vapour pressure** (s.v.p.) whose value depends only on the temperature; it is independent of any external pressure. If the volume of the space is reduced, some of the vapour liquefies, but there is no change in the pressure.

A *saturated* vapour does *not* obey the gas laws. An unsaturated vapour obeys them reasonably well.

Mixtures of gases and vapours

A vapour can be liquefied by pressure alone, whereas a gas cannot be unless it is first cooled (p. 231). To solve problems involving a mixture of the two:

- (a) apply *Dalton's law* (p. 200) to separate the total pressure into the partial pressures of the gas and vapour respectively,
- (b) apply the gas equation (p. 197) to evaluate what happens to the gas and any *unsaturated* vapour,

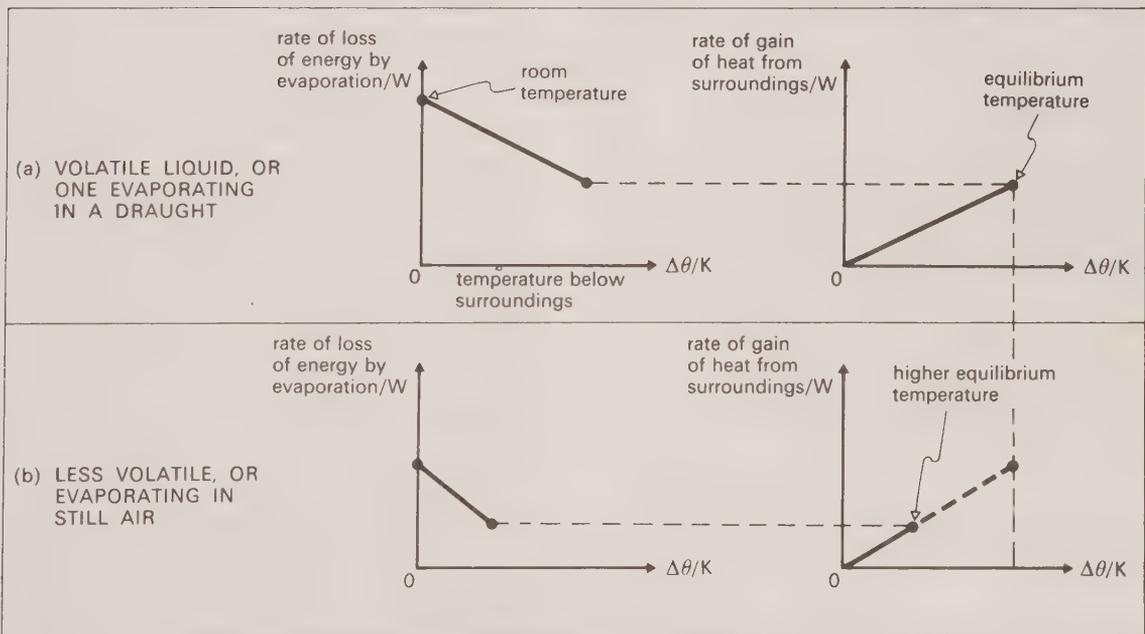


Fig. 27.3. To illustrate schematically why the steady-state temperature of an evaporating liquid is below that of its surroundings.

(c) calculate the pressure in any space saturated by vapour by reference to its s.v.p. at the temperature concerned, and

(d) calculate the final pressure by further application of Dalton's law.

Cooling due to evaporation

If it were equally likely for *any* molecule to escape, the liquid would lose energy and mass in the same proportion, and its temperature would not change. Because the *faster* molecules escape, the average molecular k.e. (and hence temperature) decreases.

The fall of temperature brought about by evaporation depends on the rate of evaporation and thus on the nature of the liquid, and on external factors. As soon as the liquid temperature falls below that of the surroundings, they give heat to the liquid.

A steady state is reached when

$$\left(\begin{array}{l} \text{rate of gain of heat} \\ \text{from surroundings} \end{array} \right) = \left(\begin{array}{l} \text{rate of loss of energy} \\ \text{by evaporation} \end{array} \right)$$

Then the temperature remains steady at a value below that of the surroundings, while the liquid mass continues to decrease. This is shown by fig. 27.3 (p. 227).

The lowering of temperature will be different at different temperatures, and can be increased by

- using a volatile liquid, and
- a strong draught.

27.4 BOILING

Boiling occurs when molecules escape in the form of bubbles of vapour from the *body* of the liquid. If we ignore surface tension effects (p. 158), this occurs when the pressure inside such a bubble is equal to the pressure in the liquid at that depth. The pressure inside a bubble will be the s.v.p. at that temperature, so

A liquid boils at the temperature at which its s.v.p. is equal to the external pressure.

Thus boiling occurs

- at a temperature controlled by the external pressure (fig. 27.4), but
- at a rate controlled by the power of the heat source.

Note that a graph which plots s.v.p. against temperature, also represents external pressure against boiling point.

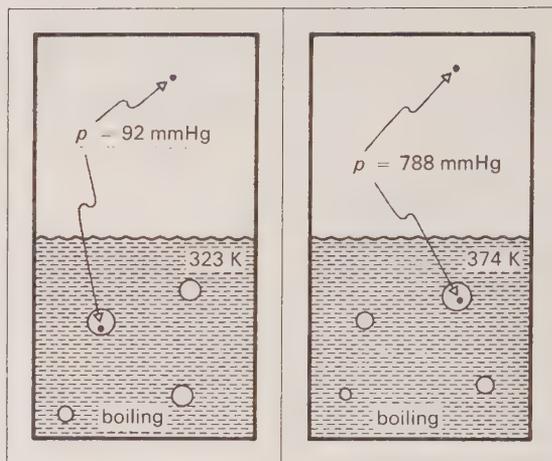


Fig. 27.4. Water boiling at different temperatures when the external pressures are different.

Standard boiling point is the temperature at which the s.v.p. of a liquid equals standard pressure (101325 Pa, which is equivalent to that exerted by a 760 mm column of mercury).

The standard boiling point of pure water is 373 K: this is in effect (for elementary practical purposes) a *definition* and not an experimental result, since it is a fixed point on the International Practical Temperature Scale.

27.5 MEASUREMENT OF S.V.P.

(A) Static (barometer) method

Usually used for s.v.p. in the range 50–400 mmHg. The sources of error include

- change with temperature of the density of mercury,
- surface tension effects from the required surplus of liquid,
- the pressure exerted by this surplus, and
- the difficulty of measuring accurately the height difference involved.

The method *cannot* measure an s.v.p. greater than atmospheric pressure.

(B) Dynamic (boiling) method

Usually used for s.v.p. from 50 mmHg upwards.

The apparatus shown in fig. 27.5 is a means of boiling liquid under a variable pressure. We measure

- the boiling point, and
- the corresponding external pressure (which

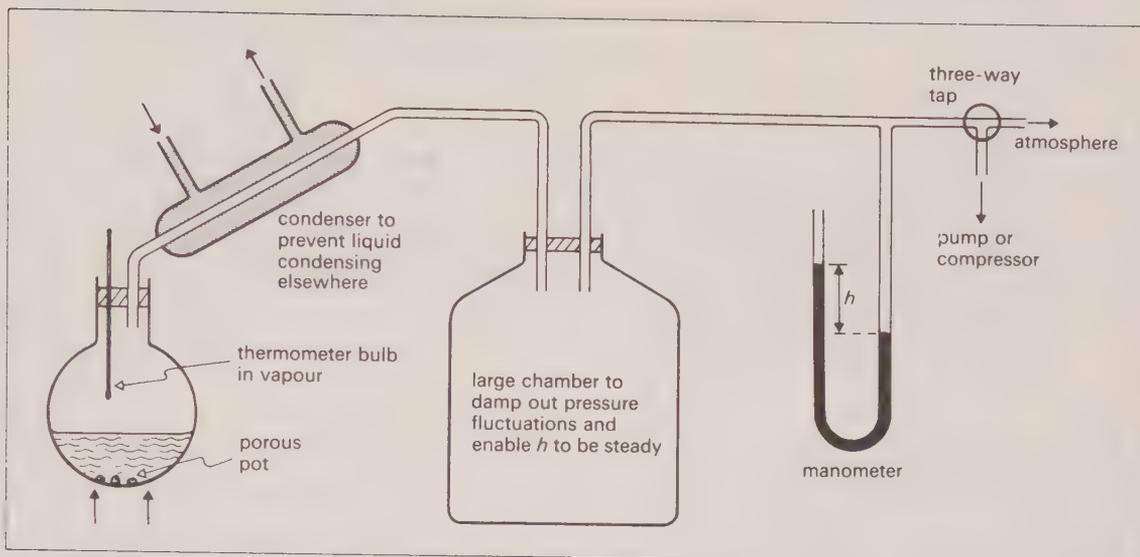


Fig. 27.5. Measurement of s.v.p. by the boiling method.

equals the s.v.p.): this is calculated from h and the atmospheric pressure.

► 27.6 VARIATION OF VAPOUR PRESSURE WITH TEMPERATURE

The laws of thermodynamics lead to the **Clausius–Clapeyron equation**

$$\frac{dp}{dT} = \frac{L_m}{T(V_f - V_i)_m}$$

L_m is the **molar heat of transformation** accompanying a phase change in which a substance has initial and final volumes V_i and V_f . p is the equilibrium vapour pressure and T is the thermodynamic temperature. Consider the phase change from liquid to vapour and make the following assumptions:

- L_m for vaporization is independent of temperature.
- $V_{f,m} \gg V_{i,m}$.
- The vapour shows ideal-gas behaviour, so that $pV_{f,m} = RT$.

The **Clausius–Clapeyron equation** becomes

$$\frac{1}{p} \frac{dp}{dT} = \frac{L_m}{RT^2}$$

Integrating,

$$\ln p = -\frac{L_m}{R} \frac{1}{T} + \text{constant}$$

(a graph of $\ln p$ against $1/T$ gives a straight line of gradient $-L_m/R$)

$$\therefore \ln \frac{p}{p_0} = -\frac{L_m}{R} \frac{1}{T}$$

The vapour pressure therefore varies with temperature according to

$$p = p_0 e^{-L_m/RT}$$

where p_0 is a constant of proportionality. This approximate relationship shows that a small temperature rise results in a large increase in vapour pressure. The expression $e^{-L_m/RT}$ is a measure of the fraction of liquid molecules that have energy greater than the amount required to overcome the attractive forces of neighbouring molecules.

27.7 p – T GRAPHS AND THE TRIPLE POINT

Fig. 27.6 (overleaf) shows the graphs obtained when experimental values of pressure are plotted against temperature (as in simple boiling point experiments). We can identify three curves.

(a) The **sublimation curve** connects points at which vapour and solid exist in equilibrium.

(b) The **vaporization curve** shows vapour and liquid existing in equilibrium, and is more widely known as the graph plotting s.v.p. against temperature. The curve begins at the triple point, and ends at the critical point (p. 231).

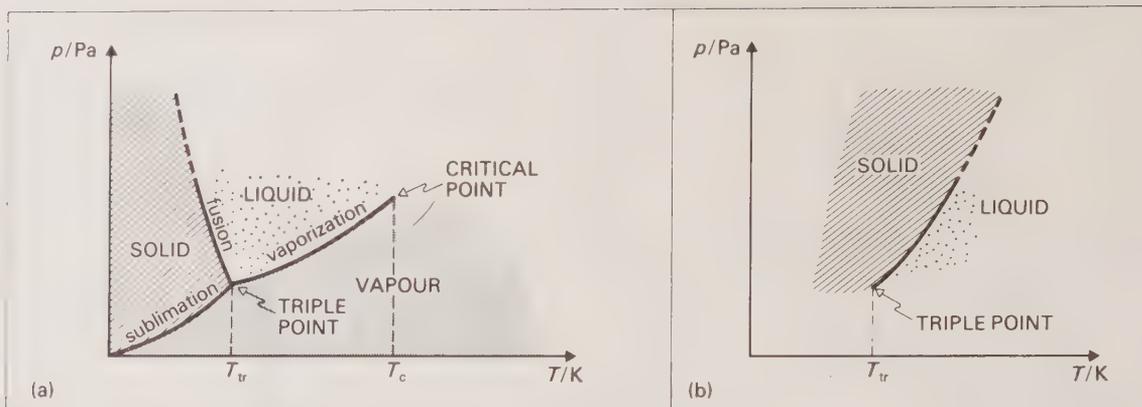


Fig. 27.6. (a) A p - T graph for a substance (such as water) which contracts on melting. (b) The fusion curve for a substance which expands on melting. (Neither graph to scale.)

(c) The **fusion curve** shows liquid and solid existing in equilibrium. It has no definite upper end-point.

The triple point

(a) The sublimation, vaporization and fusion curves meet at a single point called the **triple point**.

(b) That *unique* temperature at which water

exists in solid, liquid and gaseous phases simultaneously in equilibrium (i.e. its triple point) is chosen as the upper fixed point of the thermodynamic temperature scale. For historical reasons it is *allotted* the value 273.16 K.

(c) On a p - V diagram such as fig. 28.2 the triple 'point' would appear as a *line* rather than as a point.

(d) The triple point occurs at a unique pressure (611 Pa in the case of water), unlike m.p. and b.p.

28 REAL GASES

28.1 ANDREWS'S EXPERIMENTS (1863)	231
28.2 DEVIATIONS FROM THE GAS LAWS	232
28.3 INTERNAL ENERGY OF REAL GASES	233

28.4 VAN DER WAALS'S EQUATION	234
28.5 PRINCIPLES OF GAS LIQUEFACTION	235

28.1 ANDREWS'S EXPERIMENTS (1863)

Using the apparatus of fig. 28.1 *Andrews* measured the volume of the trapped carbon dioxide for different values of the applied pressure. The pressure was calculated by assuming that air behaved as an ideal gas.

The experiment was repeated over a wide range of temperatures, and the results represented as a series of isothermals on an indicator diagram.

In fig. 28.2 (overleaf) the thick lines (full or dotted) separate the areas where it is possible for the carbon dioxide to exist as gas, vapour, liquid, etc.

(A) The 321.2 K isothermal

The curve is approximately hyperbolic, as would be the curve for an ideal gas.

(B) The 304.2 K isothermal

This is called the **critical isothermal**, because it represents the dividing line between those temperatures at which CO_2 can be liquefied, and those at which it cannot. (*Andrews* measured the **critical temperature** to be 304.0 K.)

(C) The 286.2 K isothermal

At A the vapour is about to be compressed. The volume decreases as the pressure is increased (similar to the ideal-gas behaviour). At B the s.v.p. for that temperature has been reached: this means that further pressure increases are not possible while saturated vapour is present. As the volume is reduced, so the vapour continues to condense, with the pressure remaining constant at the s.v.p. until the point C, where all is liquid. Liquids are not easily compressed, so further great increases of pressure cause very little decrease in volume.

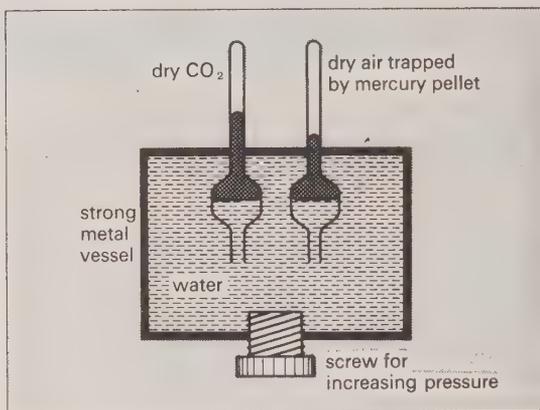


Fig. 28.1. *Andrews's apparatus.*

Some conclusions

(a) CO_2 cannot be liquefied above a certain temperature: it is reasonable to suppose that every gas has its own **critical temperature** T_c , above which it cannot be liquefied by increasing the pressure.

A fluid in the gaseous phase is named a **gas** above its critical temperature, and a **vapour** below.

The **critical pressure** p_c is the pressure necessary to produce liquefaction at the critical temperature.

	He	H ₂	O ₂	CO ₂	NH ₃	H ₂ O
critical temperature T_c/K	5	33	155	304	406	647
critical pressure $p_c/10^5 \text{ Pa}$	2.2	13	50	73	111	218

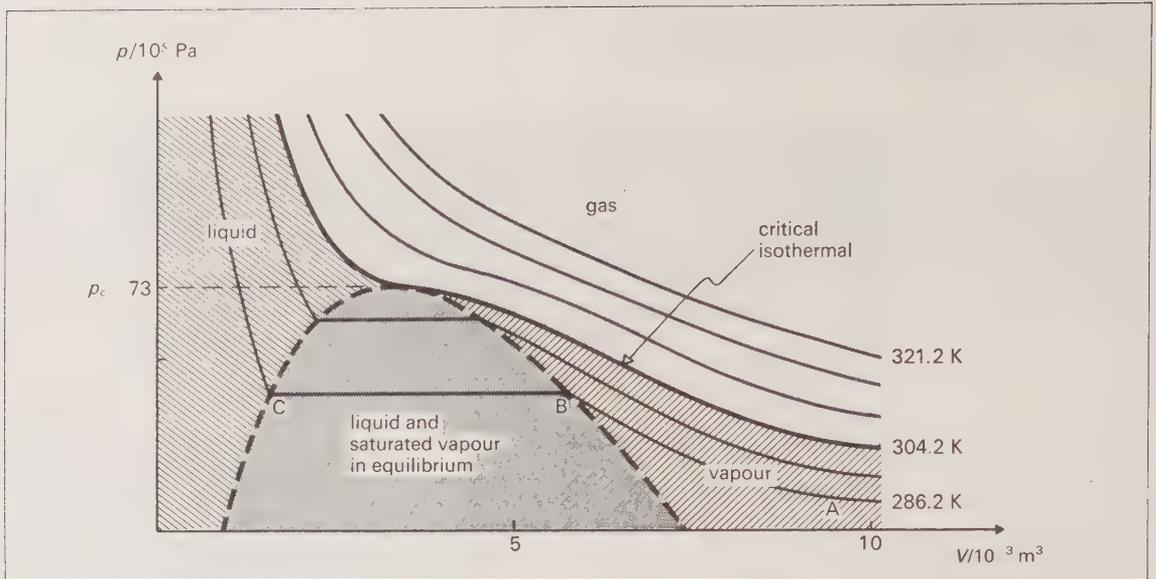


Fig. 28.2. Andrews's isothermals for 1 kg of carbon dioxide.

One can also define a **specific critical volume** v_c for a gas, which is the volume of 1 kg of gas subjected to its critical pressure at its critical temperature. The value for CO_2 is $2.2 \times 10^{-3} \text{ m}^3 \text{ kg}^{-1}$.

(b) The isothermals emphasize that carbon dioxide does not show the behaviour of an ideal gas unless either

- (i) the temperature is high, and/or
- (ii) the pressure is very low.

An ideal gas would never liquefy.

28.2 DEVIATIONS FROM THE GAS LAWS

For all *real* gases

(a) experiments (such as those of *Andrews* and *Amagat*) show that $pV/\mu T$ (p. 197) does vary with temperature and pressure, and

(b) the assumptions of the theory that led to $p = \frac{1}{3}\rho\bar{c}^2$ (p. 199) do not hold.

Deviations from $pV/\mu T = \text{constant}$ are small for the so-called *permanent* gases at normal pressures because normal temperatures are far above their critical temperatures.

Fig. 28.3 illustrates how, for all real gases, pV increases with p , provided the pressure is large enough.

We will consider two possible breakdowns for our kinetic theory assumptions:

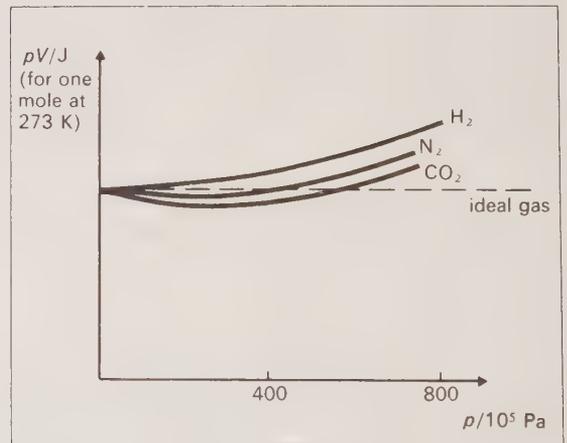


Fig. 28.3. Deviations from the gas laws shown by three different gases. (Not to scale.)

(i) The *repulsive* intermolecular forces are operative at relatively large distances. (In other words the volume of the actual molecules may be comparable with the volume of the space occupied by the gas.)

(ii) The *attractive* intermolecular forces are operative at relatively large distances. They may cause the formation of **molecular complexes** (p. 234), and the result is a reduction in the pressure exerted by the gas on the faces of its container.

(The effect of the container's walls is considered in detail on p. 234.)

(i) is more important at small volumes (and so high pressures), when effect (ii) may be masked. For some gases (depending on the temperature) the two effects may counterbalance near a particular value of p .

The Boyle temperature

The **Boyle temperature** is the temperature at which pV is most nearly constant over a wide range of pressures, because the repulsive and attractive intermolecular forces produce equal but opposing effects.

Suppose that in fig. 28.4 we can write

$$pV = A + Bp + Cp^2 + \dots \text{ for a real gas.}$$

Then $A = \mu RT$, and B and C etc. are progressively smaller factors, all of which are zero for an *ideal* gas. They depend on temperature, but not pressure.

For a *real* gas $B = 0$ at the Boyle temperature, and the value of C is negligible.

28.3 INTERNAL ENERGY OF REAL GASES

(A) Joule's experiment (1845)

Suppose a mass of real gas expands adiabatically (and irreversibly) into a vacuum. Then

- (a) the gas does zero *external* work, but
- (b) the energy required to separate the gas molecules (the *internal* work) is taken from the random molecular translational k.e. This process is accompanied by a temperature decrease.

Joule tried an experiment of this kind, but the high heat capacity of the apparatus made it too insensitive for the very small temperature change to be detected.

(B) The Joule–Kelvin experiment (1852)

Suppose a mass of gas at volume V_1 and (high) pressure p_1 is forced through a *porous plug* to a region where its volume becomes V_2 and the (low) pressure is p_2 (fig. 28.5).

Two effects are superimposed:

- (a) since $V_2 > V_1$, there will be an increase in intermolecular p.e., which will cause cooling, and
- (b) the *net* work done *by* the gas is

$$(p_2V_2 - p_1V_1)$$

Thus there may be either a cooling or a heating effect, depending on whether the real gas is below or above its *Boyle* temperature. (For an *ideal* gas $p_1V_1 = p_2V_2$.)

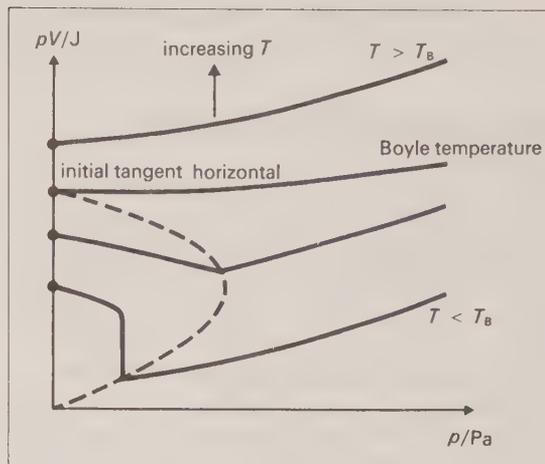


Fig. 28.4. The Boyle temperature illustrated by the variation of pV with p for a fixed mass of one gas.

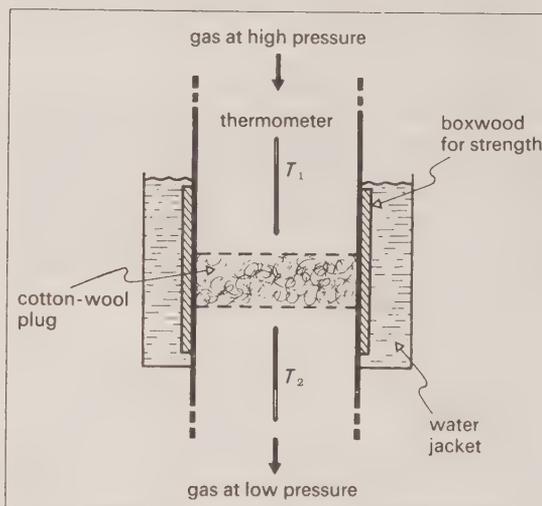


Fig. 28.5. The Joule–Kelvin porous-plug apparatus.

The observed temperature change is the *net* result of effects (a) and (b).

Experimentally *Joule* and *Kelvin* found that hydrogen showed a heating effect (because of effect (b) above), whereas *most* gases showed a cooling effect which was greater than could be accounted for by effect (b). This is evidence for attractive intermolecular forces.

The **inversion temperature** is that temperature at which a gas shows no change of temperature. Above its inversion temperature a gas will show a *net* heating effect, and below its inversion tempera-

ture a *net cooling* effect. (The inversion temperature obviously exceeds the *Boyle* temperature.)

28.4 VAN DER WAALS'S EQUATION

$$p_i \left(\frac{V_i}{\mu} \right) = RT$$

is the **equation of state** for an amount μ of an *ideal* gas. *Van der Waals's* equation is the result of one attempt to find such an equation for a *real* gas.

(A) Effect of repulsive intermolecular forces

Consider 1 mol of a real gas, and let it occupy a volume V_m . The actual volume of the gas molecules will reduce by b (the molar **covolume**) the free volume in which random molecular motion can take place.

Then $(V_m - b) = V_i/\mu$ is the molar volume which can be altered by changes of temperature and pressure, and which corresponds to the molar volume of an ideal gas.

(B) Effect of attractive forces

Attractive forces between non-ideal gas molecules may cause two or more molecules to become bound into loose complexes. Thus if (for example) *two* molecules are bound together, a **dimer** is formed. Since the total number of separate particles present is thereby reduced, so also is the number of moles of separate particles. We could indicate this *decrease* in μ by writing

$$\mu_{\text{real}} < \mu_{\text{ideal}}$$

but it is more common to *add* a correction factor to the pressure.

(Note that because a dimer and a monomer are in thermal equilibrium, on average they have equal translational k.e.: a dimer would *not* have twice the average momentum of a monomer.)

Suppose we write

(a) p_i = the pressure that *would* have been measured at the walls of the container if the real gas had been ideal,

(b) p = the pressure as normally understood (i.e. that *actually* measured for the real gas),

(c) p_a = the correction that must be added to p to take account of the formation of complexes.

Then
$$p_i = (p + p_a)$$

A consideration of the chemical equilibrium between monomers and dimers (beyond the scope of this book) indicates that we should write:

$$p_a = a/V_m^2$$

where the 'constant' a depends on temperature. We conclude

$$p_i = \left(p + \frac{a}{V_m^2} \right)$$

(C) The effect of the container walls

A molecule approaching very close to the wall of the container will experience an attractive force which will increase its momentum towards the wall. During this process the wall molecules will gain an equal but oppositely directed momentum, so that the overall effect of the collision at the wall is to impart the *same net momentum to the wall* as that which would have been imparted in the absence of the attractive force.

The fact that the range of the forces exerted by the wall is very small compared to the dimensions of the gas container thus causes the *walls* to have zero effect on the pressure.

Summary

Experiment shows that the 'constants' a and b vary from gas to gas, as we should expect. We now rewrite the ideal-gas equation of state

$$p_i \left(\frac{V_i}{\mu} \right) = RT$$

in the following form for a real gas

$$\left(p + \frac{a}{V_m^2} \right) (V_m - b) = RT$$

This is called **van der Waals's equation of state**.

Orders of magnitude

Substance	$a/\text{N m}^4 \text{ mol}^{-2}$	$b/\text{m}^3 \text{ mol}^{-1}$
hydrogen	2.5×10^{-2}	2.7×10^{-5}
nitrogen	1.4×10^{-1}	3.9×10^{-5}
carbon dioxide	3.6×10^{-1}	4.3×10^{-5}

Note carefully the units that we allot to a and b . Their dimensions are $[\text{ML}^5\text{T}^{-2}\text{N}^{-2}]$ and $[\text{L}^3\text{N}^{-1}]$ respectively.

The predictions of van der Waals's equation

Fig. 28.6 (opposite) shows the isothermals that are predicted by *van der Waals's* equation, which is a cubic in V_m .

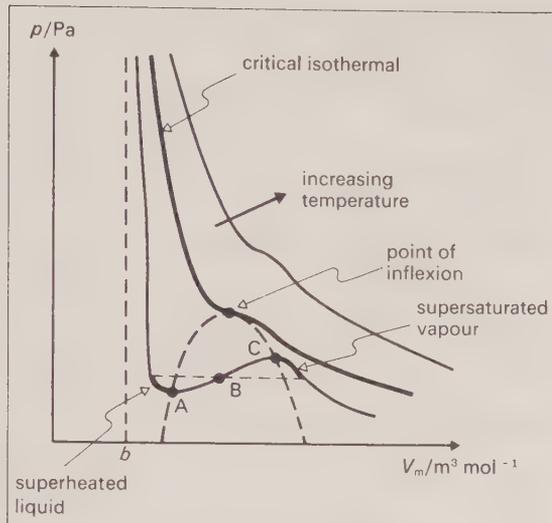


Fig. 28.6. Isotherms for a van der Waals gas.

The predicted curve ABC cannot be realized experimentally, but otherwise the isotherms resemble those plotted by Andrews for CO_2 .

At large values of V_m the equation reduces to the ideal-gas equation, so the isotherms are hyperbolic.

28.5 PRINCIPLES OF GAS LIQUEFACTION

A gas may be cooled by

- mixing it with a cool fluid;
- making it do external work in an adiabatic expansion;
- making it (while below its inversion temperature) expand through a porous-plug from a region of high pressure to one of low pressure (the Joule–Kelvin effect).

A gas may be liquefied by

- reducing the temperature of a vapour to below its boiling point at normal pressures;
- using an evaporating liquid to cool a gas below its critical temperature, and then increasing the pressure, as in the **cascade process** (Pictet);
- cooling the gas by Joule–Kelvin expansion, and using the cool gas to reduce the temperature of gas which is about to expand—this is called **regenerative cooling**;
- making the gas do external work in a nominally reversible adiabatic expansion, as in *Claude's process*;
- combining a Joule–Kelvin expansion with *Claude's process*, as in the *Collins process* for liquefying helium.

The table shows some temperatures of importance in gas liquefaction, together with some characteristic values for comparison.

Name of temperature	Origin of definition	Page	Value, in kelvins, for		
			N_2	H_2	He
standard melting point	liquid and solid in equilibrium at standard pressure	225	63	14	3.5
triple point	solid, liquid and vapour in equilibrium	230	63	13.8	2.2
standard boiling point	liquid and vapour in equilibrium at standard pressure	225	77	20.3	4.27
critical temperature	coexisting liquid and vapour phases have the same density—no pressure can cause liquefaction at higher temperatures	231	121	33	5.2
Boyle temperature	initial tangent of $pV-p$ curve is horizontal, and pV is most constant over a wide range of pressures	233	420	114	16
inversion temperature	gas shows no change of temperature in a Joule–Kelvin throttling process	233	650	190	25

29 THERMAL CONDUCTION

29.1	THE MECHANISMS OF CONDUCTION	236
29.2	TEMPERATURE GRADIENT	237
29.3	DEFINITION OF THERMAL CONDUCTIVITY λ	237
29.4	THERMAL RESISTANCE	238
29.5	COMPARISON WITH ELECTRICAL CONDUCTIVITY	239

29.6	THE COMPOUND SLAB	239
▶ 29.7	USE OF THE DEFINING EQUATION	240
29.8	MEASUREMENT OF CONDUCTIVITY	241
29.9	CONDUCTIVITY OF A GOOD CONDUCTOR	241
29.10	CONDUCTIVITY OF A POOR CONDUCTOR	241

29.1 THE MECHANISMS OF CONDUCTION

Thermal conduction is a transport phenomenon (p. 202) in which a temperature difference causes the transfer of thermal energy from one region of a hot body to another region of the same body which is at a lower temperature. It takes place in the direction in which the temperature decreases: it therefore tends to equalize the temperature *within the body*.

(A) In gases

Conduction is the result of collisions between fast and slow mobile molecules. Such a collision transfers k.e. from the fast to the slow molecule.

(B) In liquids and solids

There are two processes:

(a) In *metals* (which possess **conduction** electrons) the positive ions of a lattice are in thermal equilibrium with the electron gas.

The electrons have random speeds as high as 10^6 m s^{-1} , but otherwise behave rather like gas molecules, and transfer thermal energy from the hot region of the lattice to the cold relatively quickly. (They will also do this in a liquid metal.)

(b) The molecules of liquids and solids vibrate about fixed positions with a vibrational energy that increases with temperature. *Coupling* between neighbouring molecules allows transfer of part of

this vibrational energy from energetic (hot) to less energetic (cold) regions.

Conduction by the electron gas suggests a close relationship with *electrical conductivity*, which is discussed on p. 239.

(C) Phonons

The principal mechanism for the transfer of thermal energy through a non-metallic solid is by lattice vibrations, which travel through the material as extremely high frequency ($\sim 10^{13} \text{ Hz}$) acoustic waves. **Phonons** are the quanta of lattice vibrational energy just as photons are the quanta of light energy. Phonons travel through the substance at the speed of sound ($\sim 4 \text{ km s}^{-1}$) but diffusion of thermal energy takes place far more slowly because the phonons have frequent collisions with lattice atoms. This results in absorption and re-emission of phonon energy because their frequencies *match* the lattice frequencies.

In metals the conduction electrons have much higher frequencies than those of the lattice atoms and so scattering of phonons is less significant. The thermal diffusion rate in metals is rapid because of the high electron speeds ($\sim 1000 \text{ km s}^{-1}$). In general, metals are good conductors compared with non-metals, emphasizing the important contribution of electron diffusion to thermal conductivity. However, some non-metallic crystalline materials are good thermal conductors because of high phonon speeds due to strong bonds and atoms of small mass.

29.2 TEMPERATURE GRADIENT

Suppose **steady-state** conditions hold, so that the common temperature θ at any point does not vary with time (fig. 29.1).

The **temperature gradient** is defined to be $d\theta/dx$, which is

- (a) for the unlagged bar, the slope of the graph at a given point, and
- (b) for the perfectly insulated bar, $\Delta\theta/\Delta x$, a constant.

In each case it is *negative* because the heat flows in the direction in which x increases, and θ decreases in this direction. The lines of heat flow are

- (a) *divergent* for the unlagged bar, but
- (b) *parallel* for the perfectly insulated bar, since there is zero *lateral* heat loss.

29.3 DEFINITION OF THERMAL CONDUCTIVITY λ

We first define λ for the general situation (lines of heat flow divergent), and then simplify the equation for parallel heat flow.

Experiment shows that for small values of $\Delta\theta$ and Δx in fig. 29.2

- (a) $\Delta Q \propto \Delta t$, and $\Delta Q \propto A$ when $\Delta\theta$ is fixed.
- (b) $\Delta Q \propto \Delta\theta/\Delta x$, the *average* temperature gradient, when Δt and A are fixed.
- (c) The direction of heat flow (here that of x) is that in which θ *decreases*, so is *opposite* to the temperature gradient.

Combining these results

$$\frac{\Delta Q}{\Delta t} \approx -\lambda A \frac{\Delta\theta}{\Delta x}$$

where λ is introduced below.

Suppose $\Delta x \rightarrow 0$, so that we have a limiting situation in which we consider a *cross-section*.

The **thermal conductivity** λ of a substance is defined by the equation

$$\frac{dQ}{dt} = -\lambda A \frac{d\theta}{dx}$$

in which the symbols are defined in fig. 29.2.

$$\frac{dQ}{dt} [\text{W}] = -\lambda \left[\frac{\text{W}}{\text{m K}} \right] A [\text{m}^2] \frac{d\theta}{dx} \left[\frac{\text{K}}{\text{m}} \right]$$

The insertion of the negative sign ensures that the constant of proportionality λ is positive.

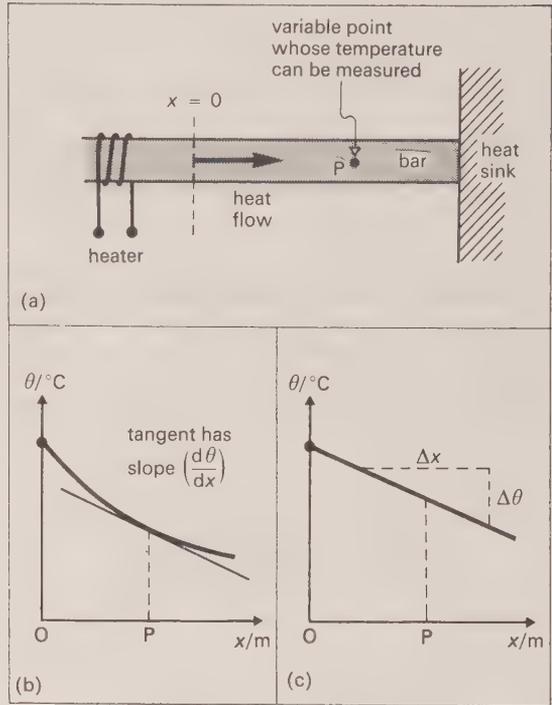


Fig. 29.1. Temperature gradient; (a) the situation, (b) and (c) graphs drawn for unlagged and perfectly insulated bars respectively.

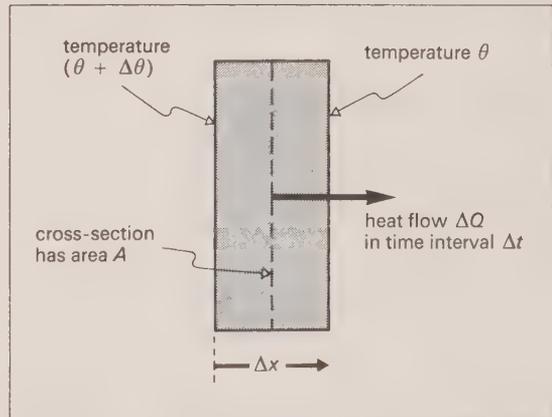


Fig. 29.2. Situation used to define λ .

Notes

(a) λ is numerically equal to the rate of heat flow through a cross-section of unit area at which there is unit temperature gradient. It has the unit $\text{W m}^{-1} \text{K}^{-1}$, and the dimensions $[\text{MLT}^{-3} \Theta^{-1}]$.

(b) Typical values

Substance	$\lambda/W\ m^{-1}\ K^{-1}$	Measured at T/K
copper	3.9×10^2	293
asbestos	8×10^{-2}	293
air	2.4×10^{-2}	273
hydrogen	14×10^{-2}	273

Although these vary by a factor of about 10^4 , the variation of electrical conductivity between good conductors and good insulators is greater by many orders of magnitude (p. 374).

(c) When the lines of heat flow are parallel, the temperature gradients at all cross-sections are equal.

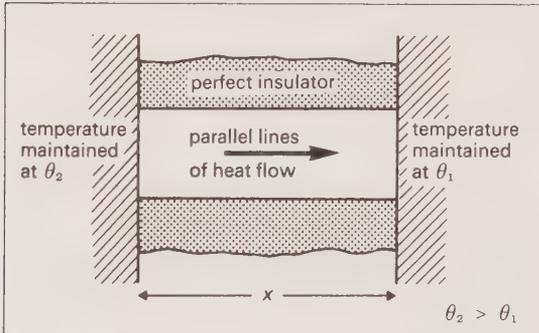


Fig. 29.3. Situation for a special case of the defining equation for λ .

For the situation of fig. 29.3

$$\frac{d\theta}{dx} = \frac{\theta_2 - \theta_1}{x}$$

and the defining equation for λ can be written

$$\frac{Q}{t} = \lambda A \left(\frac{\theta_2 - \theta_1}{x} \right)$$

a form which is applicable to most elementary problems.

Because the internal energy of the rod remains constant with time, the heat Q crossing any cross-section of the rod in time t is the same.

(d) λ varies slightly with temperature, but for calculation purposes this variation is usually negligible.

(e) The calculation of λ for an ideal gas is given on p. 202.

29.4 THERMAL RESISTANCE

The heat transfer coefficient, the thermal conductance or U -value of a particular thermal conductor is defined by the equation

$$U\text{-value} = \frac{\Delta Q / \Delta t}{A \Delta \theta}$$

$$U\text{-value} [W\ m^{-2}\ K^{-1}] = \frac{\Delta Q / \Delta t [W]}{A [m^2] \Delta \theta [K]}$$

It is also equal to $\lambda / \Delta x$ and has dimensions $[MT^{-3}\ \Theta^{-1}]$. Heating engineers make use of U -values in the design of industrial heat exchangers, central heating systems and house insulation. The following are some typical values for buildings:

Thermal conductor	U -value/ $W\ m^{-2}\ K^{-1}$
single-glazed window with metal frame	5.6
double-glazed window with metal frame	3.2
brick wall with air cavity	1.5
solid floor with two exposed edges	0.45

The thermal resistance coefficient, X , of a particular thermal conductor is defined by the equation

$$X = \frac{1}{U\text{-value}} = \frac{A \Delta \theta}{\Delta Q / \Delta t}$$

$$X [m^2\ K\ W^{-1}] = \frac{A [m^2] \Delta \theta [K]}{\Delta Q / \Delta t [W]}$$

It is also equal to $\Delta x / \lambda$ and has dimensions $[M^{-1}\ T^3\ \Theta]$. The thermal resistance, R' , of a sample of material is defined by the equation

$$R' = \frac{\Delta \theta}{\Delta Q / \Delta t}$$

$$R' [K\ W^{-1}] = \frac{\Delta \theta [K]}{\Delta Q / \Delta t [W]}$$

From above, it can be seen that $R' = X/A = \Delta x / \lambda A$ and is therefore analogous to electrical resistance, $R = l / \sigma A$.

If thermal conductors are arranged in adjacent layers, e.g. brick and plaster, the total thermal resistance is the sum of the separate resistances as the layers are *in series*. If the thermal conductors are *in parallel*, e.g. a window and a wall, the total thermal resistance is the common temperature difference across them divided by the sum of their rates of heat flow. Compare these results with the combinations of *electrical* resistance (p. 383).

$$\left(\frac{\lambda}{\sigma}\right) \propto T$$

the thermodynamic temperature, for most pure metals.

These experimental results can be explained by the theory of the behaviour of electrons in metals, which, though important, is beyond the scope of this book.

29.5 COMPARISON WITH ELECTRICAL CONDUCTIVITY

In the table below we compare some analogous quantities.

The close analogy between the defining equations for λ and σ , and the fact that electrons are largely responsible for transport of thermal energy and electric charge in a metal, both underline the close relationship that exists between these two processes.

Wiedemann and Franz (1853) pointed out that, at a given temperature, the ratio (λ/σ) is nearly the same for all pure metals.

Lorentz (1872) showed in addition that, except at low temperatures,

29.6 THE COMPOUND SLAB

Suppose that in fig. 29.4 (overleaf)

(a) a steady state has been achieved (so that θ does not depend on t , and specific heat capacities are not involved), and

(b) that the lines of heat flow are parallel.

It follows that the rate of heat flow through all cross-sections is the same.

Using
$$\frac{dQ}{dt} = -\lambda A \frac{d\theta}{dx}$$

we have

$$\frac{Q}{t} = \lambda_1 A \left(\frac{\theta_3 - \theta_2}{x_1}\right) = \lambda_2 A \left(\frac{\theta_2 - \theta_1}{x_2}\right)$$

from which θ_2 (for example) could be calculated.

Heat conduction		Electrical conduction	
quantity	symbol	symbol	quantity
temperature difference	$\Delta\theta$	ΔV	potential difference
rate of heat flow	$\Delta Q/\Delta t$	$I = \Delta Q/\Delta t$	rate of charge flow
temperature gradient	$\Delta\theta/\Delta x$	$\Delta V/\Delta x$	potential gradient (numerically equal to field strength)
thermal resistance	$R' = \frac{\Delta\theta}{\Delta Q/\Delta t}$ $= \frac{\Delta x}{\lambda A}$	$R = \frac{\Delta V}{\Delta Q/\Delta t}$ $= \frac{l}{\sigma A}$	electrical resistance
thermal conductivity	λ	σ	electrical conductivity
under conditions of zero lateral heat loss		if the electric current is steady	
$\frac{Q}{t} = \lambda A \left(\frac{\theta_2 - \theta_1}{x}\right)$		$I = \sigma A \left(\frac{V_2 - V_1}{l}\right)$	

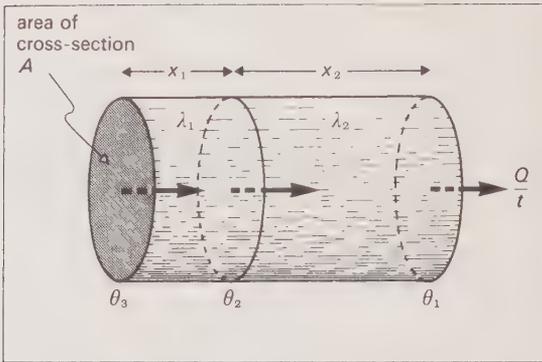


Fig. 29.4. Steady heat flow through a compound slab.

Notes

(a) Suppose $\lambda_1 > \lambda_2$, so it follows that

$$\left(\frac{\theta_3 - \theta_2}{x_1}\right) < \left(\frac{\theta_2 - \theta_1}{x_2}\right)$$

The temperature gradient is *greater* within the poorer conductor.

(b) The situation is of practical importance when a poor conductor (such as an air-film) controls the rate of heat flow to a good conductor (such as the copper base of a saucepan). It may then be convenient to replace the compound slab for calculation purposes by its equivalent thickness of one material.

► 29.7 USE OF THE DEFINING EQUATION

It is emphasized that the equation

$$\frac{Q}{t} = \lambda A \left(\frac{\theta_2 - \theta_1}{x}\right)$$

is a simplification that is not applicable in most real situations. The situation at a given cross-section is described by

$$dQ/dt = -\lambda A(d\theta/dx)$$

and in general this equation must be integrated to find the rate of heat flow. As illustrations of possible complications, note that

- (a) at low temperatures λ may vary significantly with θ ,
- (b) the value of A may depend upon x (as in the example below), and
- (c) lines of heat flow are generally not parallel, which means that $d\theta/dx$ is usually *not a constant*.

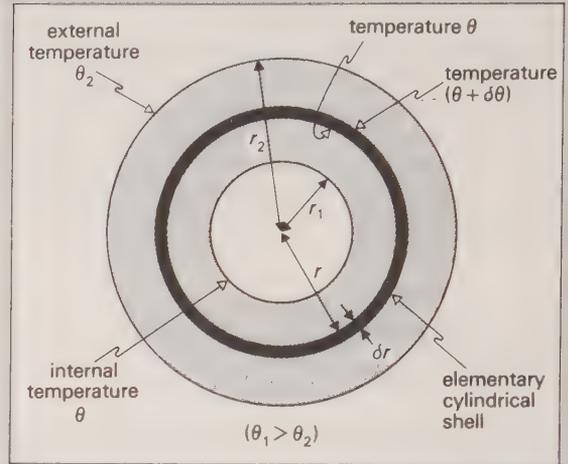


Fig. 29.5. Radial heat flow between a pair of coaxial cylinders.

Radial lines of heat flow

Refer to fig. 29.5. The aim is to find the rate at which heat is conducted from the inner cylinder to the outer cylinder under steady-state conditions. Both cylinders have length l . We assume that λ is independent of θ .

To apply

$$dQ/dt = -\lambda A(d\theta/dx)$$

to the elementary cylindrical shell shown, we note that

- (a) $A = 2\pi rl$,
- (b) the lines of heat flow are radial, so that $\delta\theta/\delta r$ is dependent upon r , and
- (c) $\delta\theta$ is negative.

Substituting
$$\frac{dQ}{dt} = -\lambda(2\pi rl) \frac{d\theta}{dr}$$

$$\frac{dQ}{dt} \int_{r_1}^{r_2} \frac{dr}{r} = -2\pi\lambda l \int_{\theta_1}^{\theta_2} d\theta$$

$$\frac{dQ}{dt} \ln\left(\frac{r_2}{r_1}\right) = 2\pi\lambda l(\theta_1 - \theta_2)$$

which gives
$$\left(\frac{dQ}{dt}\right) = \frac{2\pi\lambda l(\theta_1 - \theta_2)}{\ln(r_2/r_1)}$$

The result is relevant to both the electrical and the thermal insulation of a cable dissipating electrical energy.

29.8 MEASUREMENT OF CONDUCTIVITY

We aim to use

$$\frac{Q}{t} = \lambda A \left(\frac{\theta_2 - \theta_1}{x} \right)$$

which implies:

- (a) We must prevent lateral heat flow so that the lines of heat flow are parallel.
- (b) We must use a well-insulated rod for a good conductor. This means that the large x and small A make $(\theta_2 - \theta_1)$ sufficiently big to measure.
- (c) We must use a disc for a bad conductor. This reduces lateral conduction, and the large A and small x make Q/t sufficiently big to measure.

For fluids we adapt the disc method for a bad conductor, but it is difficult to obtain an accurate value for λ , because

- (i) there will nearly always be convection, and
- (ii) thermal energy transferred by convection and radiation, for which it is difficult to account, becomes large relative to that transferred by conduction (since λ is small). One can correct for radiation through a gas by a control experiment in a vacuum.

29.9 CONDUCTIVITY OF A GOOD CONDUCTOR

Using the apparatus shown in fig. 29.6, we measure the following:

The mass m of cooling water that flows in a time t

Temperature increase of water = $\Delta\theta$
Average temperature gradient

between P and $Q = \left(\frac{\theta_2 - \theta_1}{x} \right)$

Cross-sectional area = A

Then using

$$\frac{Q}{t} = \lambda A \left(\frac{\theta_2 - \theta_1}{x} \right)$$

$$\left(\frac{m}{t} \right) c \Delta\theta = \lambda A \left(\frac{\theta_2 - \theta_1}{x} \right)$$

in which c is the specific heat capacity of water. λ is then the only unknown.

The difference between the electrical rate of energy supply, and the rate of energy absorption by the cooling water provides a useful check on power losses by lateral conduction. (We have assumed these to be zero.)

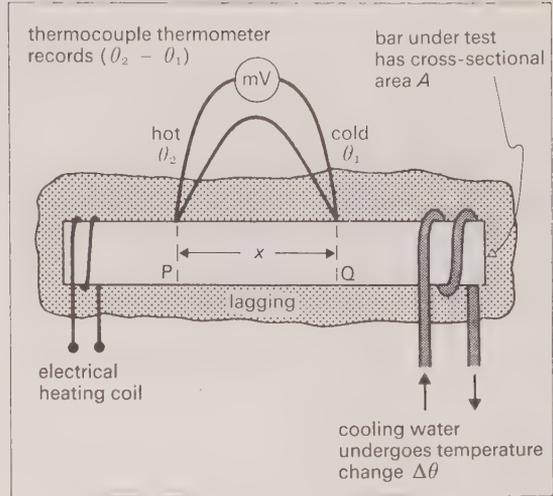


Fig. 29.6. The principle of Searle's method.

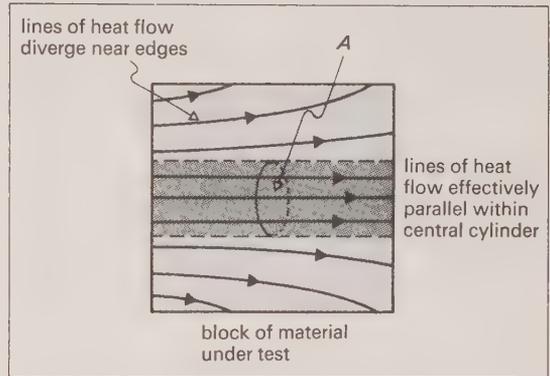


Fig. 29.7. The principle of the guard-ring.

Improvement

A better approximation to parallel heat flow is obtained by using a guard-ring, shown in fig. 29.7.

The idea behind the guard-ring should be compared to that used in the parallel-plate capacitor (p. 360). In this example the outer layers of the block constitute the guard-ring for the central cylinder.

29.10 CONDUCTIVITY OF A POOR CONDUCTOR

These symbols refer to fig. 29.8 (overleaf).

θ_0 = room temperature,

A = cross-sectional area of sample and identical copper discs,

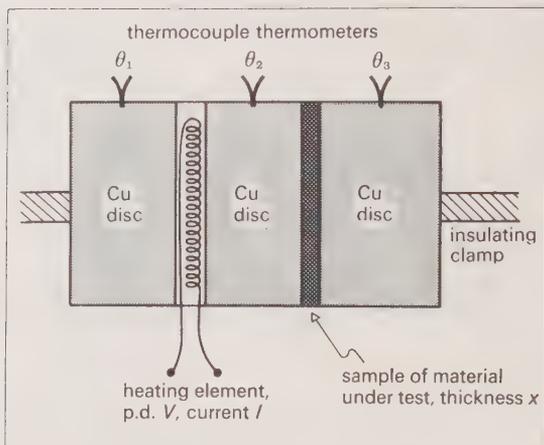


Fig. 29.8. One version of Lees's disc method. (Not to scale.)

A_1 , A_2 , A_3 and A_s = emitting surface areas of discs and sample, all with identical enamelled surfaces,

P = power emitted from unit area of these surfaces, for unit temperature difference between surface and surroundings. (P would be measured in $\text{W m}^{-2} \text{K}^{-1}$.)

Allow a steady state to be attained.

(a) Power supplied electrically equals rate of heat loss from surfaces:

$$VI = P \left[A_1(\theta_1 - \theta_0) + A_2(\theta_2 - \theta_0) + A_s \left(\frac{\theta_2 + \theta_3}{2} - \theta_0 \right) + A_3(\theta_3 - \theta_0) \right]$$

in which P , the only unknown, can be calculated.

(b) Effective rate of heat flow across sample equals average of heat entering and heat leaving in each second:

$$\begin{aligned} \lambda A \left(\frac{\theta_2 - \theta_3}{x} \right) &= \frac{1}{2} P \left\{ \left[A_s \left(\frac{\theta_2 + \theta_3}{2} - \theta_0 \right) + A_3(\theta_3 - \theta_0) \right] \right. \\ &\quad \text{ENTERING} \\ &\quad \left. + \left[A_3(\theta_3 - \theta_0) \right] \right\} \\ &\quad \text{LEAVING} \end{aligned}$$

Since P has been found in (a), λ is the only unknown.

Notes

(a) Good thermal contact between the sample and copper is aided by using glycerol.

(b) To reduce convection currents in liquids, one

(i) uses the apparatus in a horizontal position, and

(ii) heats from above.

30 THERMAL RADIATION

30.1 THE NATURE OF RADIATION	243
30.2 DETECTION OF RADIANT ENERGY	243
▶ 30.3 THE BLACK BODY	244
▶ 30.4 IMPORTANT TERMS	245
30.5 PRÉVOST'S THEORY (1792)	246

▶ 30.6 KIRCHHOFF'S LAW (1858)	246
▶ 30.7 STEFAN'S LAW (1879)	247
▶ 30.8 ENERGY IN THE BLACK-BODY SPECTRUM	248
30.9 RADIATION PYROMETERS	248

30.1 THE NATURE OF RADIATION

Thermal radiation is energy that travels from one place to another by means of an electromagnetic wave motion (p. 282). When absorbed by matter it may increase the vibrational or translational k.e. of atoms or molecules: this increase of internal energy will usually become apparent as a temperature increase. *All* frequencies of the electromagnetic spectrum produce such a heating effect when waves are absorbed.

The vibrational frequencies of atoms at room temperature are about 10^{14} Hz (p. 325). A wave of frequency about 10^{14} Hz would cause *resonance* to occur, and in general would thus be efficient at transferring the electromagnetic wave energy to the electrically charged particles of which matter is composed.

Waves whose frequencies are close to 10^{14} Hz are called **infra-red** waves. They are both radiated and absorbed by bodies at normal temperatures.

For convenience the infra-red part of the spectrum can be classified thus:

	Near	Intermediate	Far
$\lambda/10^{-6}$ m	0.8–3	3–10	10 – (1000)
ν/Hz	4×10^{14}	$10^{14} - 3 \times 10^{13}$	3×10^{11}

Infra-red radiation is emitted when thermal agitation causes changes in the vibrational and rotational energy states of molecules. A study of this radiation helps us to elucidate molecular structure.

It has all the usual *properties* of an electromagnetic wave motion (p. 288).

A **diathermanous** body, such as a calcium fluoride prism, is one that absorbs little of the radiation passing through it (cf. a body *transparent* to visible light).

An **adiathermanous** body, such as a mass of water, is one that absorbs strongly (cf. a body *opaque* to visible light).

For example glass is diathermanous when

$$0.4 \times 10^{-6} \text{ m} < \lambda < 2.5 \times 10^{-6} \text{ m}$$

but adiaathermanous for longer wavelengths such as 10^{-5} m. This fact can be used to explain the action of a greenhouse.

30.2 DETECTION OF RADIANT ENERGY

Photodetectors depend upon the photoelectric effect, and photoconductivity. They respond if a **photon** of incident radiation has a certain minimum energy (and so maximum wavelength (p. 505)).

Thermal detectors are *thermometers*, that is they respond to the temperature change that accompanies the absorption of *all* frequencies.

The human skin is one example; others include

(a) The sensitive differential air thermometer.
 (b) *Boys's radiomicrometer* (1889), in which a thermocouple is incorporated into a moving-coil galvanometer suspended by a quartz fibre.

(c) *The thermopile*, which consists of about 25 or more thermocouples joined in series. Radiation falls on the blackened hot junction, while the cold junction remains shielded.

(d) *Langley's bolometer* (1881) is a differential resistance thermometer (fig. 30.1, overleaf).

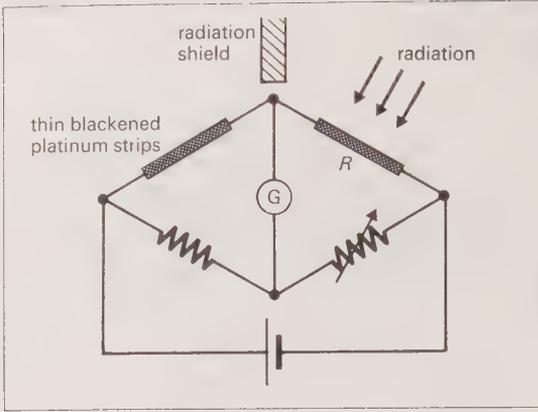


Fig. 30.1. The principle of Langley's bolometer.

Procedure

- (1) Balance the bridge.
- (2) Allow the incident radiation to change R to $(R + \Delta R)$ using one arm.
- (3) Rebalance the bridge, and calculate ΔR and hence the temperature change ΔT .
- (4) Repeat, using the arm that was previously shielded.

The bolometer is made more sensitive if the platinum is replaced by a semiconductor of high negative temperature coefficient of resistance (p. 382).

► 30.3 THE BLACK BODY

Fig. 30.2 shows what may happen when radiation is incident on a body.

One must distinguish carefully between

- (a) absorbed and re-radiated energy,
- (b) transmitted energy, and
- (c) scattered and reflected energy.

The concept of a black body

By analogy with visible optics, a **black body** is one that absorbs *all* the radiation (of all wavelengths) falling on it. It is (like the ideal gas of kinetic theory) an *idealized concept* that can nearly be realized in practice, and which is invaluable for the formulation of the laws of radiation.

All the energy incident on the hole in the cavity of fig. 30.3 is absorbed.

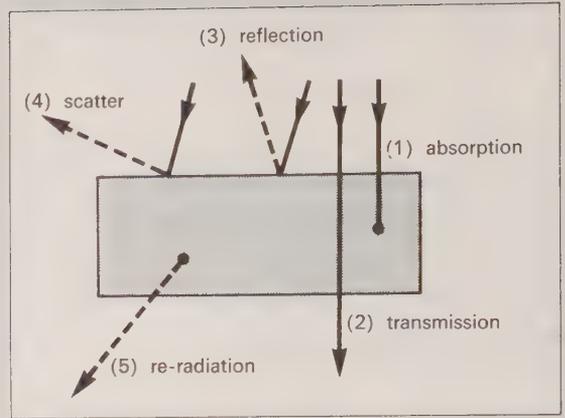


Fig. 30.2. The possible destination of radiation incident on matter.

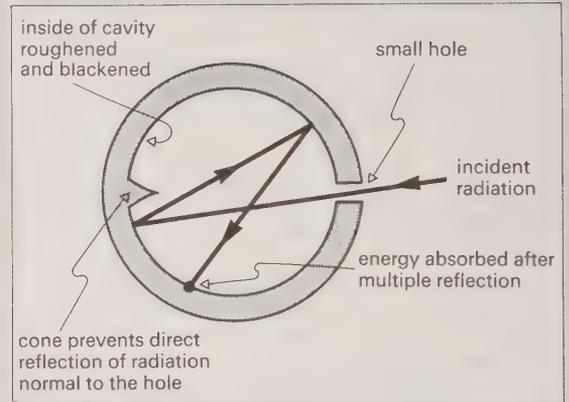


Fig. 30.3. Practical realization of a black body (absorber).

Black-body radiation

The radiation inside an **equal temperature enclosure** depends only upon the *temperature*: it is *not* controlled by the *nature* of the walls.

A *small hole* in the side of the *cavity* of fig. 30.4 then approximates to a *perfect radiator*.

The radiation emitted is called variously **black-body radiation**, *full radiation*, *cavity radiation* or *temperature radiation*, and has the characteristic feature that the intensity of each frequency has some well-defined value which is determined by the temperature (p. 247).

If the cavity is at (say) 2000 K, the small hole (the 'black' body) emits visible radiation whose colour is characteristic of that temperature.

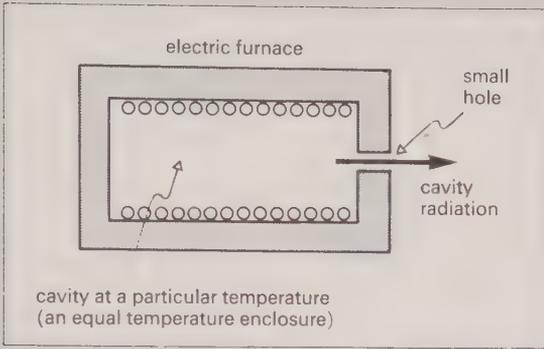


Fig. 30.4. Practical realization of a black body (radiator).

Black-body temperature is the temperature of a black body which emits radiation of the same description as the body under consideration.

► 30.4 IMPORTANT TERMS

(A) Reflectance and transmittance

Suppose radiant power Φ_0 is incident on a surface, and power Φ_r is reflected.

Then the **reflectance** (reflexion factor) $\rho = \Phi_r/\Phi_0$; τ , the **transmittance** (transmission factor), is defined in a similar way.

Neither has a unit, and they are not discussed further in this book.

(B) Spectral radiant exitance M_λ

Suppose a surface of area A radiates power Φ_λ in the waveband λ to $(\lambda + \delta\lambda)$.

The **spectral radiant exitance** M_λ of the surface for this waveband is defined by the equation

$$M_\lambda \delta\lambda = \frac{\Phi_\lambda}{A}$$

$$M_\lambda \left[\frac{\text{W}}{\text{m}^2} \right] \delta\lambda[\text{m}] = \frac{\Phi_\lambda [\text{W}]}{A[\text{m}^2]}$$

Spectral radiant exitance has the unit W m^{-3} .

(C) Total radiant exitance M

Suppose a surface of area A radiates power Φ over all wavelengths (so that Φ is the total power radiated).

The **total radiant exitance** M of the surface is defined by the equation

$$M = \frac{\Phi}{A}$$

$$M \left[\frac{\text{W}}{\text{m}^2} \right] = \frac{\Phi[\text{W}]}{A[\text{m}^2]}$$

Notes

- (a) M_λ and M have different dimensions.
- (b) It follows that

$$M = \int_0^\infty M_\lambda d\lambda$$

(c) M_λ and M for a black body will be written in this book as $M_{\lambda,B}$ and M_B respectively.

(D) Spectral emissivity ϵ_λ

The **spectral emissivity** ϵ_λ of a surface for the waveband λ to $(\lambda + \delta\lambda)$ is defined by

$$\epsilon_\lambda = \frac{M_\lambda}{M_{\lambda,B}}$$

Notes

- (a) ϵ_λ is dimensionless, and has no unit.
- (b) $\epsilon_\lambda = 0$ for a perfect transmitter or a perfect reflector.
- (c) $\epsilon_\lambda = 1$ for a perfect emitter (such as a hole in a cavity).
- (d) $\epsilon_\lambda \leq 1$ for all non-luminescent bodies.
- (e) $M_\lambda = \epsilon_\lambda M_{\lambda,B}$

$$\left(\begin{array}{c} \text{spectral} \\ \text{radiant} \\ \text{exitance of} \\ \text{a surface} \end{array} \right) = \left(\begin{array}{c} \text{spectral} \\ \text{emissivity} \end{array} \right) \left(\begin{array}{c} \text{spectral radiant} \\ \text{exitance of a} \\ \text{black body at the} \\ \text{same temperature} \end{array} \right)$$

(E) Spectral absorptance α_λ

Suppose radiant power $\Phi_{\lambda,0}$ in the waveband λ to $(\lambda + \delta\lambda)$ is incident on a surface, and that the surface absorbs power $\Phi_{\lambda,a}$.

The **spectral absorptance** (absorption factor) α_λ of the surface is defined for this waveband by

$$\alpha_\lambda = \frac{\Phi_{\lambda,a}}{\Phi_{\lambda,0}}$$

Notes

- (a) α_λ is dimensionless, and has no unit.
- (b) $\alpha_\lambda = 0$ for a perfect transmitter or a perfect reflector.
- (c) $\alpha_\lambda = 1$ for a perfect absorber (i.e. a black body).

Compare these notes with the corresponding notes (a) to (c) for ϵ_λ .

(F) The solar constant

Suppose Φ_0 is the total power radiated from the Sun to an area A placed normal to the radiation at the edge of the Earth's atmosphere, while the Earth is at its mean distance from the Sun. (Φ_0 is called the **radiant flux**.)

The quantity Φ_0/A is called the **solar constant**, and is the **irradiance** E of this particular surface. Its value is found by experiment to be 1.35 kW m^{-2} , and can be used (together with *Stefan's Law*) to give an estimate of the surface black-body temperature of the Sun.

30.5 PRÉVOST'S THEORY (1792)

The theory states:

Every body radiates energy at a rate controlled by its thermodynamic temperature T .

We deduce

- (a) when $T_{\text{body}} > T_{\text{surroundings}}$, there is a net loss of radiant energy,
- (b) when $T_b < T_s$, there is a net gain of radiant energy, but
- (c) when $T_b = T_s$, equilibrium is attained, but the energy exchange continues—the equilibrium is dynamic.

Notes

- (a) If the temperature of a good absorber is not to rise spontaneously above that of its surroundings, it must also be a good emitter of radiation. (See *Kirchhoff's Law*, below.)
- (b) A thermometer immersed in a diathermanous medium (such as air) will record a temperature influenced by the environment from which it receives radiation, as well as by the medium.

► **30.6 KIRCHHOFF'S LAW (1858)**

The law which will be deduced applies for a specified waveband λ to $(\lambda + \delta\lambda)$, and a specified temperature T .

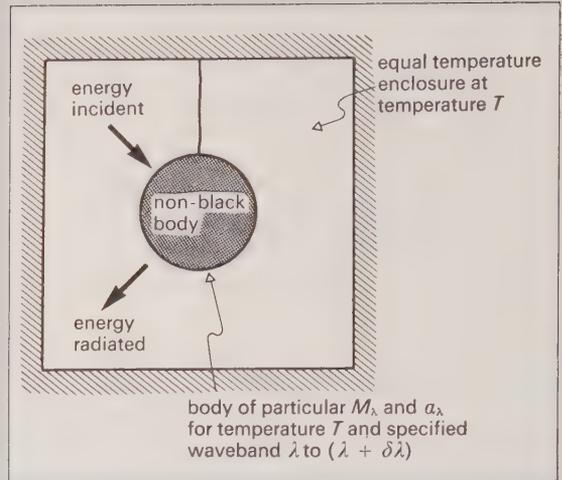


Fig. 30.5. Derivation of Kirchhoff's Law.

Suppose $E_{\lambda,0}$ is the radiant power in the specified waveband incident on unit area of the body of area A (fig. 30.5).

The rate of absorption of this energy by the body is

$$E_{\lambda,a}A = \alpha_\lambda(E_{\lambda,0}A) \quad (\text{def. of } \alpha_\lambda)$$

The rate of emission of radiant energy is

$$A(M_\lambda \delta\lambda) \quad (\text{def. of } M_\lambda)$$

Since the body is in an equal temperature enclosure, these are equal, so

$$\alpha_\lambda E_{\lambda,0}A = AM_\lambda \delta\lambda$$

$$\frac{M_\lambda}{\alpha_\lambda} = \left(\frac{E_{\lambda,0}}{\delta\lambda} \right)$$

Now $E_{\lambda,0}$ and $\delta\lambda$ are controlled by the waveband $\delta\lambda$ and the temperature T , but not by the nature of the body.

It follows that M_λ/α_λ has the same value (for a given waveband and temperature) for all bodies. This is **Kirchhoff's Law**.

For a black body the constant of Kirchhoff's Law is given by

$$\frac{M_\lambda}{\alpha_\lambda} = \frac{M_{\lambda,B}}{1}$$

We deduce

(a) this constant is the spectral radiant exitance $M_{\lambda,B}$ of a black body, and

$$(b) \quad \alpha_\lambda = M_\lambda/M_{\lambda,B} \quad (\text{from Kirchhoff's Law})$$

$$= \epsilon_\lambda \quad (\text{def. of } \epsilon_\lambda)$$

For a specified waveband and temperature, the spectral emissivity of a body equals its spectral absorptance.

Demonstration

Refer to fig. 30.6.

(A) Emission

Excited sodium atoms in a bunsen flame emit characteristic yellow light (p. 273).

(B) Absorption

The observed D-lines are less bright than the rest of the spectrum, as Na vapour (at the same temperature as in (A)) has absorbed the energy of this wavelength, and reduced its intensity by re-radiation in all directions.

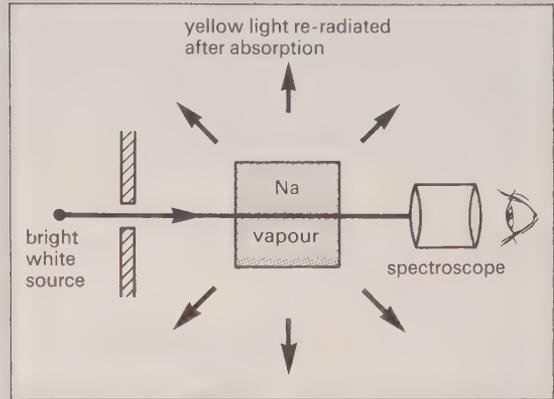


Fig. 30.6. Observation of Fraunhofer lines.

Net power loss or gain

For a black body at temperature T in an environment at T_0

$$\begin{aligned} \Phi_{\text{net}} &= A\sigma T^4 - A\sigma T_0^4 \\ &\quad \text{(loss)} \quad \text{(gain)} \\ &= A\sigma(T^4 - T_0^4) \end{aligned}$$

For a non-black body, net power loss is

$$\epsilon A\sigma T^4 - \alpha A\sigma T_0^4 = \epsilon A\sigma(T^4 - T_0^4)$$

where we have assumed $\epsilon = \alpha$.

Special cases

(a) If $T \gg T_0$, then T_0^4 is negligible, so for a black body the loss P_{net} is given by

$$P_{\text{net}} \approx A\sigma T^4$$

(b) If $(T - T_0)$ is small, for a black body

$$\begin{aligned} \Phi_{\text{net}} &\approx A\sigma(T^2 + T_0^2)(T^2 - T_0^2) \\ &\approx A\sigma(2T_0^2)(2T_0)(T - T_0) \\ &\approx (4A\sigma T_0^3)(T - T_0) \end{aligned}$$

The rate of loss of energy by radiation is approximately $\propto (T - T_0)$ when $T \approx T_0$. One sometimes assumes this result when making a simple cooling correction (p. 190).

► 30.8 ENERGY IN THE BLACK-BODY SPECTRUM

The problem is to measure the way in which the energy radiated from a black body is distributed between the different wavelengths, and to explain the experimental results.

► 30.7 STEFAN'S LAW (1879)

This law is also called the *Stefan-Boltzmann Law*; it was derived by *Boltzmann* from thermodynamic reasoning (1884).

The law states that the power radiated over all wavelengths from a black body is proportional to the fourth power of the thermodynamic temperature T .

Thus total radiant exitance $M_B \propto T^4$,

$$M_B = \sigma T^4$$

but $M = \Phi/A$ so

$$\Phi = A\sigma T^4$$

$$\Phi[\text{W}] = A[\text{m}^2]\sigma\left[\frac{\text{W}}{\text{m}^2\text{K}^4}\right]T^4[\text{K}^4]$$

σ is called the **Stefan-Boltzmann constant**, and has the value $5.67 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$.

$$[\sigma] = [\text{MT}^{-3}\Theta^{-4}]$$

For a non-black body

$$\Phi = \epsilon A\sigma T^4$$

where ϵ is the total emissivity.

The measurements were taken by *Lummer and Pringsheim* (1893), but can be explained only by using the quantum theory (*Planck* 1900).

Distinguish carefully between the situations represented by fig. 30.7(a) and (b).

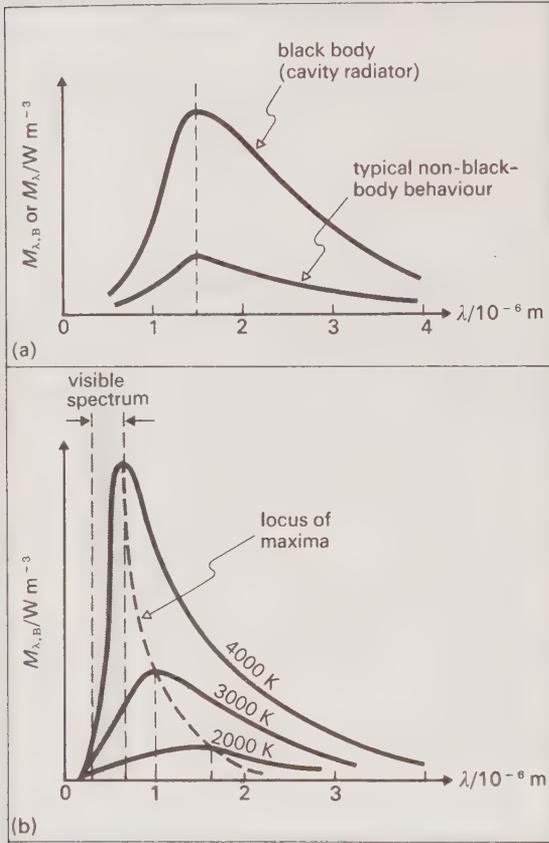


Fig. 30.7. Distribution of radiant energy for various spectra: (a) variation between bodies at a given temperature (about 2000 K), and (b) variation between different temperatures for a given (black) body. $M_{\lambda, B}$ is proportional to the intensity of black-body radiation at wavelength λ .

Notes on fig. 30.7(b)

(a) At each temperature T there is a wavelength λ_{max} for which the intensity is a maximum.

These are related by **Wien's displacement law**

$$\lambda_{max} T = \text{constant}$$

The law was based on thermodynamic reasoning. The constant, found by experiment, is $2.9 \times 10^{-3} m K$ for a black-body radiator.

The law illustrates why the colour of the visible light radiated by a very hot body changes from red to yellow as the temperature increases.

(b) The total area under each curve

$$M_B = \int_0^{\infty} M_{\lambda, B} d\lambda$$

This represents the total radiant exitance for a black body at that particular temperature.

Stefan's Law states $M_B = \sigma T^4$. We find experimentally that (the area under the curve for each temperature) $\propto T^4$, which agrees with the law.

Calculation of the Sun's temperature

For the Sun $\lambda_{max} \sim 5 \times 10^{-7} m$

so using $\lambda_{max} T_S = 2.9 \times 10^{-3} m K$

we deduce that the *black-body* temperature of the Sun is about 6000 K. The Sun's temperature varies, being $> 10^7 K$ at the centre. 6000 K is the temperature of the radiating surface.

30.9 RADIATION PYROMETERS

These are generally used to measure $T > 1250 K$.

(A) Total radiation pyrometer (Féry)

This responds to both visible and invisible radiation and is shown in fig. 30.8.

Notes

(a) It can be calibrated by using gold, freezing at 1338 K.

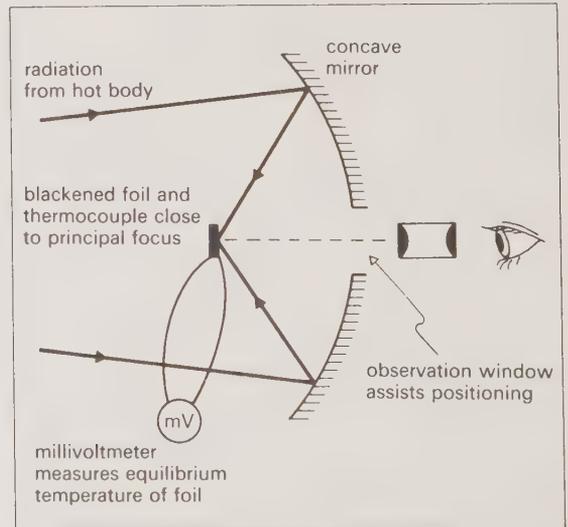


Fig. 30.8. A total radiation pyrometer.

(b) The recording of temperature can be made automatic.

(c) Black-body temperatures can be calculated by using *Stefan's Law*. This will give reliable temperatures when the radiation is from a hole in the side of a furnace.

The current I is adjusted until the image of the filament cannot be distinguished.

The instrument is simpler and more accurate than the total radiation instrument. It is used to establish the international temperature scale from the gold point to about 2300 K.

(B) Optical pyrometer (disappearing filament)

This responds to visible radiation only and is shown in fig. 30.9.

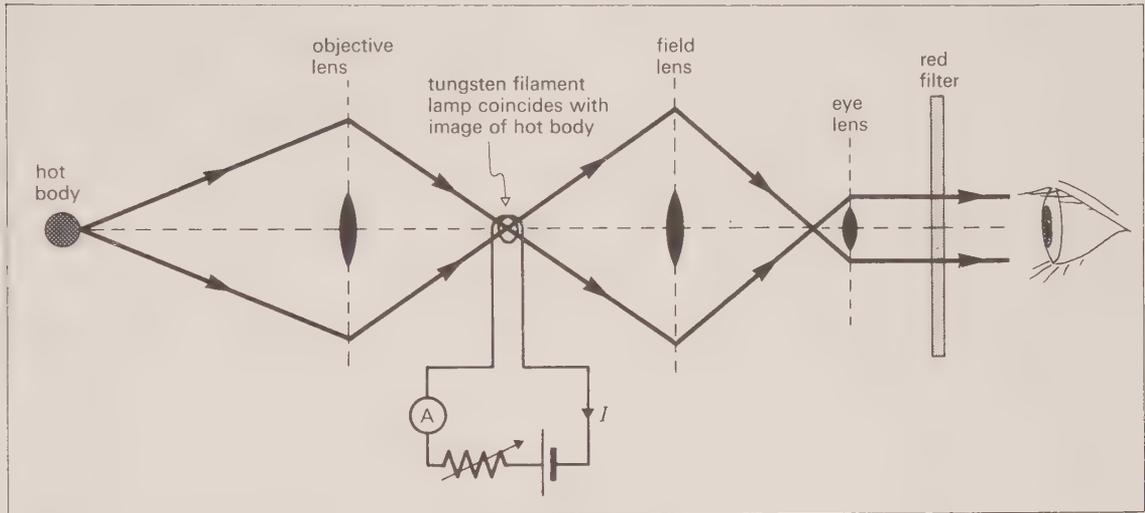


Fig. 30.9. The disappearing filament pyrometer.

VI GEOMETRICAL OPTICS

31	PRINCIPLES OF GEOMETRICAL OPTICS	252
32	THE PRISM AND THIN LENS	257
33	SPHERICAL MIRRORS	266
34	LENS AND MIRROR ABERRATIONS	267
35	DISPERSION AND THE SPECTROMETER	269
36	OPTICAL INSTRUMENTS	275

31 PRINCIPLES OF GEOMETRICAL OPTICS

31.1	INTRODUCTION AND THE RAY CONCEPT	252
31.2	THE LAWS OF REFLECTION AND REFRACTION	252
31.3	REFRACTIVE INDEX AND SNELL'S LAW	253
31.4	APPROXIMATIONS IN OPTICS	253

31.5	REAL AND APPARENT DEPTH	254
31.6	TOTAL INTERNAL REFLECTION AND CRITICAL ANGLE	254
31.7	METHODS OF MEASURING REFRACTIVE INDEX	255
31.8	THE PRINCIPLES OF IMAGE FORMATION	256

31.1 INTRODUCTION AND THE RAY CONCEPT

By *light* we mean the part of the electromagnetic spectrum that has the correct frequency range to stimulate the nerve endings of the human retina. To study the *production* and *absorption* of these waves one must use the *quantum theory*. To study the interaction of the waves with each other one must use *wave theory*, and this is done in part VII.

The wavelength of all visible light is about 10^{-7} m.

When we consider objects and apertures of

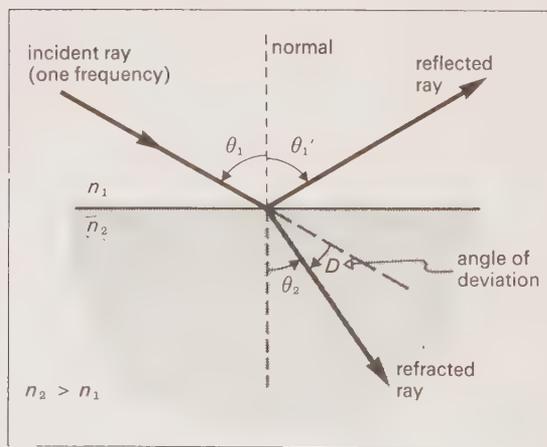


Fig. 31.1. Reflection and refraction at a plane interface.

dimensions greater than about 10^{-5} m, diffraction effects are often considered as relatively unimportant. This will enable us to treat the behaviour of light using *rays* rather than waves.

A *ray* is the path taken by light energy in travelling from a source to a receiver: in an isotropic medium the path is a straight line.

31.2 THE LAWS OF REFLECTION AND REFRACTION

Refer to fig. 31.1.

I The incident ray and the normal define a plane which contains both the reflected ray and the refracted ray.

II For reflection:
The angle of incidence equals the angle of reflection.

$$\theta_1 = \theta_1'$$

III For refraction:
For a wave of given frequency, and a given pair of media

$$\frac{\sin \theta_1}{\sin \theta_2} = \text{a constant}$$

Law III is **Snell's Law**. Wave theory shows that the constant is the (relative) **refractive index** for light passing from medium 1 to medium 2 (p. 101).

Note that $\theta_1 = \theta'_1$ for both *regular* (specular) and *diffuse* reflection, but that when *diffuse* reflection occurs the normals at many different points on the surface are not parallel.

31.3 REFRACTIVE INDEX AND SNELL'S LAW

Absolute refractive index n of a medium for a given frequency is defined by

$$n = \frac{\text{speed of light in a vacuum}}{\text{speed of light in the medium}}$$

$$n = \frac{c}{c_n}$$

n varies with the frequency of the light wave.

When a ray is refracted at the interface between two media, the *relative* refractive index determines the deviation. We can avoid the use of relative refractive index as follows (fig. 31.2):

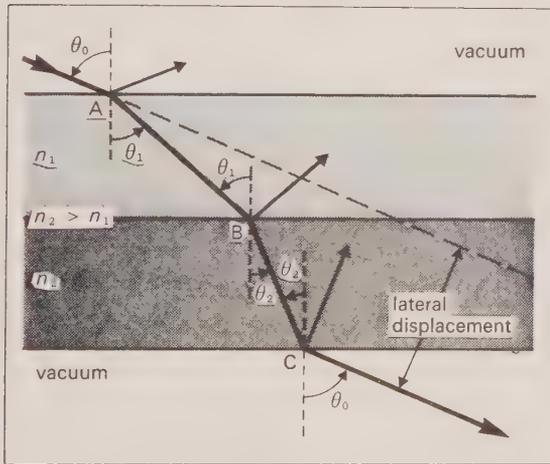


Fig. 31.2. The symmetrical form of Snell's Law.

At A
$$n_1 = \frac{c}{c_1} = \frac{\sin \theta_0}{\sin \theta_1}$$

at C
$$n_2 = \frac{c}{c_2} = \frac{\sin \theta_0}{\sin \theta_2}$$

Dividing,
$$\frac{n_2}{n_1} = \frac{c_1}{c_2} = \frac{\sin \theta_1}{\sin \theta_2} = {}_1n_2 \text{ (refraction at B)}$$

The relative refractive index ${}_1n_2 = n_2/n_1$, and *Snell's Law* can be written in the symmetrical form

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

where θ_1 and θ_2 are the angles between the ray and the normal in the media n_1 and n_2 .

By definition the (absolute) refractive index of a vacuum is 1 exactly. That of air at 101325 Pa pressure and 293 K is 1.0003, which is usually taken as 1.

31.4 APPROXIMATIONS IN OPTICS

Many equations used in geometrical optics are derived on the assumption that

$$\sin \theta \approx \tan \theta \approx \theta [\text{rad}]$$

where θ is measured in *radians*. Fig. 31.3 shows some examples.

This approximation is valid, for most purposes, provided θ is small. When $\theta = 10^\circ \approx 0.17$ rad, then the error is about 1%.

In the situation of fig. 31.3(iii) we might have a lens of 50 mm diameter (h) placed 200 mm (d) from a light-box. Then $\frac{1}{2}\theta \approx 0.12$ rad $\approx 7^\circ$, and the difference between the exact and approximate value for θ is likely to be less than the experimental error.

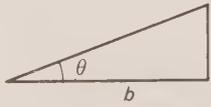
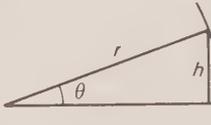
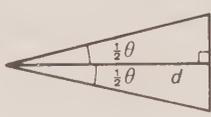
SITUATION	EXACT STATEMENT	BUT WE CAN SAY
(i) 	$\tan \theta = \frac{a}{b}$	$\theta = \frac{a}{b}$
(ii) 	$\sin \theta = \frac{h}{r}$	$h = r\theta$
(iii) 	$\tan \frac{\theta}{2} = \frac{h/2}{d}$	$h = d\theta$

Fig. 31.3. Approximations.

Diagrams are usually drawn with an exaggerated vertical scale in order to show clearly the related lengths. For the same reason we usually exaggerate small angles.

31.5 REAL AND APPARENT DEPTH

In fig. 31.4 the image position I is displaced relative to the object position O. We will find the position I for paraxial rays (rays for which θ_a and θ_n are small).

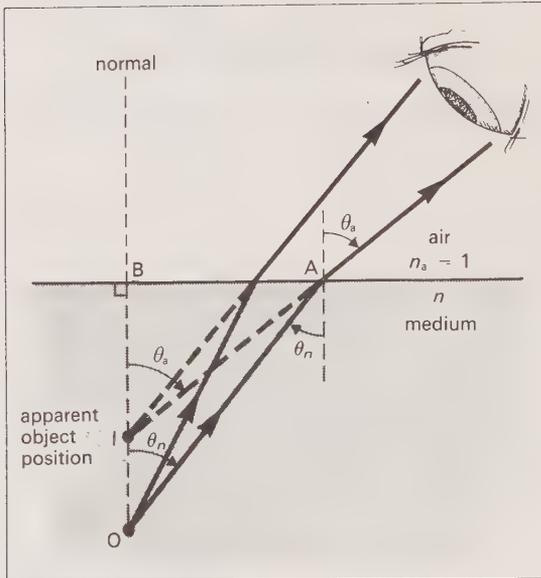


Fig. 31.4. Calculation of apparent depth.

For refraction at A

$$\begin{aligned}
 n_a \sin \theta_a &= n \sin \theta_n \\
 n &= \frac{\sin \theta_a}{\sin \theta_n} \quad (n_a = 1) \\
 &\approx \frac{\tan \theta_a}{\tan \theta_n} \\
 &= \frac{AB/IB}{AB/OB} \\
 &= \frac{OB}{IB} \\
 IB &= \frac{OB}{n} = \text{constant}
 \end{aligned}$$

so that the position of I is the same for all rays making a small angle with the normal. This shows

that I is the *image* of O. (For the concept of an image, see p. 256.)

The result is usually stated by

$$n = \frac{\text{real depth of object}}{\text{apparent depth of object}}$$

Notes

- (a) If the eye position is changed so that θ_a becomes larger, the point I describes a *caustic curve*, whose *cusps* is at the limiting position.
- (b) The distance OI is called the *displacement*.
- (c) When an object is viewed through several media, the *net displacement* is the sum of the displacements that each medium would have produced in the absence of the others.

31.6 TOTAL INTERNAL REFLECTION AND CRITICAL ANGLE

When light travels from a medium 1 to an optically denser medium 2, the maximum angle of refraction occurs when $\theta_1 = 90^\circ$. Then

$$\begin{aligned}
 n_1 \sin \theta_1 &= n_2 \sin \theta_2 \\
 n_1 \sin 90^\circ &= n_2 \sin \theta_{\text{max}}
 \end{aligned}$$

This maximum angle is called the **critical angle**, *c*, given by

$$\sin c = n_1/n_2 \quad (n_1 < n_2)$$

The exact value of *c* is controlled by the frequency of the wave, and hence by the colour of the light.

Suppose (fig. 31.5) light travels from medium 2 toward the interface with medium 1.

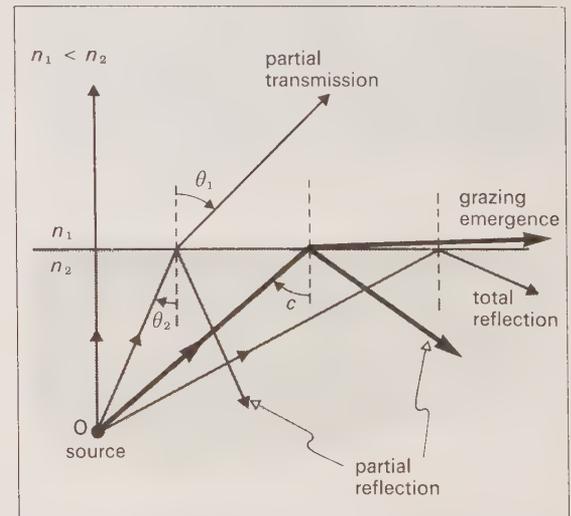


Fig. 31.5. Total internal reflection and the critical angle.

When $\theta_2 \leq c$, we see *partial* reflection and *partial* transmission (refraction).

When $\theta_2 > c$, refraction *cannot* occur, and the ray is *totally* reflected back inside medium 2.

Note the *conditions* for total internal reflection

- (1) $n_2 > n_1$,
- (2) $\theta_2 > c$, where $\sin c = n_1/n_2$.

Total internal reflection is a phenomenon associated with all types of wave motion. In light it finds application in the light pipe, in prisms for optical instruments, and in the measurement of critical angle.

Fibre optics

Light falling upon one end of a solid glass rod will be totally internally reflected at the cylindrical surface of the rod wherever the angle of incidence is greater than the critical angle. Leakage through the sides is reduced by polishing the rod and surrounding it with a glass 'cladding' of lower refractive index. The light emerges at the far end of the rod where the incident angle is less than the critical angle. A flexible **light pipe** (or **light guide**) consists of large numbers ($\sim 10^3$) of very thin, long

glass fibres bound together at random. If an *image*, rather than merely *light*, is to be transmitted, fibres of smaller diameter are required and they must be *coherent*, i.e. maintain the same position in the bundle relative to each other. The **fibre-optic endoscope** is a medical instrument with a long flexible shaft that allows the transmission of images from inaccessible areas such as joints, the stomach, intestines and heart. In telecommunications, optical fibres are now used for the transmission of modulated laser beams carrying telephone, radio, television and computer signals.

31.7 METHODS OF MEASURING REFRACTIVE INDEX

For solids

- (a) Measure θ_1 and θ_2 , and use

$$n_1 \sin \theta_1 = n_2 \sin \theta_2$$

- (b) Measure real and apparent depth using a travelling microscope.

- (c) Use a prism of the substance on the table of a spectrometer (p. 271). This is a very accurate method.

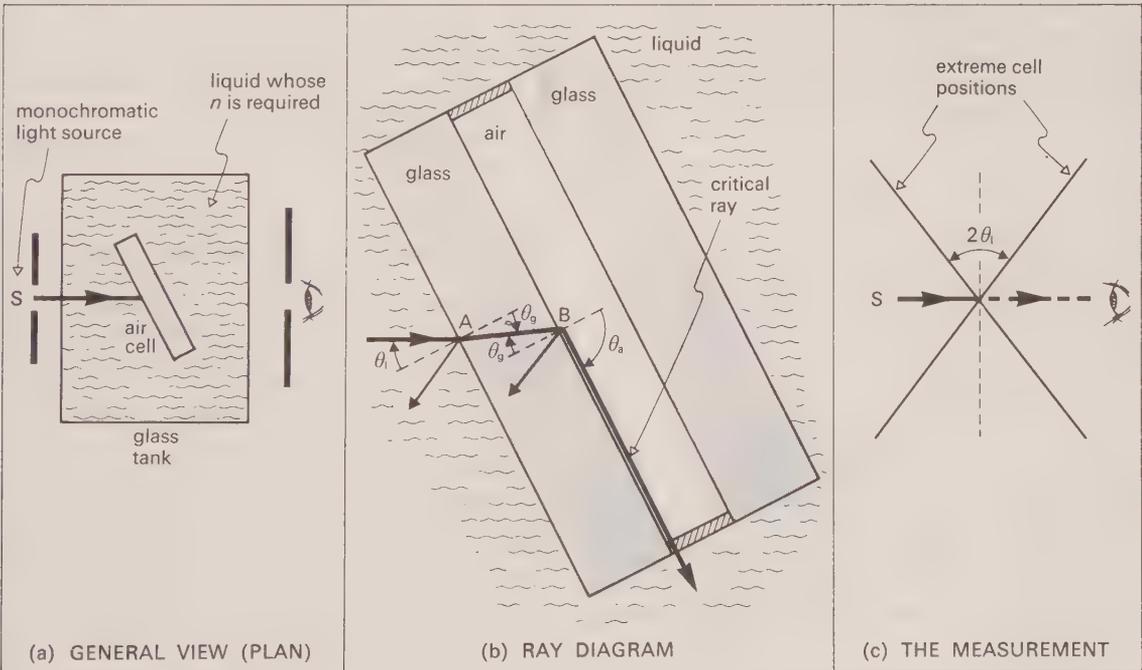


Fig. 31.6. The air-cell method for n_1 .

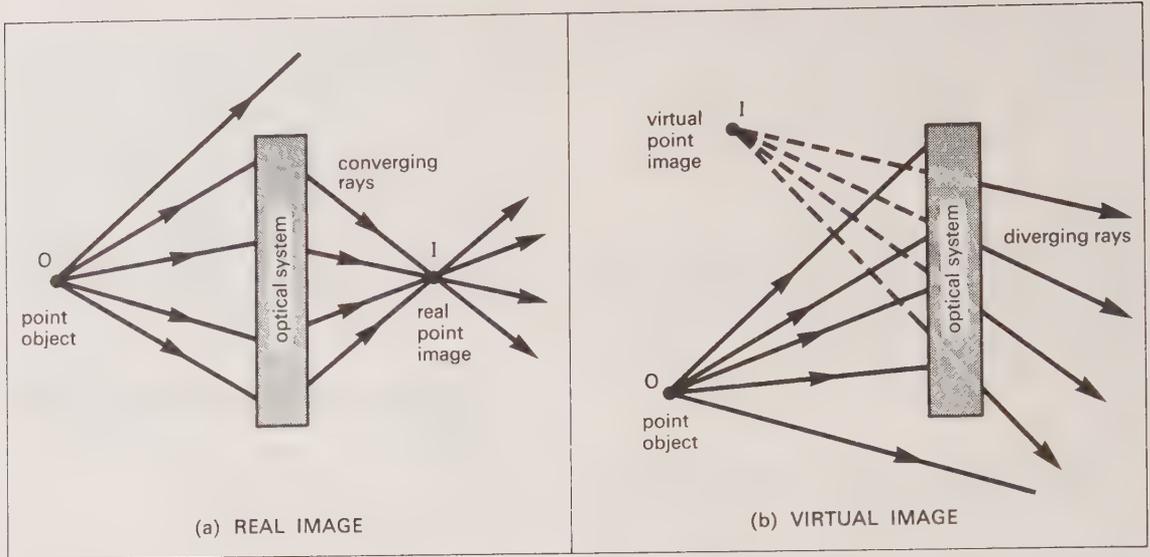


Fig. 31.7. Formation of images.

For liquids

- (a) Measure the critical angle, e.g. by
 - (i) the air-cell method (see below), or
 - (ii) the **Pulfrich refractometer**, useful when very little liquid is available.
- (b) As method (b) for solids.
- (c) As method (c) for solids: the liquid is confined to a hollow prism whose sides are made accurately parallel.

The air-cell method

The cell of fig. 31.6 (p. 255) is rotated about a vertical axis until the field of view is half bright and half dark. For refraction at A and B

$$n_1 \sin \theta_1 = n_g \sin \theta_g = n_a \sin \theta_a = 1$$

so
$$n_1 = \frac{1}{\sin \theta_1}$$

Notes

- (a) The glass surfaces are parallel and a knowledge of n_g is not needed.
- (b) $2\theta_1$ is measured, in practice, by a second rotation to cut-off in the opposite sense. This simplifies the procedure, and reduces the percentage error in θ_1 .

31.8 THE PRINCIPLES OF IMAGE FORMATION

Refer to fig. 31.7.

A **point image I** of the point object O is formed if *all* the rays intercepted by the system either

- (a) pass through a single point I, after emergence—this is called a **real image**—or
- (b) *seem* to have come from a single point I—this is called a **virtual image**.

Both types of image can be seen and photographed.

A *real image* can be photographed by simply placing a photographic emulsion at the site of the image.

Virtual images can only be photographed with the help of a converging optical system which serves to collect the divergent light energy, and concentrate it upon a single point (i.e. to form a real image by using the original virtual image as a virtual object (p. 261)).

The image of an *extended object* is built up from the point images of all the point objects which constitute the object. If the image is to be undistorted the point images must maintain the same relative spacings as their respective point objects.

An **aberration** is the phenomenon of an object not giving rise to an image as defined above. (See p. 267.)

32 THE PRISM AND THIN LENS

32.1 REFRACTION THROUGH A PRISM	257
32.2 MINIMUM DEVIATION	257
32.3 SMALL-ANGLED PRISMS	258
32.4 LENSES AND THEIR CONSTRUCTION	259
32.5 A LENS FORMULA	259
32.6 SIGN CONVENTION	260
32.7 THE LENS-MAKER'S FORMULA	260
32.8 THE VIRTUAL OBJECT	261
32.9 MAGNIFICATION	262

32.10 THIN LENSES IN CONTACT	262
32.11 THE POWER OF A LENS	262
32.12 NEWTON'S FORMULA	263
32.13 OBJECT-IMAGE DISTANCES	263
32.14 MEASUREMENT OF LENS RADIUS OF CURVATURE	263
32.15 THE CAMERA	264
32.16 THE TELEPHOTO LENS	265

32.1 REFRACTION THROUGH A PRISM

These relationships follow from the geometry of fig. 32.1:

$$\text{Deviation } D = (\theta_a - \theta_g) + (\theta'_a - \theta'_g)$$

$$A = (\theta_g + \theta'_g)$$

According to the **principle of reversibility**, the light energy can travel in *either direction* along the path shown.

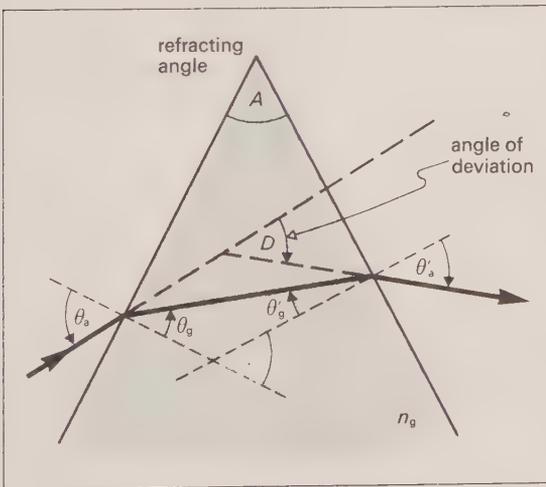


Fig. 32.1. The path of a ray through a prism.

32.2 MINIMUM DEVIATION

Suppose D is measured for various angles of incidence θ_a , as shown in fig. 32.2.

It can be proved theoretically, and demonstrated in practice, that the deviation is a *minimum* when the ray passes *symmetrically* through the prism. Then

$$\theta_a = \theta'_a \quad \text{and} \quad \theta_g = \theta'_g$$

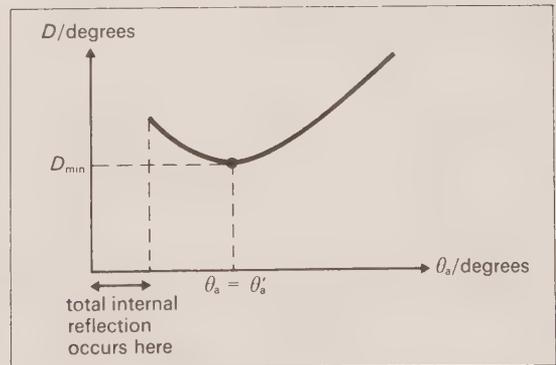


Fig. 32.2. Deviation through any prism.

From the previous section

$$\theta_a = \frac{A + D_{\min}}{2} \quad \text{and} \quad \theta_g = \frac{A}{2}$$

But $n_a \sin \theta_a = n_g \sin \theta_g$

so $n_g = \frac{\sin \theta_a}{\sin \theta_g} \quad (n_a = 1)$

$$n_g = \frac{\sin \frac{A + D_{\min}}{2}}{\sin \frac{A}{2}}$$

Note carefully that this equation is valid for a prism of *any* angle A less than twice the critical angle, but *only when the deviation is a minimum*.

The result is also proved, using wave ideas, on p. 108.

32.3 SMALL-ANGLED PRISMS

The result proved here is fundamental for the method used in this book to develop lens theory.

Suppose that (a) $\theta_a < 0.1$ rad, and
 (b) $A < 0.1$ rad (fig. 32.3).

Then it follows that each of θ_g , θ'_g and θ'_a is less

than 0.1 rad, and is therefore approximately equal to its sine.

So $n_g = \frac{\sin \theta_a}{\sin \theta_g} \approx \frac{\theta_a}{\theta_g}$

and $n_g \approx \frac{\theta'_a}{\theta'_g}$

Then $D = (\theta_a - \theta_g) + (\theta'_a - \theta'_g)$
 $= n\theta_g + n\theta'_g - (\theta_g + \theta'_g)$
 $= (n - 1)A \quad \text{using } A = \theta_g + \theta'_g$

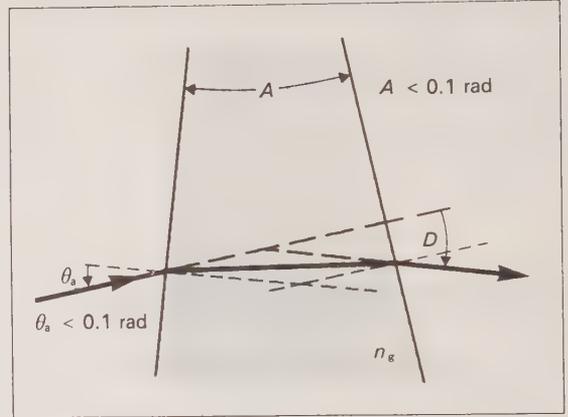


Fig. 32.3. Deviation by a small-angled prism.

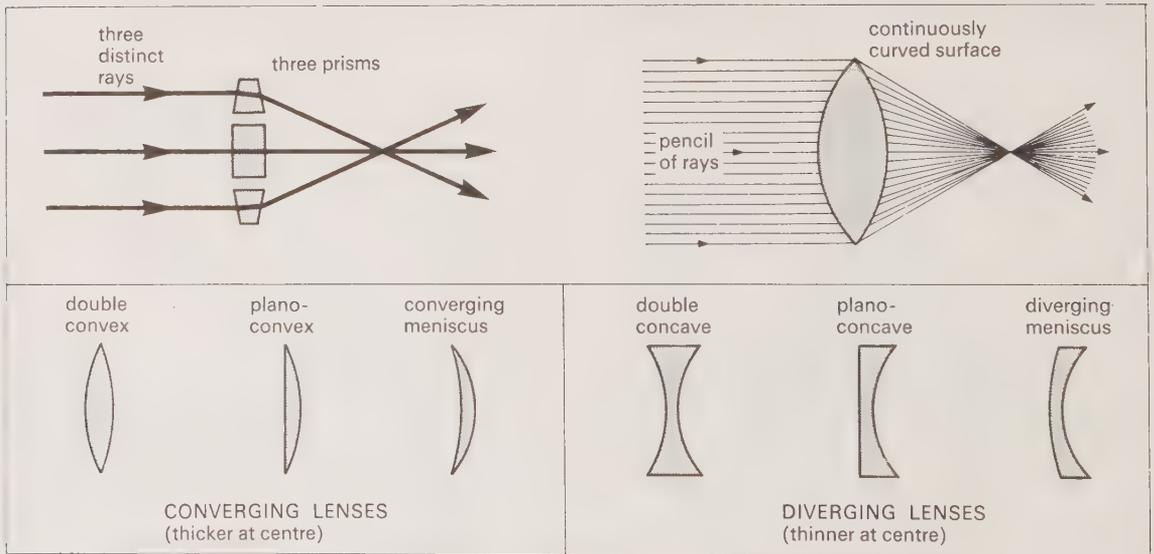


Fig. 32.4. Construction of a converging lens, and types of lenses.

$$D = (n - 1)A$$

Note carefully that this equation is valid

(a) for angles $A < 0.1$ rad (approx.), and then only

(b) for all values of $\theta_a < 0.1$ rad.

It shows that, for a prism of given angle A , all rays making small angles of incidence undergo the same deviation.

32.4 LENSES AND THEIR CONSTRUCTION

Fig. 32.4 illustrates

(a) that a series of correctly-angled prisms can be built into a system that causes several *distinct* rays to cross at a point, and

(b) that a lens of continuously curved surfaces can be made to focus a *pencil* of rays at a point.

In the next section we prove that an optical system made of a medium bounded by two *spherical* surfaces will cause such a pencil of rays to cross the axis at approximately the same point. This means that a spherical lens has the ability to form *images* (see p. 256).

Fig. 32.5 illustrates terms used for lenses. f is the image distance for an infinite object distance.

The **principal axis** joins the centres of curvature of the two spherical surfaces.

A ray passing through the **optic centre** suffers

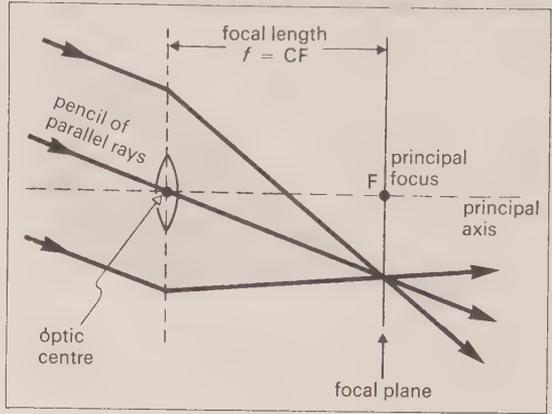


Fig. 32.5. Lens terms.

zero deviation, and, for a thin lens, negligible lateral displacement.

32.5 A LENS FORMULA

Refer to figs. 32.6 (below) and 32.7 (overleaf). The aim is to find a numerical relationship between OC, IC, CC_1 and CC_2 for this situation, in which the rays are paraxial.

We assume

(a) that the lens can be considered as a system of small-angled prisms, and

(b) that angles of incidence are small.

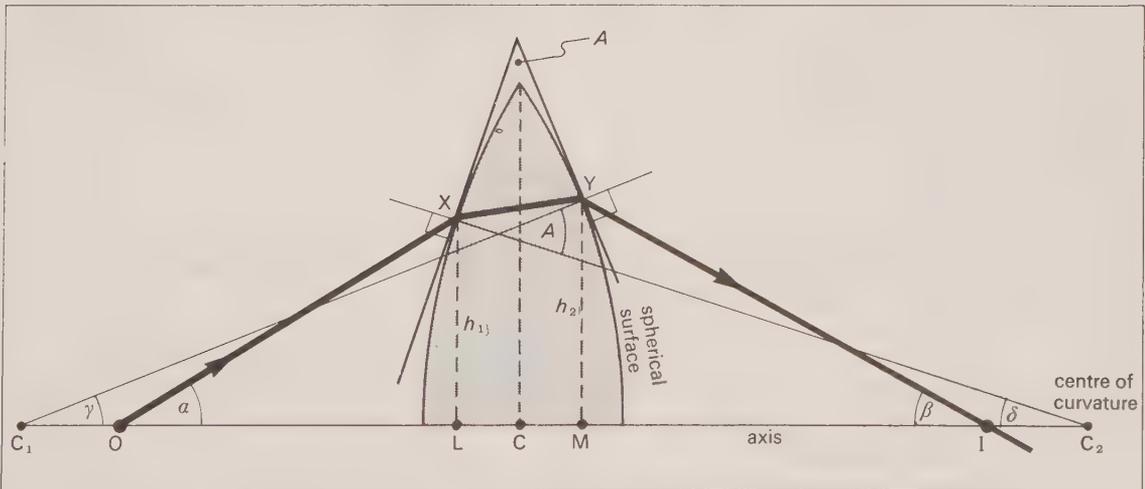


Fig. 32.6. Details of a ray passing through the top half of a thin lens (all features are considerably distorted for clarity).

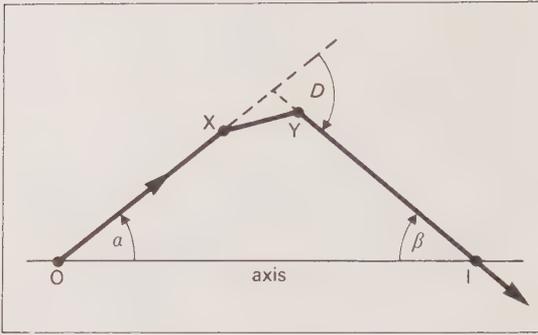


Fig. 32.7. To show $\alpha + \beta = D$. (All angles exaggerated.)

In fig. 32.7 $D = \alpha + \beta$
 and $A = \gamma + \delta$
 Using $D = (n - 1)A$
 for the small-angled prism,

$$\alpha + \beta = (n - 1)(\gamma + \delta)$$

Because α, β, γ and δ are small, we replace them by their tangents. So

$$\frac{h}{OC} + \frac{h}{IC} = (n - 1)\left(\frac{h}{CC_1} + \frac{h}{CC_2}\right)$$

in which we have put $h_1 = h_2 = h$, because the lens is to be considered thin, so that L, M and C approximately coincide.

Dividing by h

$$\frac{1}{OC} + \frac{1}{IC} = (n - 1)\left(\frac{1}{CC_1} + \frac{1}{CC_2}\right)$$

Notes

(a) For this particular lens n, CC_1 and CC_2 are constant. The equation shows that for a fixed value OC, IC is *not* dependent on the angle α . Therefore *all* rays from O that are intercepted by the lens pass through the same point I . Thus I is the *image* of O (see section 31.8).

(b) The equation is a *numerical* relationship between the lengths OC, IC, CC_1 and CC_2 for this particular type of lens. We will now adopt a *sign convention* for the distances involved; this will enable us to derive an *algebraic* relationship which will serve for all types of lens used in all possible ways.

32.6 SIGN CONVENTION

We adopt the following **symbols**:

- u = object distance, f = focal length,
- v = image distance, R = radius of curvature.

A symbol represents three things: a sign, a number and a unit. These symbols should not be given any sign until they are replaced by a numerical value. Several sign conventions are used. In this book we adopt the **Real is Positive** convention.

The convention

- (a) All distances are measured to the optic centre.
- (b) Distances are taken to be *positive* if actually traversed by the light ray (distances to *real* objects and images).
- (c) Distances are taken to be *negative* if only apparently traversed by the light ray (distances to *virtual* objects and images).
- (d) More generally the bracket $(n - 1)$ should read

$$\left(\frac{n_{\text{lens}}}{n_{\text{medium}}} - 1\right)$$

It will be taken to be positive, so that it represents the magnitude but not the sense of the deviation.

(e) The sense of the deviation will be shown by the sign allotted to R (fig. 32.8). This means that R is positive for surfaces which cause a deviation towards the axis. Some examples are given in fig. 32.9.

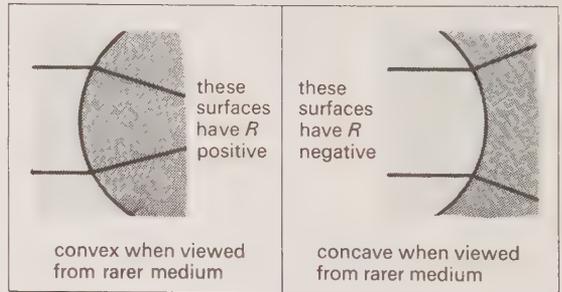


Fig. 32.8. Sign convention for R .

32.7 THE LENS-MAKER'S FORMULA

Using the sign convention we can write down an equation which can be used for any lens.

Refer to fig. 32.10. In radian measure

$$D = \alpha + \beta$$

$$\approx \frac{h}{OC} + \frac{h}{IC}$$

and because D is the same for the same value of h

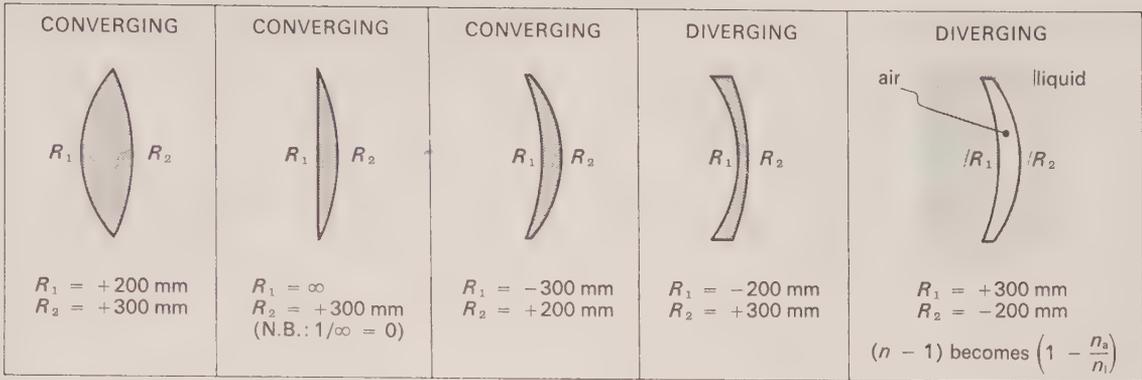


Fig. 32.9. Examples of the sign convention for R .

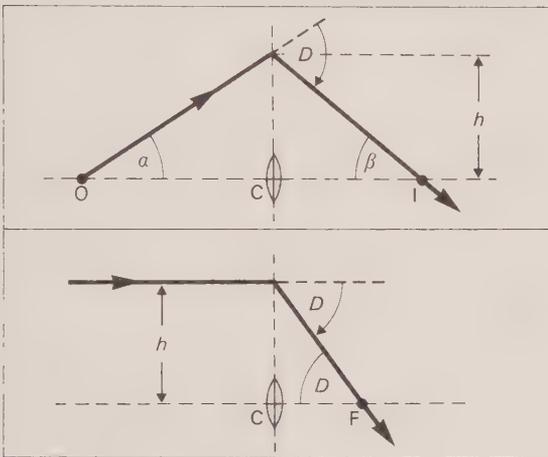


Fig. 32.10. To relate u , v and f . (Angles exaggerated.)

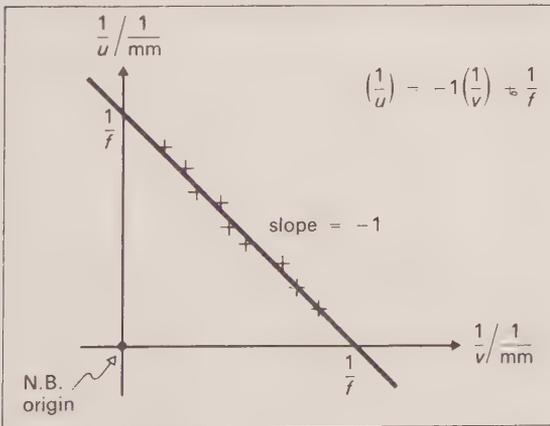


Fig. 32.11. A graph of $1/u$ plotted against $1/v$ for a converging lens.

$$D \approx \frac{h}{CF}$$

so

$$\frac{1}{OC} + \frac{1}{IC} = \frac{1}{CF}$$

We combine this equation and sections 32.5 and 32.6 to give

$$\frac{1}{u} + \frac{1}{v} = (n - 1) \left(\frac{1}{R_1} + \frac{1}{R_2} \right) = \frac{1}{f}$$

which is commonly called the **lens-maker's formula**. f will be positive for a converging lens, and negative for a diverging lens. Its value is often deduced from the graph of fig. 32.11.

32.8 THE VIRTUAL OBJECT

When rays converging to a point are intercepted by an optical system, that point is the **virtual object** for any subsequent image (fig. 32.12). An example is discussed in section 32.10.

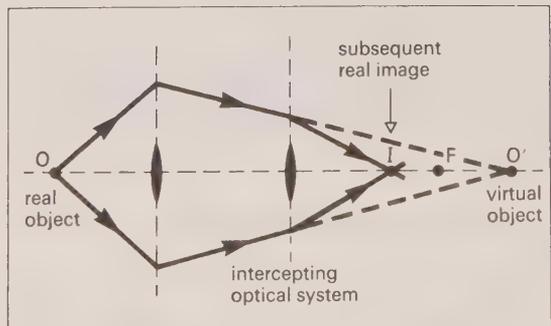


Fig. 32.12. Example of a virtual object.

32.9 MAGNIFICATION

The **transverse magnification** m is defined by

$$m = \frac{\text{height of transverse image}}{\text{height of transverse object}}$$

$$m = \frac{IM}{OB}$$

In magnitude $m = \frac{v}{u} = \frac{v}{f} - 1$

If m is negative, this means that either O or I is virtual. The negative sign has no other physical significance.

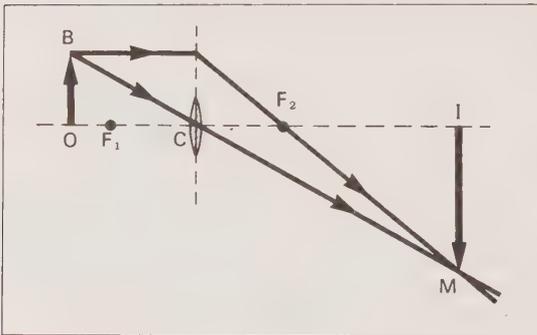


Fig. 32.13. To illustrate transverse magnification.

► The **longitudinal magnification** is defined as

$$\frac{\text{length of image}}{\text{length of object}}$$

where each lies parallel to the axis.

Its value cannot be calculated simply in terms of u and v , which are no longer uniquely defined.

32.10 THIN LENSES IN CONTACT

In fig. 32.14 a , b and c represent lengths. For lens 1,

$$\frac{1}{+a} + \frac{1}{+b} = \frac{1}{f_1} \tag{1}$$

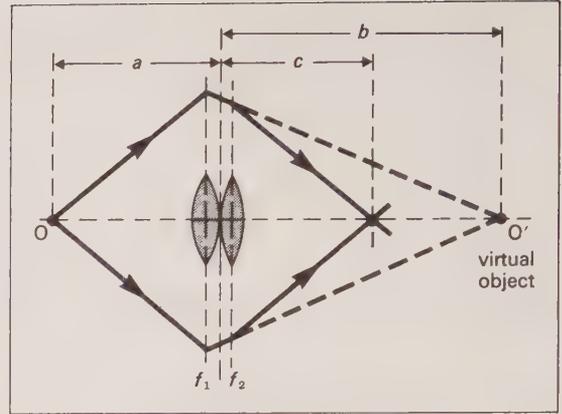


Fig. 32.14. Two thin lenses in contact.

For lens 2,

$$\frac{1}{-b} + \frac{1}{+c} = \frac{1}{f_2} \tag{2}$$

For the composite lens

$$\frac{1}{+a} + \frac{1}{+c} = \frac{1}{f}$$

Adding (1) and (2), we see

$$\frac{1}{f_1} + \frac{1}{f_2} = \frac{1}{f}$$

where f would be the focal length of the single lens equivalent to the other two *in contact*.

32.11 THE POWER OF A LENS

The **power** F of a lens is defined by the equation

$$F = \frac{1}{f}$$

$$F \left[\frac{\text{rad}}{\text{m}} \right] = \frac{1}{f[\text{m}]}$$

The equation $CF \approx h/D$ (from section 32.7) indicates why the rad m^{-1} is an appropriate unit for the power of a lens.

From the defining equation it follows that

(a) *diverging* lenses have a power which is *negative*, and

(b) for a *small* number of *thin* lenses in contact

$$F = F_1 + F_2 + F_3 + \dots$$

Note that f must be expressed in *metres*. If it is quoted in mm, then F can be calculated from

$$F \frac{\text{rad}}{\text{m}} = \frac{1000}{f/\text{mm}}$$

32.12 NEWTON'S FORMULA

This is an alternative way of representing the lens formula. In fig. 32.13 suppose

$$OF_1 = x \quad \text{so} \quad u = x + f$$

$$IF_2 = y \quad \text{so} \quad v = y + f$$

Then
$$\frac{1}{x + f} + \frac{1}{y + f} = \frac{1}{f}$$

which simplifies to

$xy = f^2$

Newton's formula

The result is helpful for finding the focal length of an inaccessible lens.

32.13 OBJECT-IMAGE DISTANCES

Experimentally it is found that if D is big enough (fig. 32.15) *two* lens positions give an image on the screen.

Rewriting the lens formula in terms of D and a ,

$$D = 2b + a \quad \text{so} \quad b = \frac{D - a}{2}$$

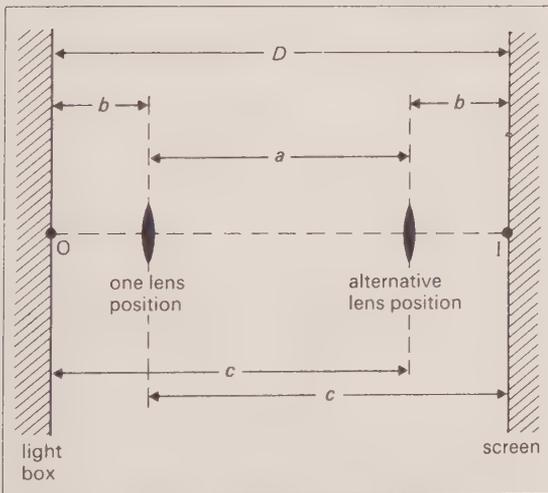


Fig. 32.15. Object-image distances.

$$D = b + c \quad \text{so} \quad c = \frac{D + a}{2}$$

Then
$$\frac{2}{D - a} + \frac{2}{D + a} = \frac{1}{f}$$

giving
$$D^2 - 4Df - a^2 = 0 \tag{1}$$

Solving the quadratic for D , we find

$$D = \frac{4f \pm \sqrt{(16f^2 + 4a^2)}}{2}$$

For a real image and object, D must be positive and greater than zero.

$$D_{\min} = \frac{4f + 4f}{2} = 4f$$

Thus the *minimum distance* between a real object and its image is $4f$. This is important in the design of the terrestrial telescope which uses a converging erecting lens.

($D = 0$ corresponds to a real object—virtual image.)

Equation (1) can be written as

$$f = \frac{D^2 - a^2}{4D}$$

This gives an accurate method of measuring f when the exact location of the *optic centre* is not accessible.

32.14 MEASUREMENT OF LENS RADIUS OF CURVATURE

Concave surface

Treat as a concave spherical mirror (p. 266).

Convex surface

Use the **Boys image** of fig. 32.16, overleaf.

When rays strike the second surface at right angles

- (a) there is a virtual image at C, and
- (b) reflection (followed by a further refraction) forms a real ghost image beside the object.

For the lens formula

$$u = +a$$

$$v = -b$$

If we know f , and measure a , we can find b .

For this lens $R = +b$ for the right-hand surface. If f , R_1 and R_2 are measured, then n can be calculated from the lens-maker's formula.

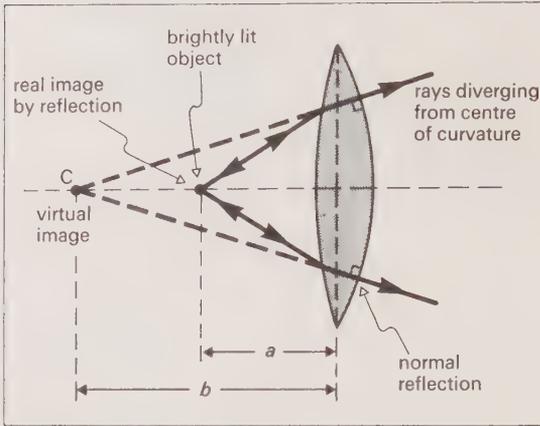


Fig. 32.16. To illustrate the Boys image.

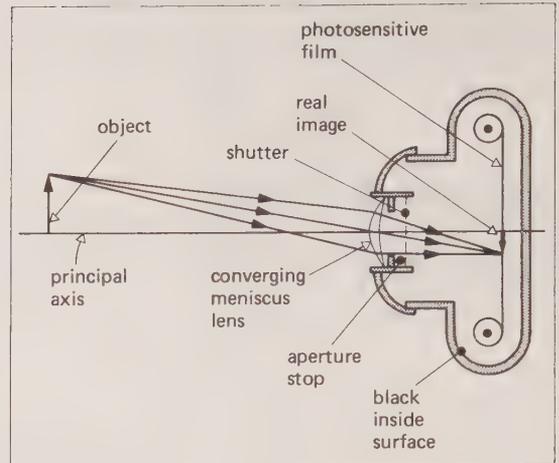


Fig. 32.17. A simple lens camera.

32.15 THE CAMERA

(A) The pinhole camera

Light from the objects being photographed passes through a very small, sharp-edged hole in an opaque diaphragm and forms an image at the back of the camera. As the hole is made smaller, the image becomes dimmer and sharper. Below a certain size, the image quality deteriorates due to diffraction effects (p. 308). Pinhole photographs need long exposure times but possess an overall softness of definition, no distortion and unlimited depth of field.

(B) The lens camera

In the simple camera of fig. 32.17, image aberrations are reduced by using a meniscus lens and an aperture stop. Better cameras have a system of lenses to minimize aberrations. The shutter controls the time (exposure time) for which the film is illuminated, and the aperture diameter determines the amount of light passing through the lens for a given shutter speed. *f*-number is defined by the following equation:

$$f\text{-number} = \frac{\text{focal length of lens}}{\text{effective diameter of aperture}}$$

The aperture control has a scale of *f*-number settings:

1.4 2 2.8 4 5.6 8 11 16 22 32

The values are selected so that changing between adjacent *f*-numbers alters the aperture area by a factor of 2 (adjacent *f*-numbers therefore differ by

a factor of $\sqrt{2}$). This means that the exposure time is also changed by a factor of 2. A large *f*-number indicates a small aperture, and therefore a large exposure time for the same degree of film illumination. The appropriate combination of exposure time and *f*-number is determined by the brightness (and motion) of the object and the type of film. The **depth of field** is the zone extending either side of the focussed distance where the blurring of the image is acceptably small. It is increased by

- (a) reducing the aperture (thereby increasing the *f*-number),
- (b) using a lens of shorter focal length,
- (c) increasing the distance from the object to be photographed.

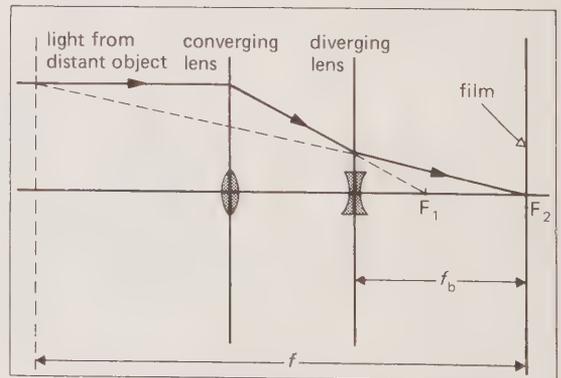


Fig. 32.18. The principle of the telephoto lens.

The separation between lens and film can be altered to cater for different object distances. The **depth of focus** is the small range of focussing movement between the film and the lens which maintains an acceptably sharp image.

32.16 THE TELEPHOTO LENS

As the size of the image of a distant object is directly proportional to the focal length of the lens

(or lens combination), a large image requires a long focal length. The camera can be kept reasonably short by using a telephoto lens which consists of a front converging lens system and a rear diverging lens system as shown in fig. 32.18. The distance f_b , the **back focal length**, from the diverging lens to the film is significantly less than f , the *effective* focal length of the combination of lenses. The camera can therefore be shorter than it would be if an equivalent single lens were used. The ratio f/f_b is known as the **telephoto ratio**.

33 SPHERICAL MIRRORS

33.1 THE MIRROR FORMULA

266

33.1 THE MIRROR FORMULA

This derivation follows the argument of sections 32.5 and 32.7 (the lens formula).

In outline,

$$\beta = \alpha + \theta$$

$$\gamma = \alpha + 2\theta$$

So

$$\alpha + \gamma = 2\beta$$

Thus when the rays are *paraxial* (when $AP \leq 1/10$ of the smallest of OP , CP and IP)

$$\frac{AP}{OP} + \frac{AP}{IP} \approx \frac{2AP}{CP}$$

which leads to the *numerical relationship*

$$\frac{1}{OP} + \frac{1}{IP} = \frac{2}{CP} \quad (1)$$

for this mirror used in this way.

We adopt a *sign convention* which includes (a), (b) and (c) of section 32.6. In addition we specify that

(d) surfaces which reflect on the *concave* side (fig. 33.1) have a *positive* radius of curvature, and

(e) surfaces that reflect on the *convex* side have a *negative* radius of curvature.

This now leads to the *algebraic* relationship

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{r}$$

in which the symbols are treated as before.

Finally we note that the *focal length* f is the image distance corresponding to infinite object distance, so

$$\frac{1}{f} = \frac{2}{r}$$

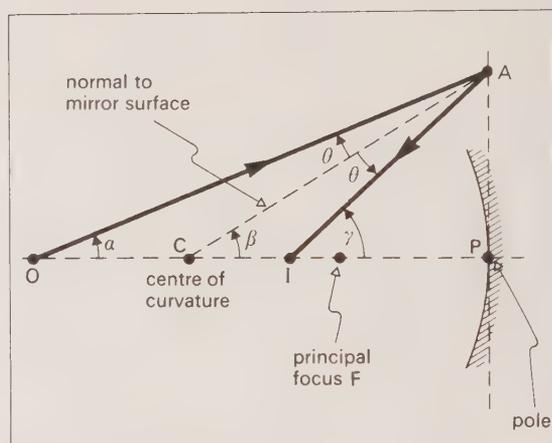


Fig. 33.1. Derivation of the mirror formula.

f is allotted the same sign as the corresponding r .
Then

$$\frac{1}{u} + \frac{1}{v} = \frac{2}{r} = \frac{1}{f}$$

is referred to as the **mirror equation**.

Notes

(a) Equation (1) above shows that the location of I is independent of α if α is small. Thus a spherical mirror will form a point *image* of a point object by reflection of paraxial rays.

(b) Note the following terms used in connection with spherical mirrors: *pole*, *principal axis*, *focal plane*.

(c) *Magnification* is defined in the same way as for lenses (p. 261).

34 LENS AND MIRROR ABERRATIONS

34.1 INTRODUCTION	267
34.2 SPHERICAL ABERRATION	267

34.3 CHROMATIC ABERRATION	268
34.4 PRACTICAL CONSIDERATIONS	268

34.1 INTRODUCTION

An **aberration** is the phenomenon of a point object not giving rise to a point image. It may cause the image of an extended object to differ from the object in colour, shape and sharpness of definition.

Because light is a wave motion, it exhibits **diffraction** effects which *cannot be eliminated*. The extent of the diffraction determines the blurring of what would otherwise be a point image from a point object. It is proportional to

$$\frac{\text{wavelength of light}}{\text{diameter of optical system}}$$

and so can be *reduced* (but never eliminated) by

- (a) increasing the diameter of an optical system such as a lens, and/or
- (b) using light of a smaller wavelength.

It follows that we do not necessarily improve the quality of an image by stopping down a lens (p. 310).

34.2 SPHERICAL ABERRATION

The mirror formula and lens-maker's formula are only approximately true. Perfectly *spherical* surfaces *cannot* produce point images of point objects simply because approximations as to the sizes of angles were made in deriving the formulae.

(A) Lenses

Fig. 34.1 shows a point object giving rise to a diffused circle as image. Note the terms **longitudinal spherical aberration** and **circle of least confusion**.

(a) The effect can be *corrected* (for a given object distance) only by grinding the lens surfaces to make them suitably **aspherical**.

(b) The effect can be *minimized* by arranging for the deviation produced by the lens to be divided

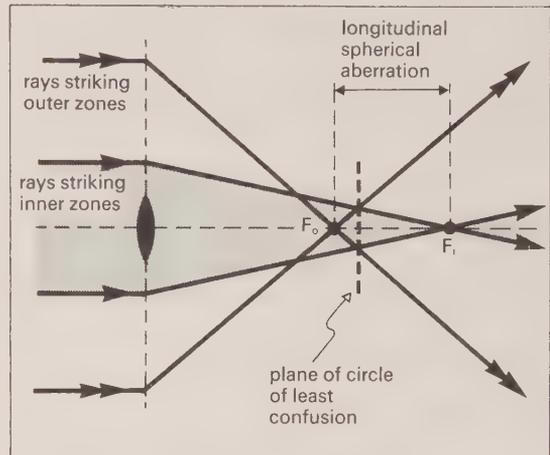


Fig. 34.1. Spherical aberration in a lens (exaggerated considerably).

equally between the surfaces. This is achieved in the plano-convex lens when the curved surface receives or emits those rays which are most nearly parallel to the axis. Thus in a microscope, the plane surface faces the incident light.

(c) The effect can be *reduced* for a given lens by stopping down the lens aperture, but this makes the image less bright and introduces greater diffraction effects. For large relative aperture (e.g. the camera), aberration is usually more serious than diffraction, whereas for small relative aperture (e.g. the telescope), the reverse is usually true.

Spherical aberration is corrected within the human eye by a decrease in refractive index away from the axis.

(B) Mirrors

(a) The effect can be *eliminated* for point objects *on the axis* by using a **paraboloidal** surface. Then all rays parallel to the axis are reflected through

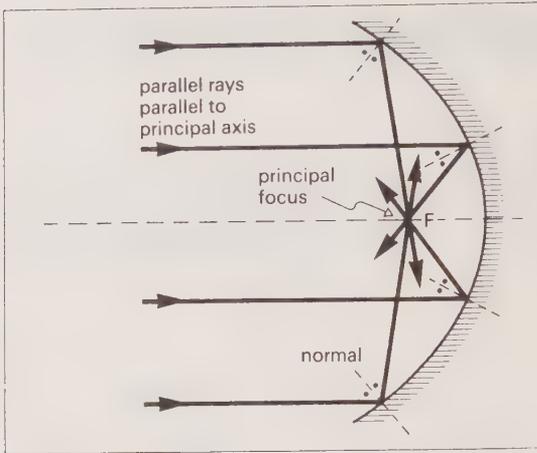


Fig. 34.2. Reflection without aberration from a paraboloidal surface.

the principal focus, whatever the mirror aperture (fig. 34.2).

(b) The effect can be reduced for a given mirror by using a smaller aperture.

34.3 CHROMATIC ABERRATION

The equation

$$\frac{1}{f} = (n - 1) \left(\frac{1}{R_1} + \frac{1}{R_2} \right)$$

shows that, for a given lens, f depends on n . n in turn is controlled by the frequency of the light.

$$n_{\text{violet}} > n_{\text{red}}$$

so

$$f_{\text{red}} > f_{\text{violet}}$$

Note the term **longitudinal chromatic aberration** in fig. 34.3.

(a) For two lenses of the same material we can make f independent of λ by putting their separation equal to the mean of their focal lengths. This system is satisfactory for viewing a virtual image.

(b) The effect can be eliminated for two colours (and improved for all) by using an **achromatic doublet**, fig. 34.4.

The crown and flint glass lenses have different powers, but produce equal and opposite dispersion (p. 270).

34.4 PRACTICAL CONSIDERATIONS

Large mirrors are constructed by silvering or aluminizing paraboloidal glass surfaces. If the re-

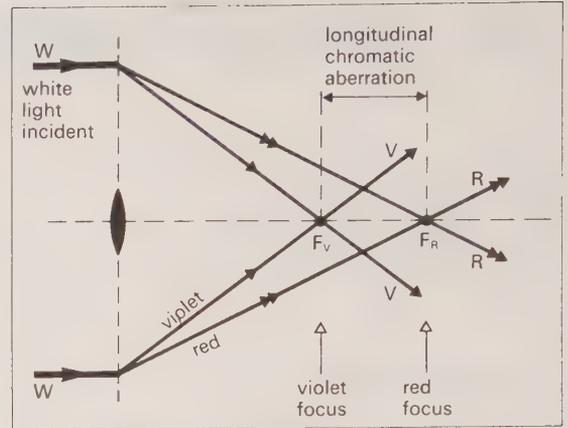


Fig. 34.3. A converging lens showing chromatic aberration (exaggerated considerably).

flecting surface is damaged or tarnished it can then be repaired more easily than if the whole mirror were metal.

Telescopes are made to have a large aperture

(a) to reduce diffraction effects, and

(b) to make the image of low power sources bright enough.

They are reflecting because

(a) there is no chromatic aberration,

(b) it is relatively easy to make a paraboloidal reflecting surface,

(c) it is difficult to maintain uniform refractive index throughout a large volume of glass,

(d) the suspension of a large lens is more difficult to achieve than that of a large mirror.

Note that the paraboloidal mirror does *not* form a point image of a point object off the axis.

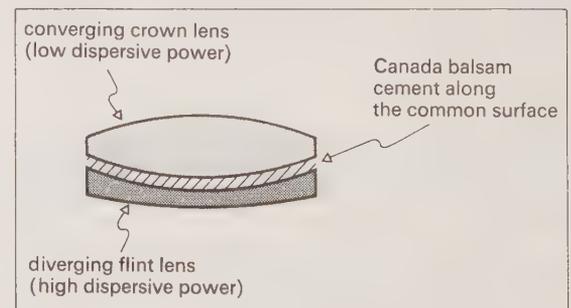


Fig. 34.4. A design of achromatic doublet.

35 DISPERSION AND THE SPECTROMETER

35.1	DISPERSION AND DISPERSIVE POWER	269
35.2	THE SPECTROMETER	270
35.3	OPTICAL SYSTEM OF THE SPECTROMETER	271

35.4	MEASUREMENT OF REFRACTIVE INDEX	271
35.5	SPECTROSCOPY	271
35.6	CLASSIFICATION OF SPECTRA	271

35.1 DISPERSION AND DISPERSIVE POWER

The deviation of a ray of light passing through a prism depends on refractive index n , which in turn is controlled by the light frequency. Light waves of different frequency (and therefore different wavelength in a vacuum) are deviated by the prism through different angles and are said to be **dispersed** (fig. 35.1).

Suppose n is measured for different wavelengths, and the graph of n against λ plotted for different glasses (fig. 35.2).

Different glasses disperse the light through different angles ΔD for a fixed mean deviation D .

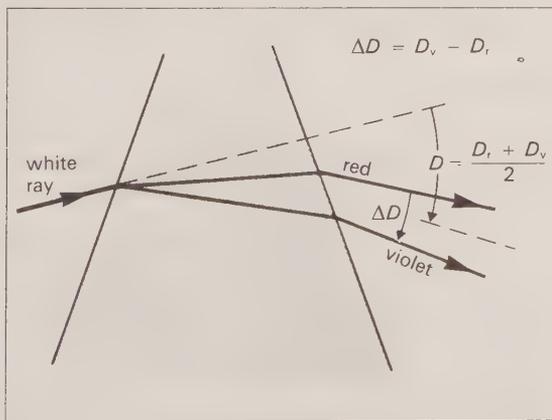


Fig. 35.1. Exaggerated dispersion of a ray of white light.

The **dispersive power** ω of a material is defined by

$$\omega = \frac{n_b - n_r}{n_y - 1}$$

and determines the size of ΔD . n_r , n_y and n_b are the refractive indices of the material for light of particular frequencies defined by the C, D and F *Fraunhofer* lines (p. 273). Thus we might find for crown glass $\omega \sim 0.016$, and for flint glass $\omega \sim 0.029$.

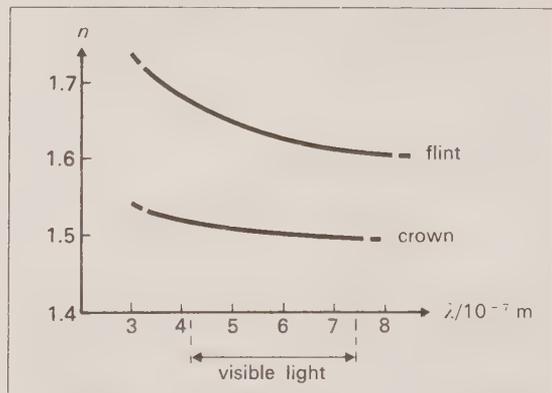


Fig. 35.2. Typical dispersion curves for flint and crown glasses.

Example

Consider a prism of small angle A working at small angles of incidence, so that

$$D = (n - 1)A$$

It follows that

$$\Delta D = \Delta n A$$

$$(D_b - D_r) = (n_b - n_r)A$$

and that the mean deviation

$$D_y = (n_y - 1)A$$

So for this special case

$$\omega = \frac{D_b - D_r}{D_y}$$

Special prism combinations

(a) *The achromatic prism* is a means of producing resultant deviation with zero resultant dispersion. Using two prisms of different glasses, rays of only two particular wavelengths can be made parallel.

In fig. 35.3 if a beam of white light is used, its top edge is coloured red, and the bottom violet. This is the basis of the achromatic lens (p. 268).

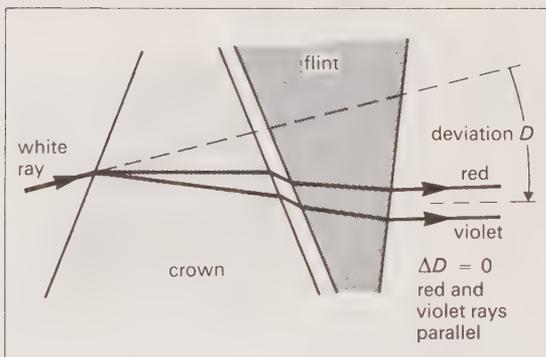


Fig. 35.3. *The achromatic prism.*

(b) *The direct vision prism* is a means of producing resultant dispersion without deviation for a particular wavelength near the middle of the spectrum (fig. 35.4).

The direct vision spectroscope uses a compound prism of this type for the rapid observation of spectra.

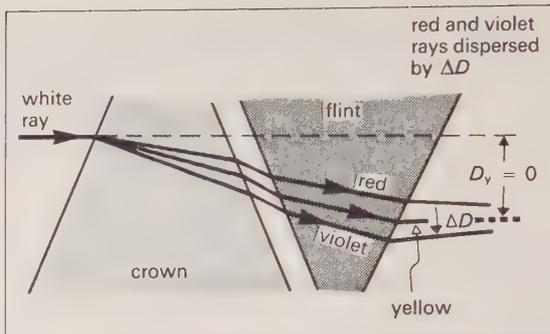


Fig. 35.4. *The principle of the direct vision prism.*

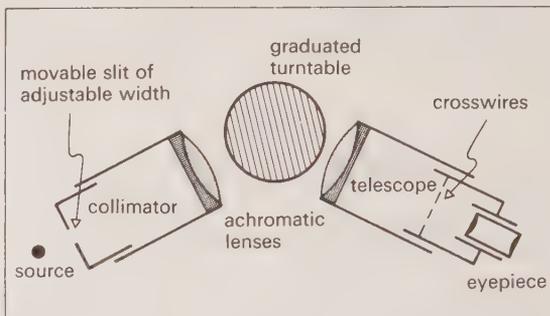


Fig. 35.5. *Parts of the spectrometer.*

35.2 THE SPECTROMETER

This is an instrument suitable for

- (a) producing, viewing and taking measurements on a *pure* spectrum (using either a prism or a diffraction grating), and
- (b) measuring accurately the refractive index of a material in the form of a prism.

Its construction

The essential parts are (fig. 35.5):

- (a) The **collimator** (which is fixed to the base of the instrument), consisting of a **slit** of variable width, and an achromatic lens.
- (b) The **table**, which can be rotated, and to which a prism or grating can be attached. The circular edge of the table has a scale graduated in degrees. (Sometimes we may need to convert from degrees to radians.)
- (c) The **telescope**, which can also be rotated. A **vernier scale** is fitted to the telescope where it adjoins the table, enabling their relative orientation to be measured to 0.1° , or less.

Adjustments

These should be made in the following order:

- (1) The **eyepiece** is focussed on the crosswires.
- (2) The objective lens of the *telescope* is focused so that the crosswires are in its focal plane.
- (3) Using a slit of width appropriate to the source brightness, the *collimator* lens is moved so that the slit is in its focal plane.
- (4) Using the table levelling screws, the axis of the *table* is made perpendicular to the plane containing the principal axes of the collimator and telescope lenses.

35.3 OPTICAL SYSTEM OF THE SPECTROMETER

Figs. 35.6 and 35.7 and their captions (p. 272) need most careful study.

In fig. 35.7, the two angles marked α are equal since they are vertically opposite.

$\alpha = \beta$, because parallel monochromatic rays are deviated through approximately equal angles by the prism when the rays from the centre of the slit undergo minimum deviation.

$\beta = \gamma$, alternate angles between parallel lines. The diagram shows construction lines drawn through the centre of the lens.

The two angles marked γ are equal since they are vertically opposite.

Because $\alpha_1 = \gamma_2$

$$\frac{x}{f_c} = \frac{y}{f_T}$$

This means that the image width y can be calculated.

35.4 MEASUREMENT OF REFRACTIVE INDEX

A knowledge of A and D_{\min} enables n to be found from

$$n = \frac{\sin [(A + D_{\min})/2]}{\sin (A/2)}$$

(A) Optical measurement of A

- (1) Place prism on table centre.
- (2) Record two telescope positions where reflected beams are observed (fig. 35.8, p. 273).
- (3) Angle between rays shown = $2(\alpha + \beta)$

$$\text{Angle of prism} = A = (\alpha + \beta)$$

Any kind of light can be used for this part of the experiment.

(B) Measurement of D_{\min}

This part of the experiment requires a source of the wavelength at which n is to be measured.

- (1) Set the telescope to accept rays that pass approximately symmetrically through the prism.
- (2) Clamp the telescope in such a position that one sees the succession of views shown in fig. 35.9, p. 273, as the table is rotated consistently in the same direction.
- (3) Clamp the table when the centre view is seen. Note the scale reading.
- (4) Remove the prism, and record the scale reading when the telescope receives light from the collimator directly. D_{\min} is the difference between these two readings.

35.5 SPECTROSCOPY

Spectroscopy is the study of emitted or absorbed electromagnetic radiation. It involves techniques of very high precision, and very high sensitivity (only a small mass of source material need be used).

A **spectrograph** is an apparatus producing dispersion, in which the spectrum is usually photographed.

A **spectroscope** lends itself for viewing by eye directly.

A **spectrometer** is designed for measurement.

Sources can be

- (a) **incandescent** when a body is at a high temperature (>1100 K or so), or
- (b) **luminescent**.
 - (i) **Fluorescence** occurs when a substance emits visible light when exposed to u.v. radiation.
 - (ii) **Phosphorescence** occurs when the emission of visible light takes place even though the u.v. source is no longer present.

35.6 CLASSIFICATION OF SPECTRA

Spectra are of two fundamental types:

- (i) An **emission spectrum** is what is seen when energy travels from the source to the observer uninterrupted.
- (ii) An **absorption spectrum** is what is seen when, before analysis, part of the emitted energy is absorbed by some intervening medium.

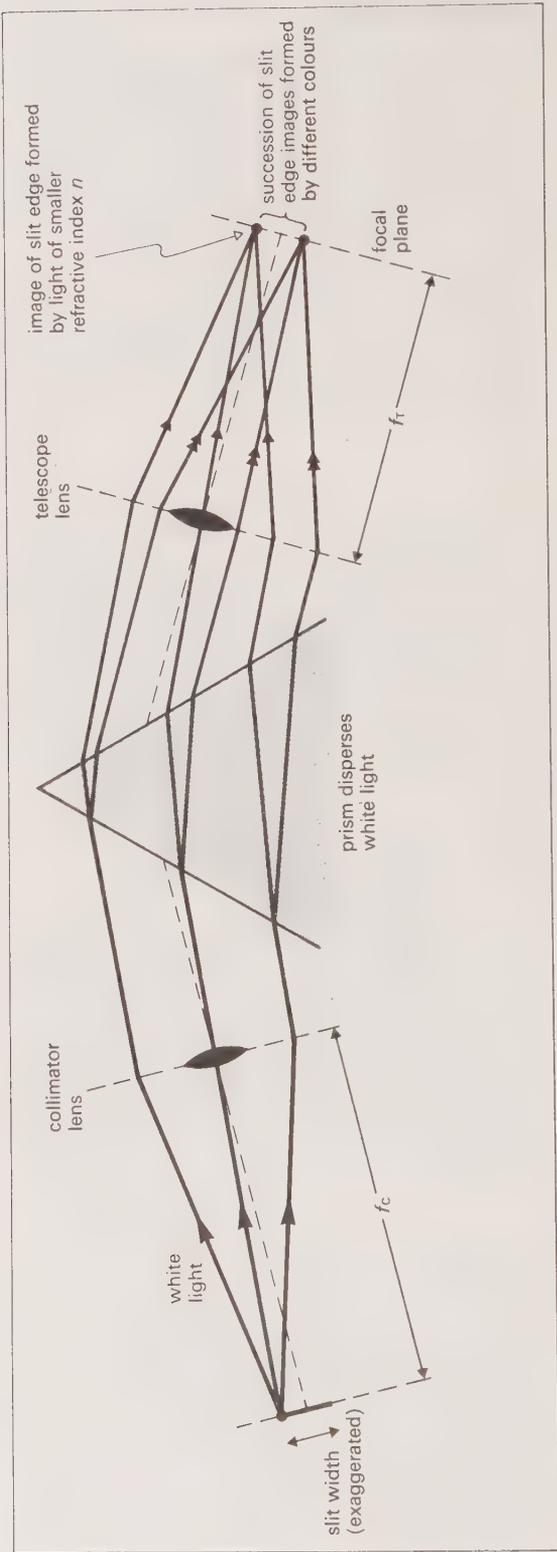


Fig. 35.6. The spectrometer, using white light, with rays drawn from one edge of the slit.

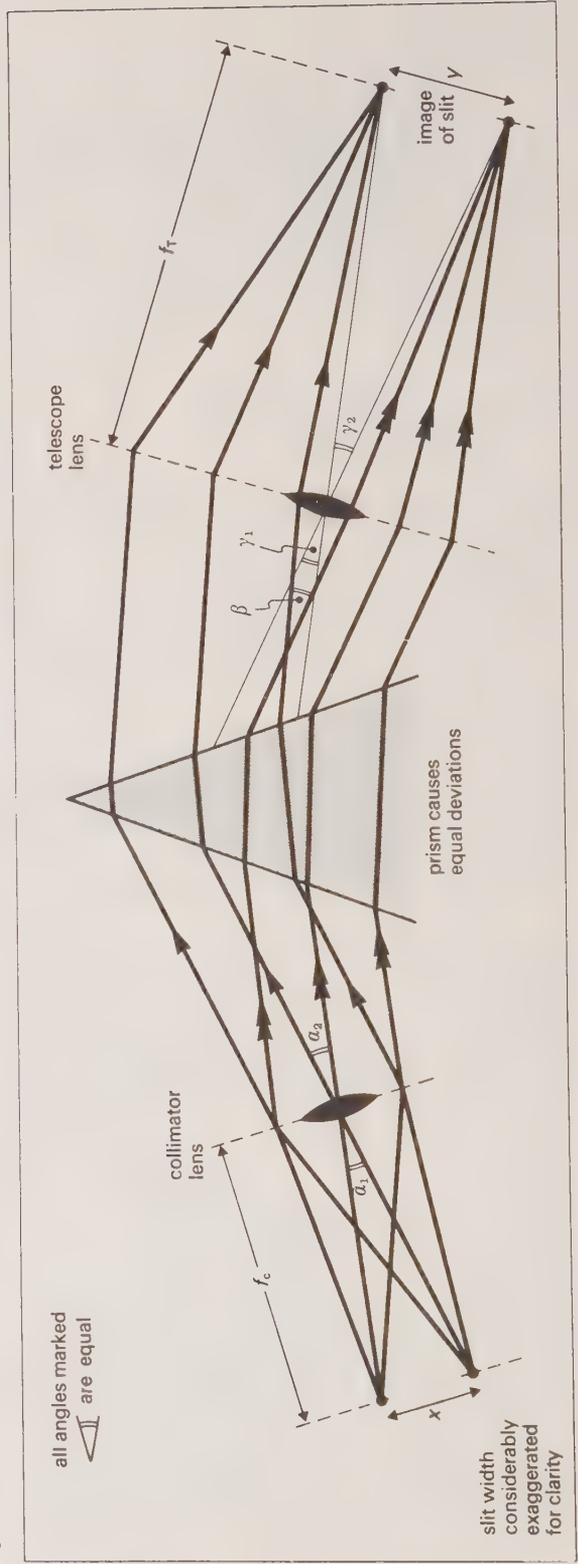


Fig. 35.7. The spectrometer, using monochromatic light, with rays drawn from both edges of the slit.

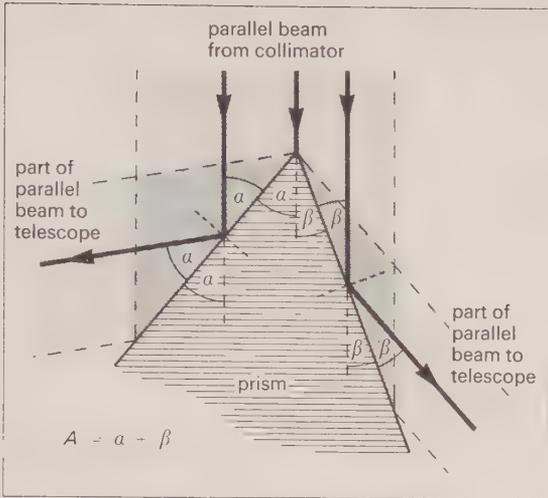


Fig. 35.8. Three rays from one beam of light which is split into two beams by reflection.

Both types are shown in fig. 35.10 (p. 274) as

- (a) a line spectrum, and
- (b) a band (fluted) spectrum, while the emission spectrum can also be
- (c) a continuous spectrum.

A continuous absorption spectrum may reduce the intensity of frequencies in one range of the spectrum, while another part of the spectrum is not absorbed.

(A) Line spectra

Emission

This is the result of exciting (giving extra energy to) substantially independent atoms. They are therefore emitted by luminous gases and vapours at low density. (See fig. 35.10(a), overleaf.)

Absorption

If light of all frequencies is passed through a gas, then the gas absorbs light of the same frequency as it would emit at the same temperature (p. 247). This is an example of resonance. The light is then re-radiated in all directions. This causes a reduction of intensity in the direction of the observer. This can be demonstrated by placing a tube of iodine vapour in front of a straight-filament white-light source and viewing it through a diffraction grating. Dark lines in the continuous white light spectrum correspond to those frequencies that have been absorbed by the iodine molecules. These frequencies have quanta with energies matching the energy differences between the various molecular energy levels. (See fig. 35.10(b), overleaf.)

In practice only a few of the lines in an emission spectrum can be matched with absorption lines.

Fraunhofer lines are dark lines crossing the otherwise continuous spectrum of the Sun. While the **photosphere** emits white light, energy of some frequencies is absorbed in the cooler gases of the **chromosphere**, resulting in a reduced intensity of these frequencies at the Earth's surface.

Fraunhofer labelled the most important lines A to H. They correspond to bright lines of emission spectra of atoms on Earth e.g.

- (a) the D line(s) correspond to the yellow light emitted by sodium, and
- (b) the C and F lines are the red and blue lines emitted by hydrogen.

These frequencies were used to define dispersive power (p. 269), but the sodium D line is now being replaced by the helium d line.

(B) Band spectra

Emission

This is the result of exciting substantially independent polyatomic molecules (e.g. molecules in a low

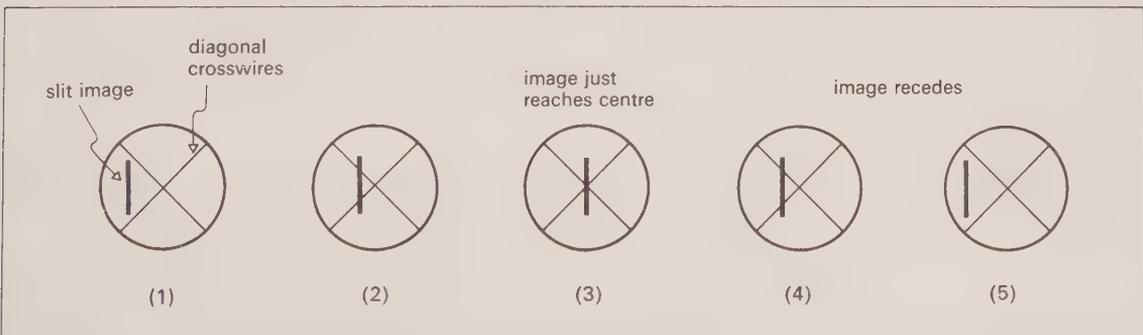


Fig. 35.9. Successive views seen as the table and prism are rotated together through the position of minimum deviation.

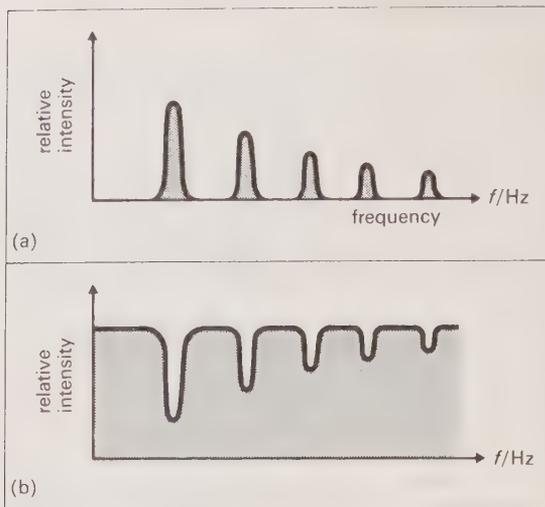


Fig. 35.10. Types of line spectra: (a) emission spectrum, (b) absorption spectrum for same material under identical conditions.

pressure gas). This could be done by electrical discharge. The emitted frequency corresponds to changes in

- (a) electron energy, giving visible light,
- (b) molecular vibrational energy, giving frequencies in the near infra-red, and
- (c) molecular rotational energy, giving frequencies in the far infra-red.

A **band** (fig. 35.11) is a large number of close lines which could be produced, for example, by superimposing rotational changes on a vibrational change.

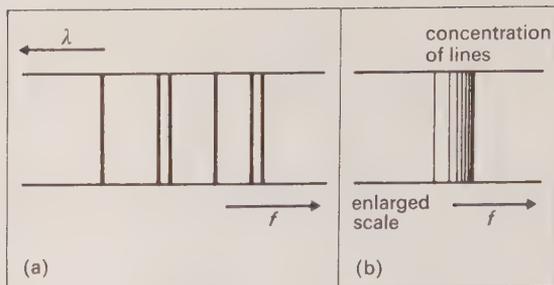


Fig. 35.11. (a) A band spectrum, and (b) details of one band.

Absorption

This is what one sees when observing a continuous spectrum after it has passed through a region which contains a band-emitting substance. It is useful for analysing molecular structure.

(C) Continuous spectra

When molecular collisions are very frequent (as in a compressed gas such as the Sun), or there is other strong molecular interaction (as in the liquid and solid phases), *all* frequencies are emitted. Thus a *continuous* spectrum is *not* characteristic of the nature of the source.

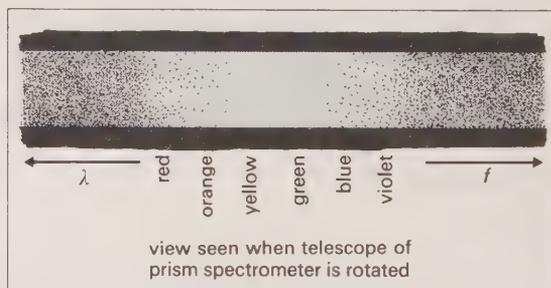


Fig. 35.12. Continuous spectrum from a carbon arc.

36 OPTICAL INSTRUMENTS

Sharda A

36.1 BASIC PRINCIPLES	275
36.2 THE SIMPLE MAGNIFYING GLASS	276
36.3 THE COMPOUND MICROSCOPE	278

36.4 ASTRONOMICAL TELESCOPES	278
36.5 THE EYE-RING	279
36.6 RESOLVING POWER OF A TELESCOPE	280

36.1 BASIC PRINCIPLES

The **near point** is the position of the nearest object that can be focussed both clearly and without strain by the unaided eye.

The **far point** is the position of the farthest object that can be focussed by the unaided eye.

For the **normal eye**

(a) the far point is at infinity, and

(b) the near point is taken to be a distance 250 mm from the eye. This is called the **least distance of distinct vision D** .

The **apparent size of an object** (and hence the discernible detail) is determined by the size of the image it forms on the retina, and so by the magnitude of the **visual angle θ** (fig. 36.1).

Angular magnification M (magnifying power) is defined by the equation

$$M = \frac{\theta_1}{\theta_0}$$

where θ_1 and θ_0 are defined in fig. 36.2. In general M is *not* the same as the linear magnification.

When calculating θ_1 , we imagine the eye to be placed close to the eye-lens of an instrument.

Diagrams

When considering optical instruments, it is often convenient to draw a diagram from which one can make calculations simply. Fig. 36.3(a) and (b) (overleaf) should be compared carefully.

Throughout this chapter diagrams similar to fig. 36.3(b) will be used, and frequently the angle to which we refer as θ is *strictly* half the visual angle. For clarity θ has been drawn grossly exaggerated.

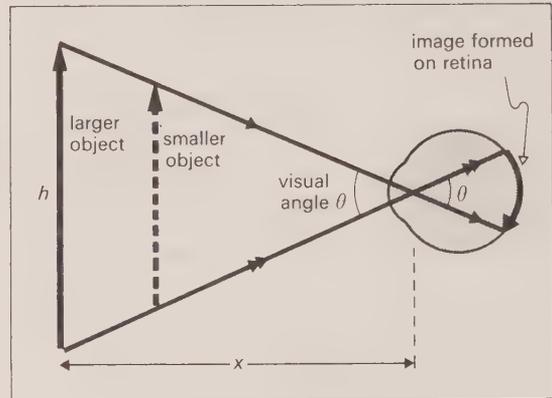


Fig. 36.1. Objects of different sizes may subtend the same visual angle, and thus be seen in equal detail.

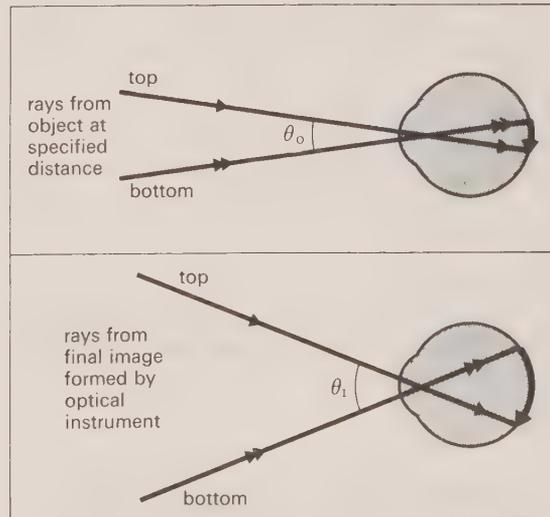


Fig. 36.2. θ_0 and θ_1 are the angles that determine magnifying power.

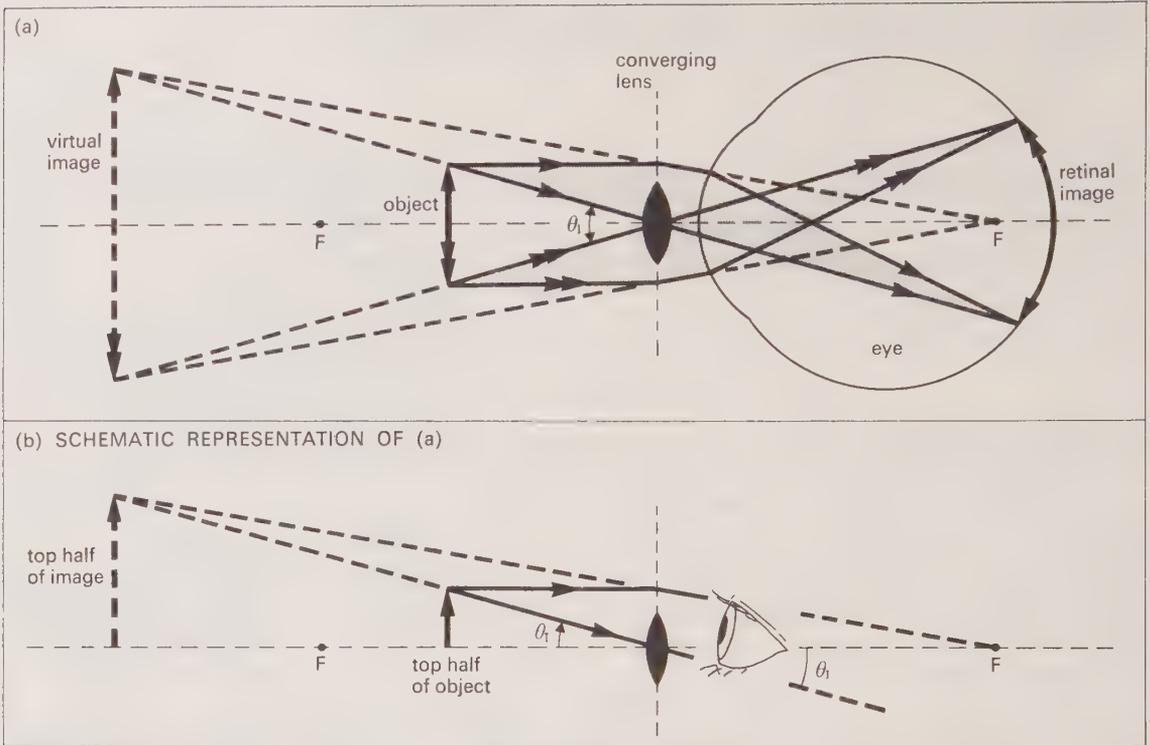


Fig. 36.3. Full and simplified ray diagrams for optical instrument theory. For calculation purposes we assume the eye to be placed very close to the eye-lens, as in (a).

Normal adjustment

An optical instrument is in normal adjustment when the final image is formed in a specified position.

(a) For a *telescope*, the image is at infinity, so that the eye is most relaxed.

(b) For a *microscope*, the image is at the near point, so that maximum angular magnification is obtained.

Approximations

We assume that all angles are small, so that angles may be replaced by their tangents. Thus in fig. 36.1, we may write $\theta = h/x$.

36.2 THE SIMPLE MAGNIFYING GLASS

By convention, θ_0 is the angle subtended when the object is placed at the near point (fig. 36.4).

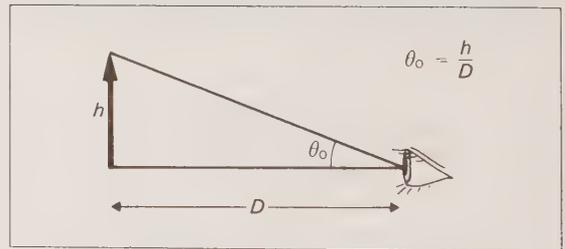


Fig. 36.4. Visual angle for object seen by unaided eye (for microscope calculations).

(A) Image at infinity (fig. 36.5)

Angular magnification

$$M = \frac{\theta_1}{\theta_0} = \frac{h/f}{h/D} = \frac{D}{f}$$

where D is the *magnitude* of the least distance of distinct vision. This is the smallest value that M can have.

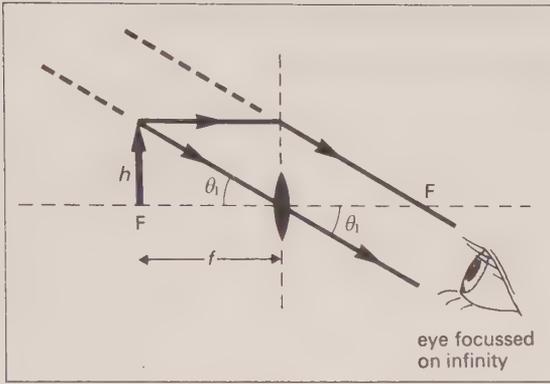


Fig. 36.5. Object placed in the focal plane of a converging lens.

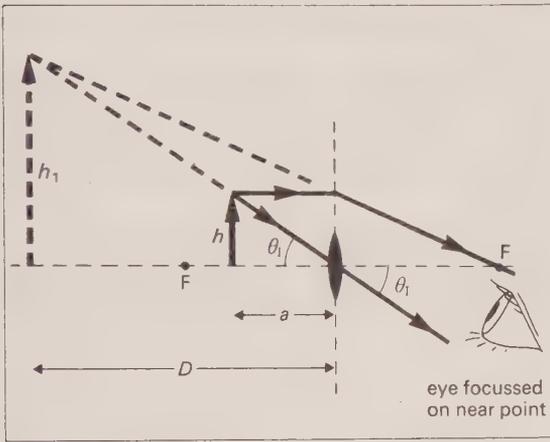


Fig. 36.6 Object placed on the lens side of a converging lens

(B) Image at near point (fig. 36.6)

$$M = \frac{\theta_i}{\theta_o} = \frac{h_1/D}{h/D} = \frac{h_1}{h} = \frac{D}{a}$$

Because $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$

$$\frac{v}{u} = \frac{v}{f} - 1$$

Here $\left| \frac{v}{u} \right| = \frac{D}{a} = \frac{D}{f} + 1$ since $v = -D$

So the magnitude of M is $\left(1 + \frac{D}{f}\right)$. This is the largest value that M can have.

M is increased by using a smaller value for f , but the aberrations involved may make the image more obscure.

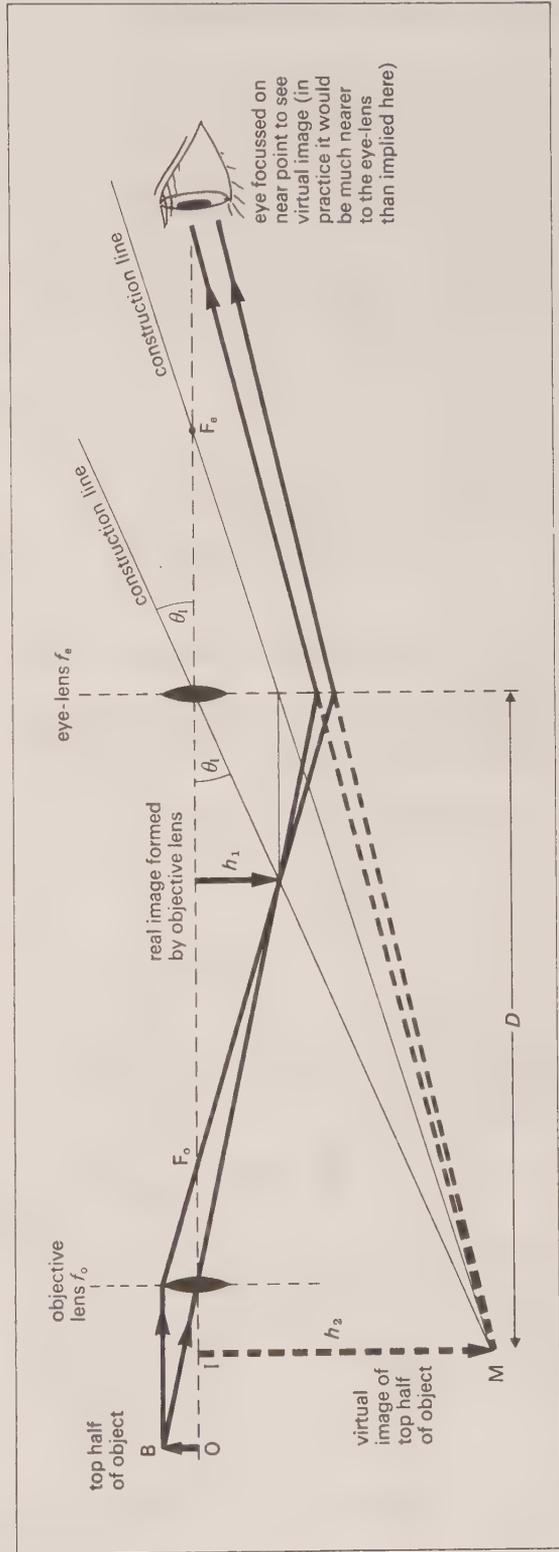


Fig. 36.7. The compound microscope in normal adjustment.

36.3 THE COMPOUND MICROSCOPE

The aim of the instrument is to produce an image on the retina larger than that obtainable by placing a small *accessible* object at the near point.

Refer to fig. 36.7. The objective lens produces an *enlarged real* image of the object. The eye uses the eye-lens as a magnifying glass to view this as a real object. In normal adjustment an *enlarged virtual* image is formed at the near point, 250 mm from the normal eye.

θ_O is evaluated when the object is placed at the near point (fig. 36.4). So

$$M = \frac{\theta_1}{\theta_0} = \frac{h_2/D}{h/D} = \frac{h_2}{h} = \frac{h_2}{h_1} \cdot \frac{h_1}{h}$$

$$M = \left(\begin{array}{l} \text{linear magnification} \\ \text{produced by eyepiece} \end{array} \right) \times \left(\begin{array}{l} \text{linear magnification} \\ \text{produced by objective} \end{array} \right)$$

This can be made larger by using smaller values for f_o and f_e .

Calculations on the microscope should *always* be carried out from first principles. Never attempt to remember formulae.

36.4 ASTRONOMICAL TELESCOPES

(A) Refracting telescope

The aim is to produce an enlarged retinal image of a distant *inaccessible* object.

Refer to fig. 36.8. Rays from the top of the object subtend an angle θ_O with rays from the bottom. Rays from any one point on the distant object are effectively parallel at the objective lens, so the objective lens forms a *diminished real* image in its focal plane. The eye-lens is used as a magnifying glass, and the final image is formed at *infinity* when the instrument is in normal adjustment.

$$M = \frac{\theta_1}{\theta_0} = \frac{h_1/f_e}{h_1/f_o} = \frac{f_o}{f_e}$$

Notes

(a) To make M large, we want a large f_o , but a small f_e .

(b) The length of the telescope tube = $(f_o + f_e)$.

(c) The resolving power depends on the diameter and quality of the objective lens. Large lenses are difficult to make, but small lenses do not deteriorate as fast as mirror surfaces. (See also p. 268.)

(d) The telescope drawn in fig. 36.8 can be converted to a **terrestrial** telescope, producing an upright image, by means of a third converging lens

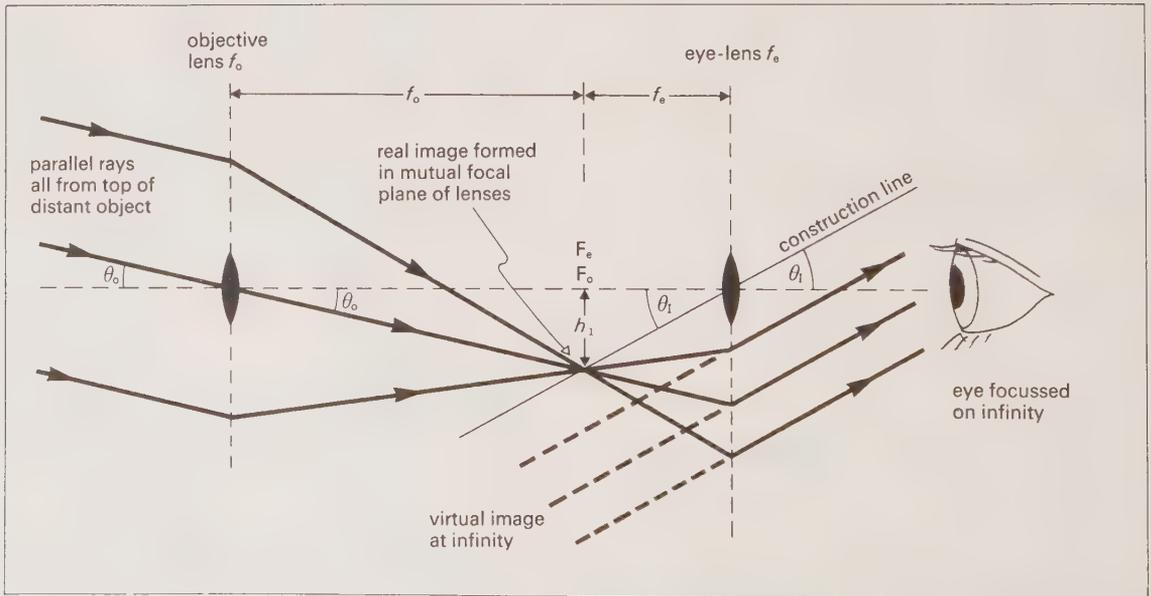


Fig. 36.8. The astronomical telescope in normal adjustment.

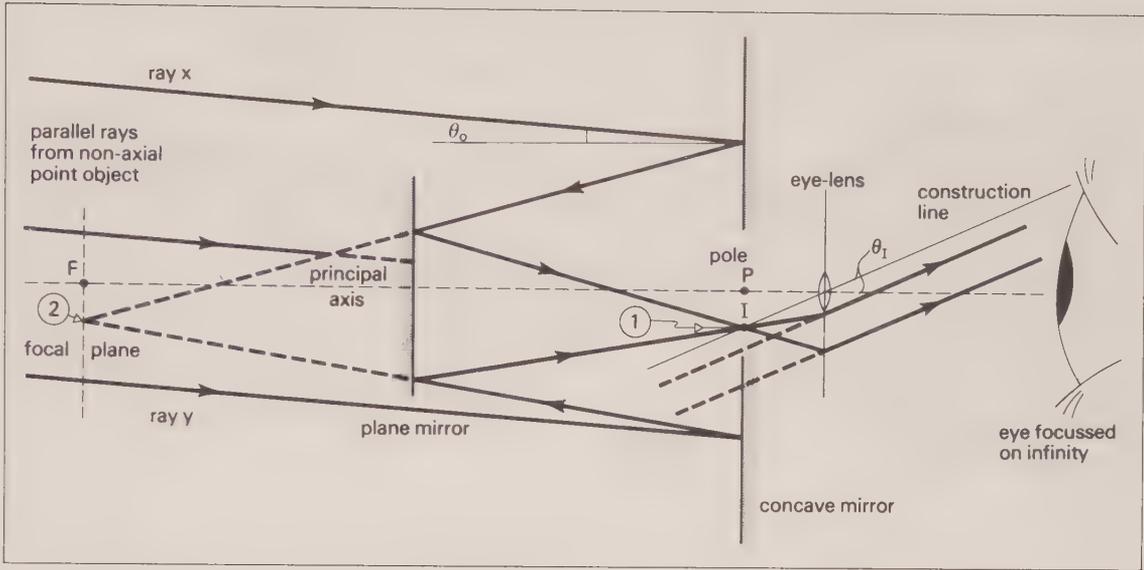


Fig. 36.9. Principles of the reflecting telescope.

(the *erecting* lens) placed between the objective and the eyepiece. The length of the telescope is increased by a minimum of four times the focal length of this lens (32.13). The *Galilean* telescope produces an upright image using only two lenses. The eyepiece is a diverging lens and the separation of the lenses is made equal to the difference of their focal lengths.

(B) Reflecting telescope

For the advantages of a reflecting system, see p. 268. The maximum aperture can be as much as 5 m, as opposed to 1 m for a refracting telescope.

If the image is to be photographed, the plate is placed at I, where a *real* image is formed (fig. 36.9). (Note that in practice the smaller mirror has an area which is small in comparison with that of the concave one.)

The *Cassegrain* reflector uses a convex mirror to bring the light to a convenient focus. The *Newtonian* reflector uses a plane mirror oriented so that the eye-lens axis is perpendicular to the mirror axis. The *Coudé* reflector is a combination of the two.

Procedure for drawing the reflecting telescope diagram (fig. 36.9)

(a) Draw the principal axis, the concave mirror and its principal focus F.

(b) Locate ① and ②, and draw the plane mirror mid-way between them.

(c) Draw the eye-lens and construction line so that $\theta_1 > \theta_0$.

(d) Draw arbitrary rays *x* and *y* reflected as to ②, then through ① and finally parallel to the construction line.

36.5 THE EYE-RING

In fig. 36.10 (overleaf) the circular area of diameter A'B' is the *smallest* area through which *all* rays from the distant object, which are refracted by both the objective lens and the eye-lens, are made to pass.

An observer should ideally

- place the eye pupil at A'B', and
- have a pupil diameter equal to A'B'.

A circular stop, of diameter not greater than that of the average eye pupil, is placed at A'B'. It is called the **eye-ring** and represents the *optimum position* for the observer's eye (fig. 36.11, overleaf).

We can show that

$$\frac{AB}{A'B'} = \frac{f_o}{f_e} = M$$

for a telescope in normal adjustment.

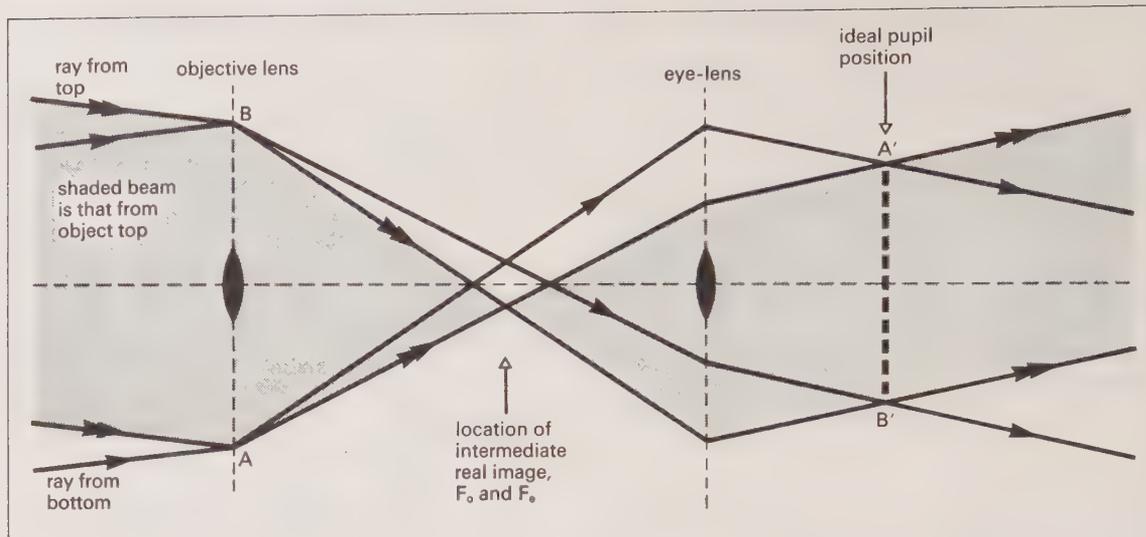


Fig. 36.10. To illustrate the behaviour of rays after they have passed through the eye-lens of a telescope in normal adjustment. (Not to scale.)

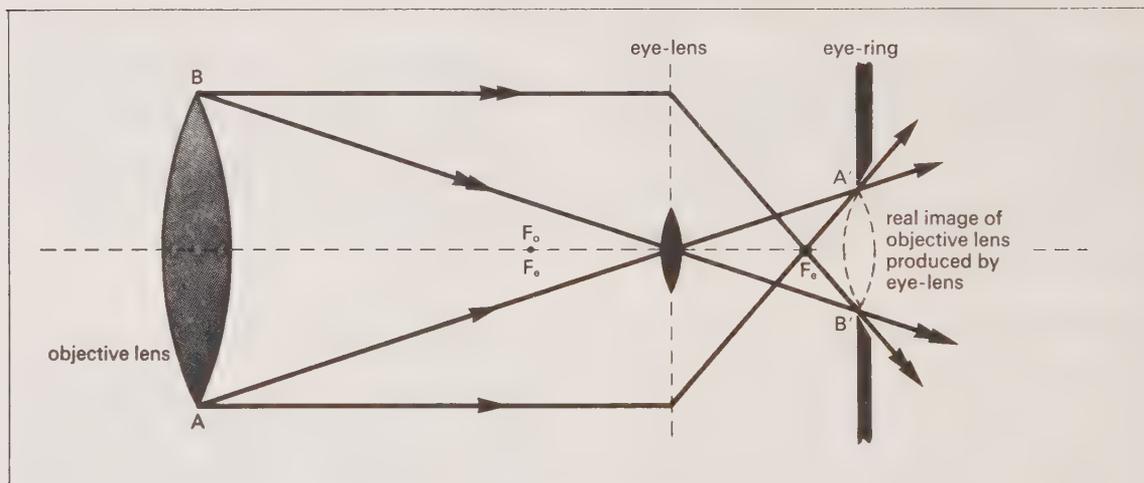


Fig. 36.11. Location of the eye-ring. (The student should note the close relationship with fig. 36.10.)

36.6. RESOLVING POWER OF A TELESCOPE

Diffraction (chapter 39) by the objective is the factor that limits the ability of a telescope to distinguish between two point objects that subtend a very small angle at the observer. Consider a double star system that can just be resolved according to the Rayleigh criterion (p. 310). If θ is the angular separation of the two stars, λ the wavelength being examined and D the diameter of the

objective, then since $\lambda \ll D$

$$\theta = 1.22 \frac{\lambda}{D}$$

The resolving power is the reciprocal of this angle which just allows resolution. The refracting telescope at the Yerkes Observatory has a diameter of 1.0 m. Thus for observations using $\lambda = 500 \text{ nm}$, we have $\theta = 0.6 \mu\text{rad}$. For the human eye $\theta \approx 0.2 \text{ mrad}$ ($200 \mu\text{rad}$).

VII WAVE PROPERTIES OF LIGHT

37	THE NATURE OF ELECTROMAGNETIC RADIATION	282
38	INTERFERENCE	292
39	DIFFRACTION	303
40	POLARIZATION	314

37 THE NATURE OF ELECTROMAGNETIC RADIATION

37.1	THE WAVE NATURE OF ELECTROMAGNETIC RADIATION	282
37.2	TYPES OF ELECTROMAGNETIC RADIATION	283
37.3	POLARIZATION AND COHERENCE	283
37.4	THE QUANTUM NATURE OF ELECTROMAGNETIC RADIATION	286

37.5	EXAMPLES OF WAVE PROPERTIES	288
37.6	THE SPEED c OF ELECTROMAGNETIC RADIATION	289
37.7	MICHELSON'S METHOD	291

37.1 THE WAVE NATURE OF ELECTROMAGNETIC RADIATION

Using the method of p. 98, we can classify electromagnetic waves as

- (a) *periodic waves*
- (b) *progressive waves* (usually, but see the comment on p. 123 and Wiener's experiment p. 289)
- (c) *transverse waves*, in which
- (d) the *disturbance* is a time-variation of the magnitude of the electric and magnetic fields E and B at a point.

Electromagnetic waves are generated by accelerated (often oscillating) electric charges, or by magnetic dipoles. Fig. 37.1 shows the nature of the disturbance.

Suppose the source of radiation is a straight vertical conductor (an antenna) that causes the radiation to travel outwards radially and symmetrically.

Fig. 37.2 is a horizontal section of such a source which shows the *wavefronts* and the associated *rays* (in this example *wave-normals*) which travel along the z -direction of fig. 37.1.

All electromagnetic waves can be propagated through a vacuum.

The relation between E , B and c can be found by considering the energy density in the electromagnetic wave. It can be shown that $c = E/B$.

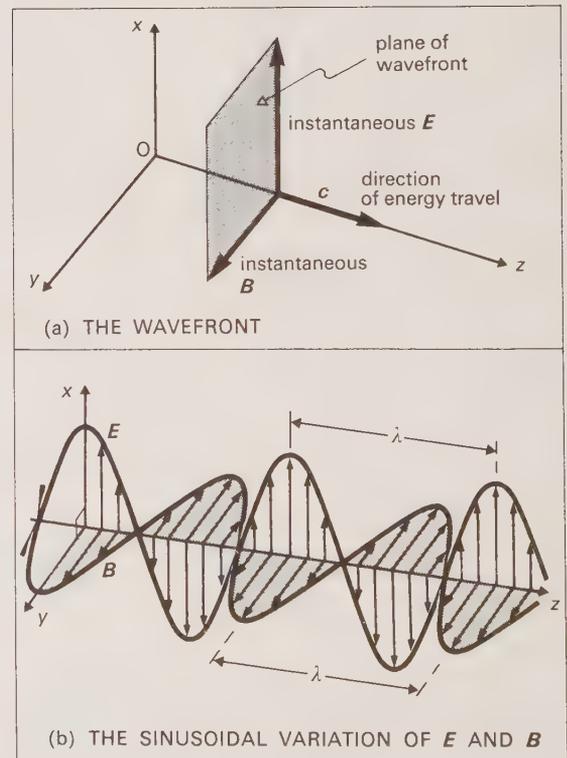


Fig. 37.1. The relative orientation of E , B and wave velocity c in an electromagnetic wave.

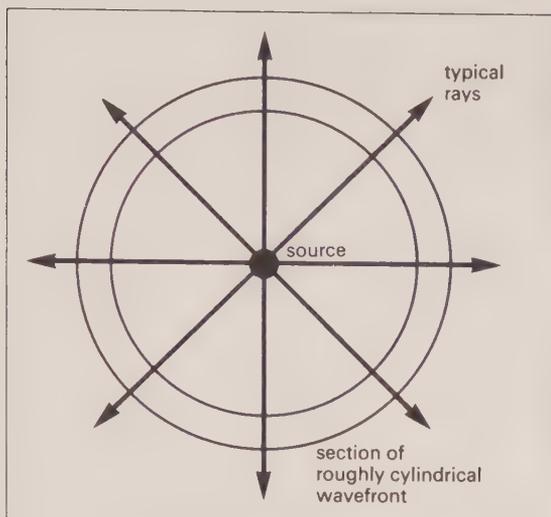


Fig. 37.2. Rays and wavefronts related.

Some orders of magnitude

Suppose a light beam has an intensity (p. 104) of 100 W m^{-2} . Then it can be shown that

$$E \approx 300 \text{ V m}^{-1}$$

$$B \approx 10^{-6} \text{ Wb m}^{-2} = 10^{-6} \text{ T}$$

Relative to typical laboratory fields, E is much larger than B . For this reason, and because the source of E (and not that of B) can exert forces on *stationary* electric charges, we are primarily concerned with the *electric vector* for optical effects. (Nevertheless the energies associated with the magnetic and electric fields are equal.)

Guided and free waves

When a high-frequency oscillator is connected across the ends of two long parallel plates, an electromagnetic wave travels along the plates which form a **waveguide**. Alternating regions of excess positive and negative charge on the plates have E - and B -fields associated with them. They all move along the guide at high speed and are reflected at the far end thereby forming a standing wave. (Note that the *electrons* move at much lower speeds.) A dipole aerial receiver can be used to measure the wavelength and therefore the speed of the wave, as long as the oscillator frequency is known. If the long parallel plates are removed and a transmitting dipole is attached to the oscillator, a strong signal is still detected by the dipole receiver. The electromagnetic wave is therefore travelling *freely* without a waveguide. These *guided*

and *free* electromagnetic waves have the same properties.

37.2 TYPES OF ELECTROMAGNETIC RADIATION

(A) Summary of parts of electromagnetic spectrum (see table overleaf)

The names given to different parts of the spectrum are for convenience only: they do not imply a rapid change of properties. The ranges of γ -rays and X-rays overlap because they are classified by *origin* rather than frequency. Note that the three sets of numbers are presented on logarithmic scales. (Compare with the graph of fig. 37.5, in which the scale is linear.)

(B) Properties of electromagnetic radiation

(See table on p. 285.)

37.3 POLARIZATION AND COHERENCE

Much electromagnetic radiation (e.g. that from radio and radar antennae) is **polarized** (p. 102) and relatively continuous: the plane containing the electric vector retains the same orientation. Infra-red, visible and ultra-violet waves are emitted *independently* by the individual atoms and molecules that make up the source. Since the atoms do not act **coherently**, the separate wavetrains (p. 293) that each produces are completely unrelated to those emitted by their neighbours. Their planes of vibration and instants of emission are random.

Polarization is discussed more fully in chapter 40, and coherence in chapter 38.

The laser

The **laser** (light amplification by stimulated emission of radiation) produces unidirectional, monochromatic, intense, *coherent visible* light. The **maser** operates in the microwave region of the electromagnetic spectrum. The solid-state ruby laser produces a **pulsed** beam whereas the helium-neon gas laser beam is **continuous**.

In a normal population of atoms in thermal equilibrium, there are far more atoms in the **ground state** (energy E_0) than in higher *excited* states (energies E_1, E_2 , etc.). The action of the laser is described in fig. 37.3 (p. 286). **Optical pumping**, using light from a xenon arc lamp for example, can lead to a **population inversion** where the number of atoms in energy state E_1 exceeds that in state E_0 . Photons of energy ($E_1 - E_0$) *stimulate* the excited

Summary of parts of the electromagnetic spectrum

Name and (approximate) range of radiation	*Energy per photon W/J	Frequency f/Hz	Wavelength λ /m	Common length units for comparison
	10^{-11}	10^{22}		
	10^{-12}	10^{21}	10^{-13}	
	10^{-13}	10^{20}	10^{-12}	picometre pm
	10^{-14}	10^{19}	10^{-11}	
	10^{-15}	10^{18}	10^{-10}	
	10^{-16}	10^{17}	10^{-9}	nanometre nm
	10^{-17}	10^{16}	10^{-8}	
	10^{-18}	10^{15}	10^{-7}	
	10^{-19}	10^{14}	10^{-6}	micrometre μ m
	10^{-20}	10^{13}	10^{-5}	
	10^{-21}	10^{12}	10^{-4}	
	10^{-22}	10^{11}	10^{-3}	millimetre mm
	10^{-23}	10^{10}	10^{-2}	
	10^{-24}	10^9	10^{-1}	
	10^{-25}	10^8	1	metre m
	10^{-26}	10^7	10^1	
	10^{-27}	10^6	10^2	
	10^{-28}	10^5	10^3	kilometre km
	10^{-29}	10^4	10^4	
	10^{-30}	10^3	10^5	

*W can be expressed relative to the quantity *electronvolt* (p. 350) by dividing these numbers by 1.6×10^{-19}

Properties of electromagnetic radiation

(N.B. All have the wave properties associated with a transverse wave motion.)

Name	Generated by	Dispersed by	Detected by	Particular properties
<i>γ-rays</i>	Changes of energy levels in the nucleus			(a) Penetrate matter (e.g. radiography) (b) Ionize gases (c) Cause fluorescence
<i>X-rays</i>	(a) Rapid deceleration of fast-moving electrons (e.g. by a tungsten target) (b) Changes in energy of innermost orbital electrons		(a) Photography (b) Ionization chamber (c) Phosphorescence (e.g. ZnS)	(d) Cause photoelectric emission from metals (e) Reflected and diffracted by crystals, enabling ionic lattice spacing and N_A (or wavelength λ) to be measured
<i>ultra-violet</i>	Orbital electrons of atoms, as in high-voltage gas discharge tubes, the arc, the Sun, and the mercury vapour lamp	quartz fluorite (CaF ₂)	(a) Photography (b) Photoelectric cell (c) Fluorescence	(a) Absorbed by glass (b) Can cause many chemical reactions, e.g. the tanning of the human skin (c) Ionize atoms in atmosphere, resulting in the ionosphere
<i>visible light</i>	Re-arrangement of outer orbital electrons in atoms and molecules, e.g. gas discharge tube, incandescent solids and liquids	glass	(a) Eye (b) Photography (c) Photocell	(a) Detected by stimulating nerve endings of human retina (b) Can cause chemical action
<i>infra-red</i>	(a) Outer electrons in atoms and molecules (b) Change of molecular rotational and vibrational energies e.g. incandescent matter	rock salt	(a) Photography by special plate (b) Special heating effect (e.g. radiometer, bolometer, etc.) (c) Photoconductive cells (e.g. PbS)	(a) Useful for elucidating molecular structure (b) Less scattered than visible light by atmospheric particles—useful for haze photography
<i>microwaves</i>	Special electronic devices such as klystron tube	paraffin wax	Valve circuit arranged as microwave receiver Point-contact diodes Thermistor bolometers	(a) Radar communication (b) Analysis of fine details of molecular and atomic structure (c) Since $\lambda \sim 3 \times 10^{-2}$ m useful for demonstration of all wave properties on macroscopic scale
<i>radio waves</i>	Oscillating electrons in special circuits coupled to radio aerials		Tuned oscillatory electric circuit (i.e. radio receiver)	Different wavelengths find specialized uses in radio communications

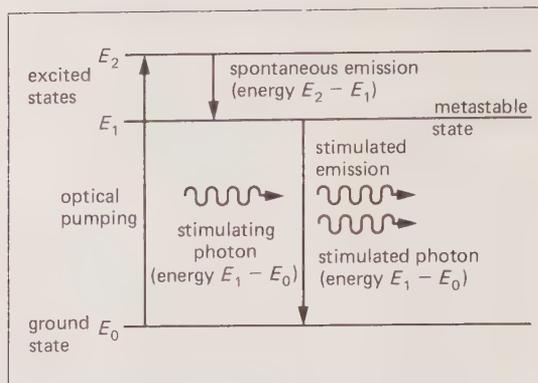


Fig. 37.3. The laser action.

atoms to return to the ground state. The emitted photons are in phase so that the resulting light beam is *coherent* and intense. Silvered surfaces or mirrors at the ends of the material increase the amplification because of the stimulation of photons by those undergoing multiple reflections. The laser beam emerges through a section at one end that is only partially silvered.

Uses of lasers

(a) *Medicine.* Applications include the treatment of malignant tumours, 'spot-welding' detached retinas, cutting through and cauterizing blood vessels, dental drilling. Recently they have been combined with fibre optics to vaporize obstructions such as blood clots and stones.

(b) *Telecommunications.* Modulated laser beams are transmitted along optical fibres and carry telephone, radio, television and computer signals.

(c) *Holography.* A **hologram** is produced by splitting a laser beam so that part (the *reference beam*) falls onto a special photographic plate directly, and part is scattered from the illuminated object. These two coherent beams form complex superposition patterns on the plate. A 3-D image is constructed in space from these patterns when the developed plate is illuminated by the laser beam from the original direction. Some holograms are viewed using transmitted light and some using reflected light. Holography has many applications including

- (i) microscopy—3-D recordings with high resolution are valuable in medicine and dentistry,
- (ii) measurements of very small movements, e.g. mechanical vibrational characteristics, design of artificial joints, detection of cracks and flaws in nuclear fuel elements,

- (iii) displays and exhibitions, e.g. valuable art objects, commercial goods,
- (iv) scanners for reading bar codes on supermarket goods,
- (v) phonecard and credit card markings to prevent forgery,
- (vi) special-purpose lenses.

(d) *Miscellaneous.* Lasers have an increasingly wide range of other applications, some of which are listed below:

- (i) detection of minute quantities of substances (single atoms),
- (ii) influencing chemical reactions,
- (iii) colourful artistic displays, pattern engraving and statue cleaning,
- (iv) welding and drilling,
- (v) military range finding and measurement of large distances, e.g. the Earth–Moon separation,
- (vi) surveying,
- (vii) printing,
- (viii) compact disc and videodisc players.

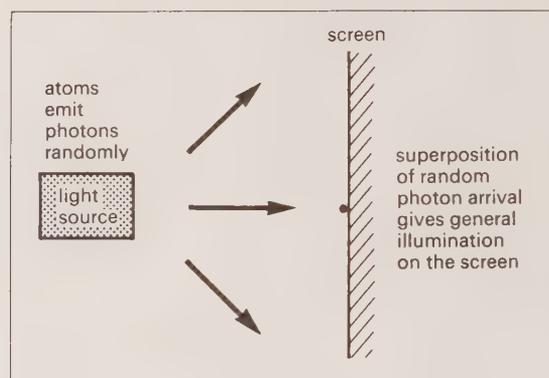


Fig. 37.4. Quantum nature of light.

37.4 THE QUANTUM NATURE OF ELECTROMAGNETIC RADIATION

Planck suggested that radiation is emitted in packets or **quanta**, which for i.r., visible light and the higher frequencies are called **photons**.

The quantum nature of light is *not* inconsistent with our observation of uniform illumination (fig. 37.4).

Orders of magnitude

The energy carried by a photon

$$W = hf$$

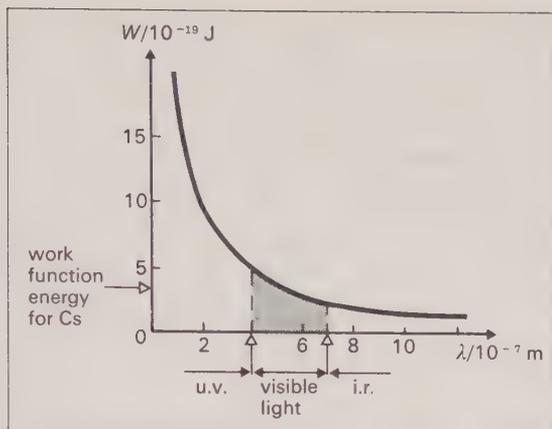


Fig. 37.5. Variation of photon energy W with the wavelength λ of electromagnetic radiation.

where h is the Planck constant (p. 504).

For u.v. photons

$$W \approx 10^{-18} \text{ J}$$

For γ -rays

$$W \approx 10^{-14} \text{ J}$$

Thus a lamp rated at 100 W will emit about 10^{20} photons in each second. (Note that this explains fig. 37.4.) The duration of emission of a photon varies from atom to atom, but typically

$$\Delta t \approx 10^{-9} \text{ s}$$

Since $c = 3 \times 10^8 \text{ m s}^{-1}$, the length of the wavetrain associated with a typical photon is only

$$\left(3 \times 10^8 \frac{\text{m}}{\text{s}}\right) \times (10^{-9} \text{ s}) = 3 \times 10^{-1} \text{ m}$$

Some photons are, however, associated with much longer wavetrains.

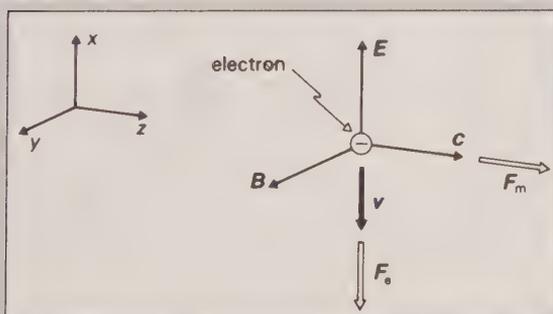


Fig. 37.6. Forces on an electron caused by the incidence of e.m. radiation.

► When electromagnetic radiation strikes an electron in a material, the E -field and B -field associated with the wave exert forces F_e and F_m as shown in fig. 37.6.

The radiation force F_r exerted on the material by the e.m. radiation is equal to the magnetic force F_m , and so is given by

$$F_r = Bev \quad (\text{p. 401})$$

$$\text{As } B = \frac{E}{c}, \quad F_r = \frac{(Ee)v}{c} = \frac{F_e v}{c}$$

F_e (and not F_m) is the force that does work on the electron, so

$$F_r = \frac{P}{c}$$

where P is the power of the wave.

This expression is for total absorption of radiation—if there is complete reflection, the radiation force is twice as great. The radiation pressure $p_r = F_r/A$ and so, for complete absorption or reflection,

$$p_r = I/c \text{ or } 2I/c$$

respectively, where I is the intensity of radiation.

The radiation pressure at the Earth's surface is about 10^{-6} Pa , whereas atmospheric pressure is about 10^5 Pa .

By writing $F_r = P/c$

in the form $\frac{dp}{dt} = \frac{dW}{dt}/c$

and integrating, we see that the linear momentum p carried by a photon becomes

$$p = W/c$$

in the direction of c .

As $W = hf$ and $c = f\lambda$, p is also given by

$$p = \frac{h}{\lambda}$$

For a photon of blue light the linear momentum

$$p \approx 10^{-27} \text{ N s}$$

Wave-particle duality

Electromagnetic radiation consists of waves emitted in packets which sometimes appear to behave like particles.

Some properties are better explained by a wave theory (refraction, diffraction, etc.), whereas others are amenable to a particle explanation (the photoelectric effect, p. 503, and the Compton effect, p. 510).

The particle aspects of behaviour are more pronounced where the photons carry a large amount of energy (the higher frequencies), and the wave aspects are pronounced where the photons have smaller energies so that the energy flow is more continuous (the lower frequencies).

► De Broglie waves

In 1924 *de Broglie* suggested that material particles such as electrons might show wave properties. The *de Broglie* wavelength λ of a material particle is given by

$$\lambda = \frac{h}{p} = \frac{h}{mv}$$

where m is the relativistic mass and v the velocity of the particle.

In 1927 *Davisson* and *Germer* confirmed the prediction of the wave nature of particles with their **electron diffraction** experiments. Theirs was a *reflection* experiment, in which a beam of electrons was diffracted by a nickel target in the form of a single crystal. (The arrangement was similar to that shown in fig. 37.7.) The measured wavelength associated with the electrons agreed well with the *de Broglie* wavelength calculated from the accelerating p.d. V .

$$\text{k.e. of electrons } E_k = Ve = \frac{1}{2}m_e v^2$$

and so
$$\lambda = \frac{h}{mv} = \frac{h}{\sqrt{2Vem_e}}$$

Diffraction effects have since been shown for various other particles such as hydrogen atoms, helium atoms and neutrons. Neutron diffraction is an excellent way of studying crystal structures. An electron in its orbit can be replaced by a *de Broglie* wave of wavelength $\lambda = h/(m_e v)$. If each circular orbit round the nucleus is to contain an integral number of waves, then orbits of only certain radii are possible. If r is the radius of the orbit and n is an integer, then

$$n\lambda = 2\pi r$$

and so
$$m_e v r = n \frac{h}{2\pi}$$

As will be seen later (p. 516) this, *Bohr's* first postulate, was an assumption rather than a wave mechanics derivation.

The angular momentum carried by a photon is $h/2\pi$, being determined by a change in a quantum number of the atom emitting the photon. This ensures that the atom-photon system conserves angular momentum when the photon is emitted.

The waves described are not real, but the particles move in a manner which is better described by a wave equation (*Schrödinger*) than by *Newtonian* mechanics.

37.5 EXAMPLES OF WAVE PROPERTIES

Electromagnetic waves exhibit all the wave properties shown on p. 99, but the effects may be difficult to demonstrate. The most striking properties are discussed in chapters 38 to 40.

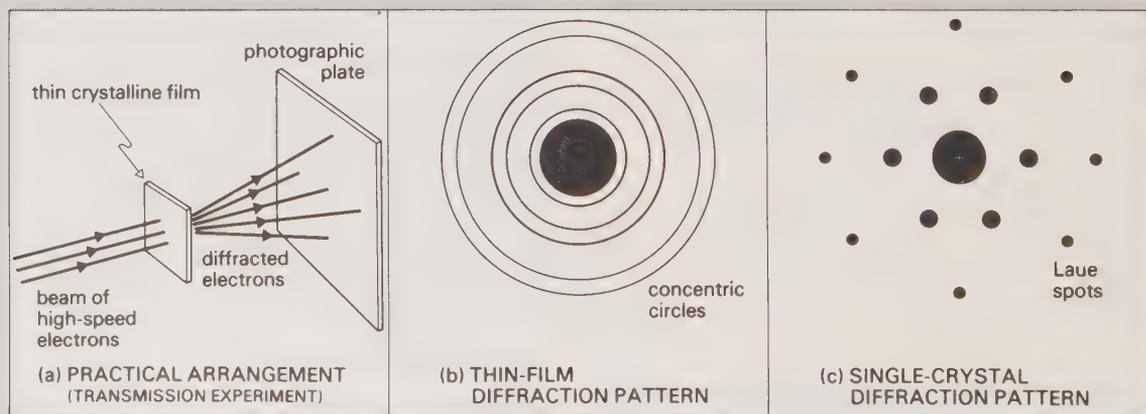


Fig. 37.7. Diffraction illustrating the wave behaviour of electrons.

Beats

Two electromagnetic waves beating together would have a frequency

$$f_{\text{beat}} = (f_2 - f_1)$$

Consider the two D-lines of sodium. Their respective frequencies are

$$f_2 = 5.090 \times 10^{14} \text{ Hz}$$

and

$$f_1 = 5.085 \times 10^{14} \text{ Hz}$$

f_{beat} is about 5×10^{11} Hz.

This effect for visible light obviously cannot be detected using separate sources when one remembers that they will be non-coherent. Nevertheless beats have been observed from a beam superposed on its own reflection from a *moving* mirror. (The *Doppler effect* (p. 119) causes the two beams to have *slightly* different frequencies.)

It is also possible to demonstrate beats from a laser being made to oscillate in two longitudinal modes. The beat frequency of about 150 MHz is easily shown by shining the laser output onto a photodiode, and feeding the output into the aerial socket of a radio.

Stationary waves

Fig. 37.8 shows how Wiener (1889) demonstrated stationary light waves. The emulsion was most exposed where it cut the *electric vector antinodes*.

Orders of magnitude

If $\lambda \sim 5 \times 10^{-7} \text{ m}$

and $\alpha \sim 2 \times 10^{-2} \text{ rad}$

then $d \sim 1.5 \times 10^{-5} \text{ m}$.

Obviously α needs to be very small.

The thickness of the emulsion was about $\lambda/20 \sim 20 \text{ nm}$.

Standing microwaves and radio waves are relatively easy to demonstrate. Fig. 37.9 shows a typical experimental arrangement.

37.6 THE SPEED c OF ELECTROMAGNETIC RADIATION

Importance of a knowledge of c

(a) For confirmation of *Maxwell's* theory of electromagnetic radiation, which had predicted

$$c = \frac{1}{\sqrt{(\epsilon_0 \mu_0)}}$$

in which μ_0 has a defined value and ϵ_0 has a value that must be found by experiment. It can be seen

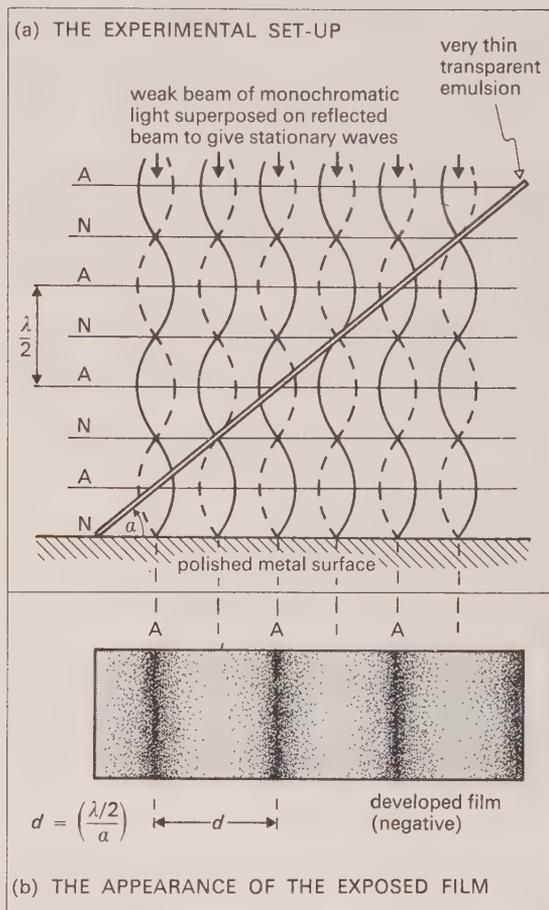


Fig. 37.8. Wiener's experiment to demonstrate stationary light waves.

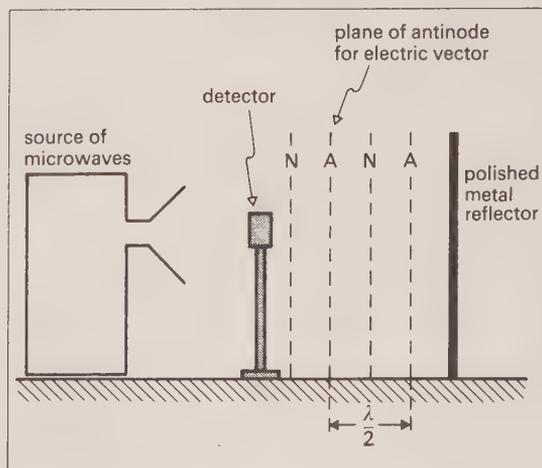


Fig. 37.9. Standing microwaves.

that this equation is dimensionally consistent since

$$[\epsilon_0] = [M^{-1}L^{-3}T^4I^2]$$

and

$$[\mu_0] = [MLT^{-2}I^{-2}]$$

N.B. Nowadays c can be measured more accurately than ϵ_0 , so one measures c , and then calculates ϵ_0 from the above equation. See also p. 361.

(b) Using $c = \text{distance}/\text{time}$, one can measure large distances (e.g. in astronomy) by timing radio pulses.

(c) The important part played by c in relativistic physics, and hence in modern atomic and nuclear physics. (For example the well-known equation $E = mc^2$.)

Historical review

(1) Roemer's method (1675)

He deduced the relatively large time taken for light to travel across the diameter of the Earth's orbit round the Sun.

(2) Bradley's method (1725)

The orbital velocity of the Earth combines vectorially with the velocity of light from a star, and may

cause an aberration between the star's true and apparent direction. The speed of light can be calculated from the necessary change in the orientation of the telescope.

(3) Fizeau's method (1849)

A rotatable *toothed wheel* served two purposes:

(a) It chopped, or *modulated*, a continuous beam of light into pulses.

(b) It measured the time interval separating adjacent pulses, and so their transit time.

An echo technique was used so that the pulses covered a known distance in a known time.

(4) Foucault's method (1860)

He used a rotating mirror to measure the speed of light through air, and through water, and discovered that the latter was smaller. This experiment was for a long time considered to be a crucial one for comparing the different explanations offered by the wave and corpuscular theories for refraction (p. 107).

(5) Michelson's method (1926)

This is discussed separately in 37.7. The modulation was achieved by a rotating *octagonal mirror*.

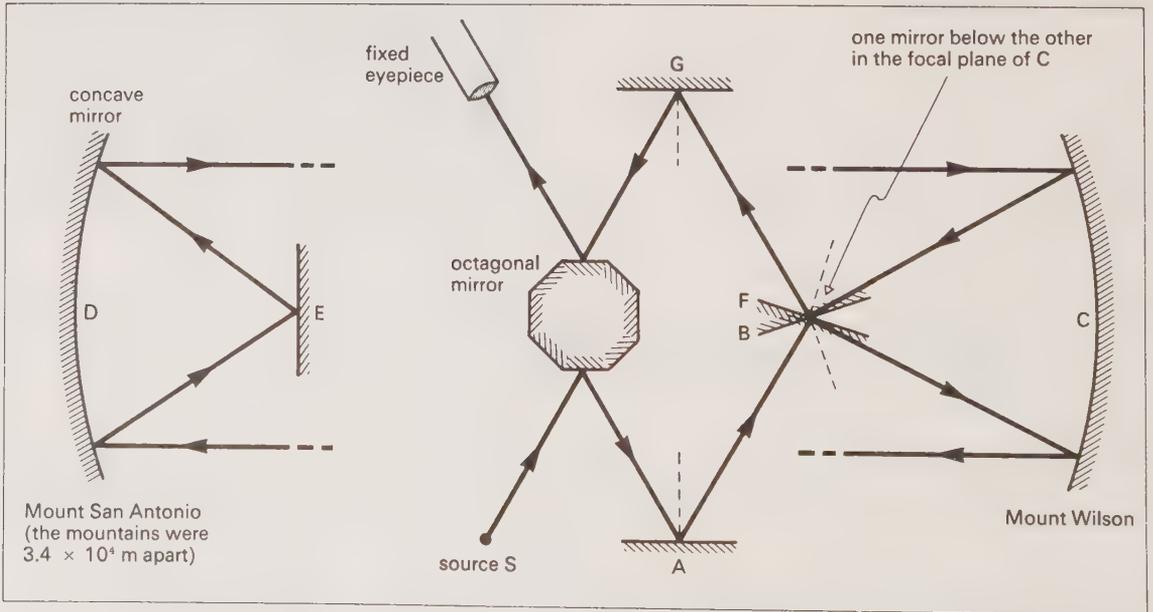


Fig. 37.10. Michelson's method for finding the speed of light.

(6) Anderson's method (1941)

Similar in principle to *Fizeau's* and *Michelson's*. The modulation was achieved by an *electronic shutter* called a *Kerr cell*.

(7) Recent methods

Measurements made more recently have been based on the standing waves set up in a cavity by short radio waves.

The value of c recommended by the International Union of Pure and Applied Physics is

$$(2.997\,924\,580 \pm 0.000\,000\,012) \times 10^8 \text{ m s}^{-1}$$

37.7 MICHELSON'S METHOD

Refer to fig. 37.10. The light from the source follows this path: S, octagonal mirror, A, B, C, D, E,

D, C, F, G, octagonal mirror, eyepiece.

During this time the mirror is made to rotate exactly one eighth of a revolution, so that an image is seen in the eyepiece. For slightly higher or lower rates of rotation, no image will be seen.

Details

(a) The rotational frequency of the octagonal mirror (528 revolutions per second) can be measured stroboscopically.

(b) The separation of the mountains was measured with an uncertainty of about 1 in 10^7 .

Calculation

$$\begin{aligned} \text{Speed} &= \frac{\text{distance travelled}}{\text{time taken}} \\ &= \frac{\text{twice mountain separation}}{\text{time for mirror to make } \frac{1}{8} \text{ rev.}} \end{aligned}$$

38 INTERFERENCE

38.1	CONDITIONS FOR WAVE INTERFERENCE TO BE VISIBLE	292
38.2	OPTICAL PATH	293
38.3	CLASSIFICATION OF FRINGE SYSTEMS	294
38.4	YOUNG'S DOUBLE-SLIT EXPERIMENT	294
▶ 38.5	OTHER TWO-SOURCE EXPERIMENTS	296

38.6	WHITE LIGHT INTERFERENCE	297
▶ 38.7	PARALLEL-SIDED THIN FILMS	298
▶ 38.8	WEDGE FRINGES	300
▶ 38.9	NEWTON'S RINGS	301
▶ 38.10	SUMMARY OF THIN FILM EFFECTS	302

38.1 CONDITIONS FOR WAVE INTERFERENCE TO BE VISIBLE

By **interference** we mean the superposition of wavetrains from a finite number of coherent sources. Two sources are frequently used.

For all types of wave motion

The following conditions must be observed if we are to see interference effects.

(a) The sources must have roughly the same amplitude.

(b) The sources must be **coherent**, i.e. they must have

- (i) a constant phase relationship, which implies
- (ii) the same frequency (and therefore the same wavelength where they are to be superposed).

Sources which have a constant phase relationship are said to be **phase-linked**.

(c) Transverse waves must be either unpolarized, or have significant resolved parts in the same plane.

When the conditions are not observed

(a) *Amplitudes unequal.* Fig. 38.1 shows the effect on the intensity–location curve. Cancellation is not complete, and the *contrast* between constructive and destructive superposition may not be detected easily.

(b) A time-dependent phase difference between the sources would result in a varying phase difference between the superposed wavetrains. The fringe pattern would be constantly moving, perhaps too rapidly for detection by an eye (whose retinal image persists for at least 0.05 s).

(c) *Polarization:* vectors must lie along the same direction if they are to be able to cancel one another.

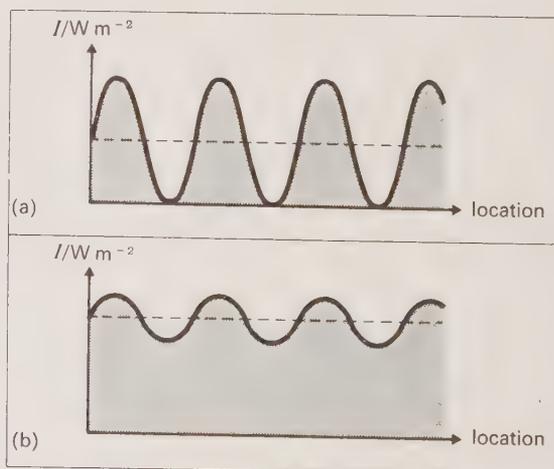


Fig. 38.1. Superposition using (a) equal and (b) unequal wave amplitudes.

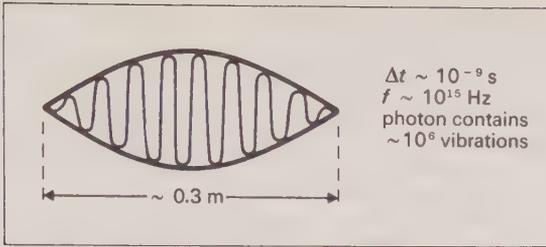


Fig. 38.2. A way of visualizing a typical photon.

Visible light wavetrains

Refer to the comments on the quantum nature of light (p. 284). Two coherent visible light wavetrains can be produced by deriving them either from a single slit source, or from a laser.

Fig. 38.2 shows a way of visualizing a photon emitted by a typical light source. The difference between the paths travelled by two wavetrains must not exceed the length of such a photon, or else the two wavetrains will not overlap. The twin sources (derived from the single source) are usually placed close together. (If they are not, the fringe width may be too small to measure.)

Reflection

There is a *phase change* of π rad when the light wave is reflected at a rare–dense interface (such as that between air and glass). This is illustrated by fig. 38.3.

This behaviour resembles that of a transverse wave on a string (p. 99).

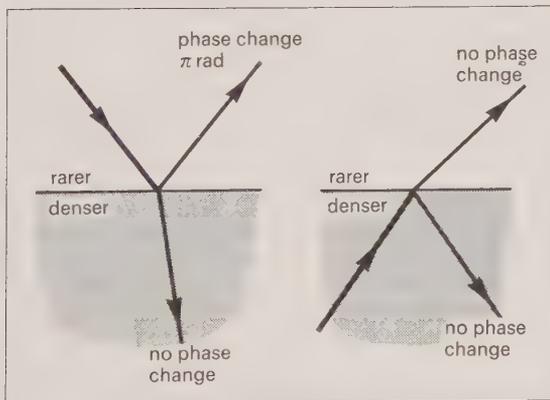


Fig. 38.3. Light suffers a phase change on reflection at a denser medium.

Brightness and the electric vector

Brightness seen by the eye corresponds to maximum intensity (rate of arrival of energy per unit area). For visible light we see

(a) *bright fringes* where the *amplitude* of the electric field vector is large (its magnitude but not its direction is important),

(b) *dark fringes* where the electric field vector is *permanently* smaller than at adjoining places.

(We discuss the electric vector for convenience: equal energy is associated with the electric and magnetic fields of the electromagnetic wave.)

38.2 OPTICAL PATH

When light of frequency f and free-space wavelength λ_0 crosses into a medium, f does not change, but λ_n becomes (λ_0/n) where n is the absolute refractive index of the medium.

The number of waves contained in a distance l is l/λ , where λ refers to the medium concerned (fig. 38.4).

In a length l in a vacuum, there are (lf/c) waves.
In a length l in a medium, there are

$$\frac{lf}{c_n} = \frac{lf}{(c/n)} = n \left(\frac{lf}{c} \right) \text{ waves}$$

The same number of waves fit into a length nl in a vacuum as into a length l in the medium. nl is called the **optical path** for that medium.

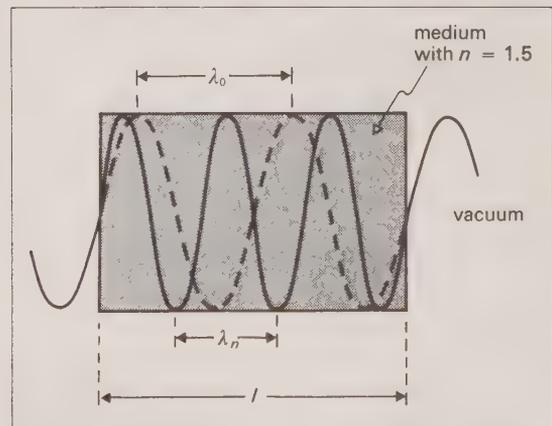


Fig. 38.4. The effect of a denser medium ($n=1.5$) on the wavelength of light: three wavelengths fit into a space previously occupied by two.

The total optical path for a wave traversing several media = $\sum nl$: it is this quantity that determines the phase of a wave at any point, since

$$(\text{phase difference}) = \left(\frac{2\pi}{\lambda}\right) \cdot \left(\frac{\text{total optical}}{\text{path difference}}\right)$$

38.3 CLASSIFICATION OF FRINGE SYSTEMS

(A) Classification by fringe location

Localized fringes

These have a position on which the eye or microscope must be focussed if they are to be seen. Fig. 38.5 gives some examples.

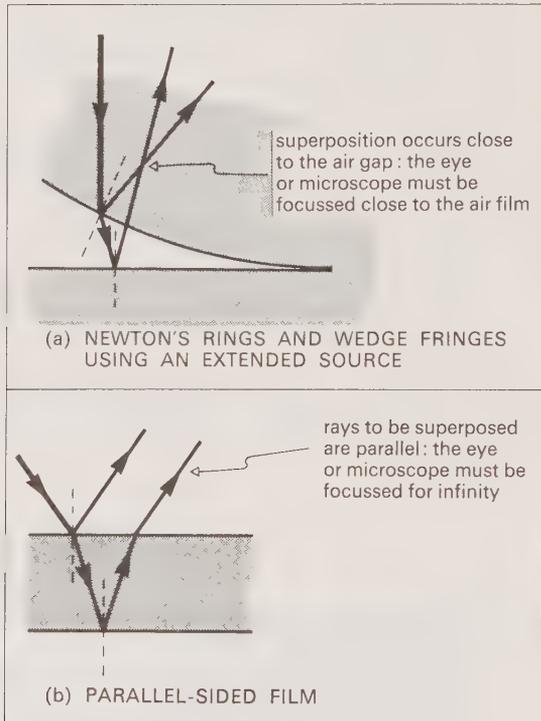


Fig. 38.5. Examples of localized fringe systems.

Non-localized fringes

These can be picked up even if one alters the location of the screen in a two-source experiment such as *Young's*, *Fresnel's* or *Lloyd's*. The fringes are non-localized in the sense that they can be seen at the intersection of the loci of antinodal lines (or surfaces) with an *arbitrary* screen (fig. 38.6).

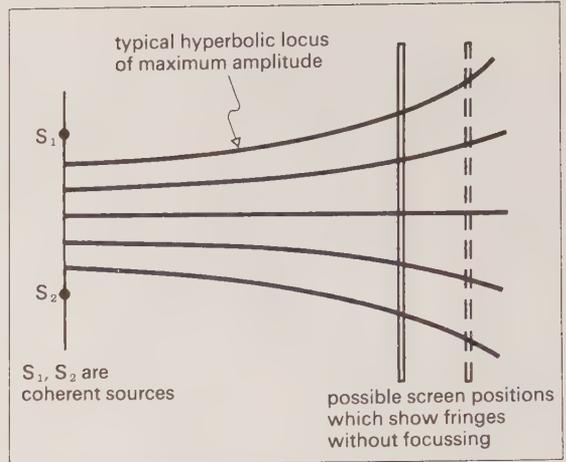


Fig. 38.6. Non-localized fringes in a two-source interference experiment.

(B) Classification by origin of wavefronts

Division of wavefront

In a typical two-source experiment the superposed wavetrains are obtained by dividing the wavefront given out by a single source (fig. 38.7).

Division of amplitude

In *Newton's* rings, wedge fringes and the effects obtained from a parallel-sided film, the incident wavefront is divided into two wavetrains by partial reflection and partial transmission. The new wavefronts each take part of the amplitude (and energy) of the incident wave.

38.4 YOUNG'S DOUBLE-SLIT EXPERIMENT

The fringe system obtained by *Young* (1801) gave positive evidence to support the theory of the wave nature of light. This experiment (and other pioneer experiments such as the *Newton's* rings experiment) were originally performed using *white* light.

The apparatus is shown in fig. 38.8 opposite.

Light from the monochromatic source is diffracted at S_0 so as to illuminate S_1 and S_2 *coherently*. *Diffraction* causes them to emit spreading coherent wavetrains which overlap *provided* S_1 and S_2 are narrow enough.

Calculation of path difference

In fig. 38.8(b)

$$\sin \theta = p/s$$

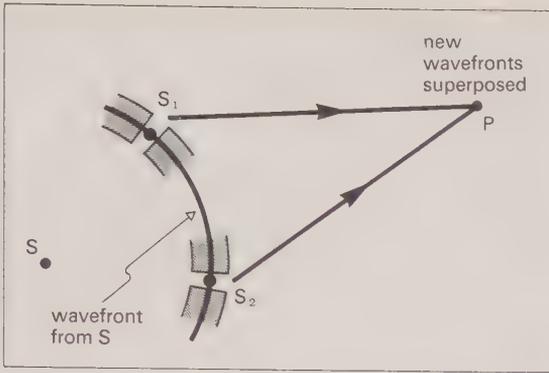


Fig. 38.7. Dividing a wavefront to obtain two coherent light sources, S_1 and S_2 (where usually $S_1S_2 \ll S_1P$ or S_2P).

where p is the path difference ($S_2P - S_1P$) for waves that are to be superposed at P . But

$$\tan \theta = d/D$$

If $\theta < 1/10$ rad, then $\tan \theta \approx \sin \theta$

$$\frac{d}{D} = \frac{p}{s}$$

so
$$p = \frac{ds}{D}$$

This is a purely geometrical relationship.

Condition for bright fringes

When

$$p = \frac{ds}{D} = m\lambda$$

the screen will show positions of reinforcement: these will appear as fringes parallel to the slits (fig. 38.9, overleaf). (m is any integer.)

Thus
$$d = \frac{m\lambda D}{s} \text{ for brightness}$$

The fringe width w is given by

$$d_{m+1} - d_m = \frac{\lambda D}{s}$$

$$w = \frac{\lambda D}{s}$$

Note explicitly that the fringe width is directly proportional to λ and D , but inversely proportional to s .

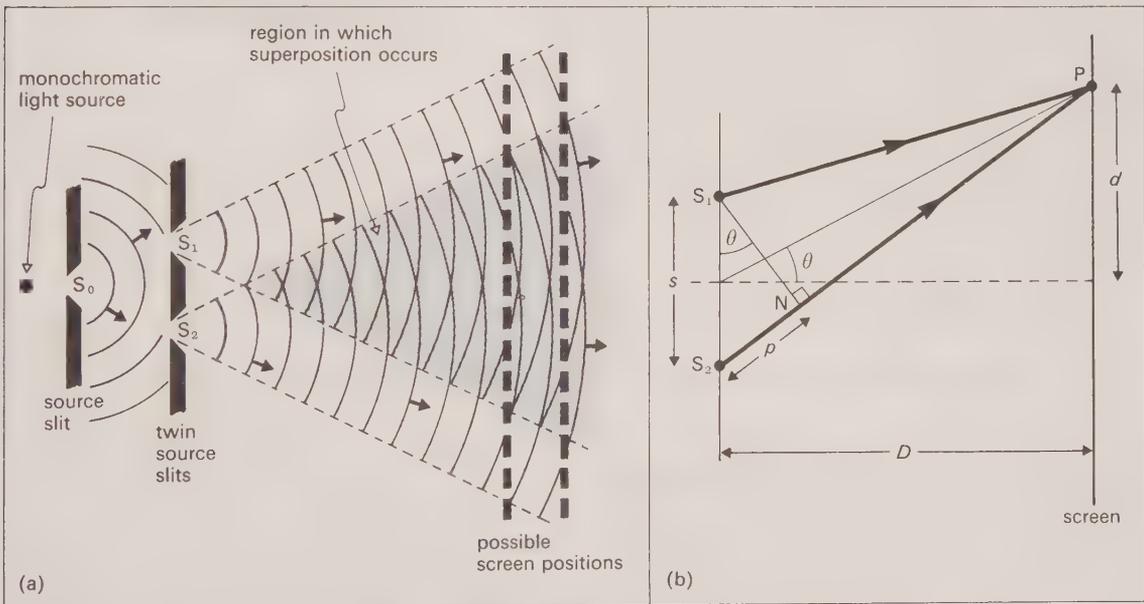


Fig. 38.8. The scheme of a Young's experiment, showing (a) why superposition occurs, and (b) the geometry to calculate the fringe separation. (In this and succeeding diagrams, the wavelength of light is grossly exaggerated. Since, in practice, $D \gg d$, S_1P can be considered parallel to S_2P .)

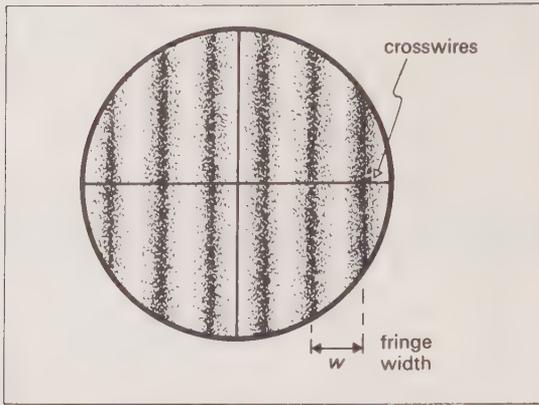


Fig. 38.9. View through an eyepiece placed at P in fig. 38.8.

Condition for dark fringes

Destructive superposition occurs where

$$p = \frac{ds}{D} = (m + \frac{1}{2})\lambda$$

so that the bright and dark fringes show the same spacing.

Measurement of wavelength λ

These orders of magnitude show why we can justify $\tan \theta \approx \sin \theta$.

$$\begin{aligned} \text{If } \lambda &\sim 5 \times 10^{-7} \text{ m} \\ s &\sim 1 \times 10^{-3} \text{ m} = 1 \text{ mm} \\ D &\sim 1 \text{ m} \end{aligned}$$

then the fringe width would be about

$$5 \times 10^{-4} \text{ m} = 0.5 \text{ mm.}$$

Measurement by travelling microscope of s and the fringe width enables λ to be calculated. The fringe width is measured by traversing ten or more fringes using a backlash-free screw.

The observer sees a field-of-view like fig. 38.9. To a first approximation the fringes are equally spaced.

The illumination varies across the field of view as shown in fig. 14.6, p. 114.

► 38.5 OTHER TWO-SOURCE EXPERIMENTS

The geometry of methods (A) and (B) below is the same as that for *Young's* method. The result

$$\text{fringe width } w = \frac{\lambda D}{s}$$

can be applied.

(A) Lloyd's mirror

The two sources are the slit S (parallel to the mirror) and its virtual image S' (fig. 38.10).

Notes

(a) If the screen is moved so that the point O touches the edge of the glass plate, the *geometrical* path difference for the two wavetrains is zero. The phase change of π rad on reflection at the denser medium then causes a *dark* fringe to be formed.

(b) This interference pattern is frequently seen in a ripple tank when one uses a wavetrain to demonstrate the law of reflection.

(c) Note the *real* waves propagating as though coming from the *virtual* source S'.

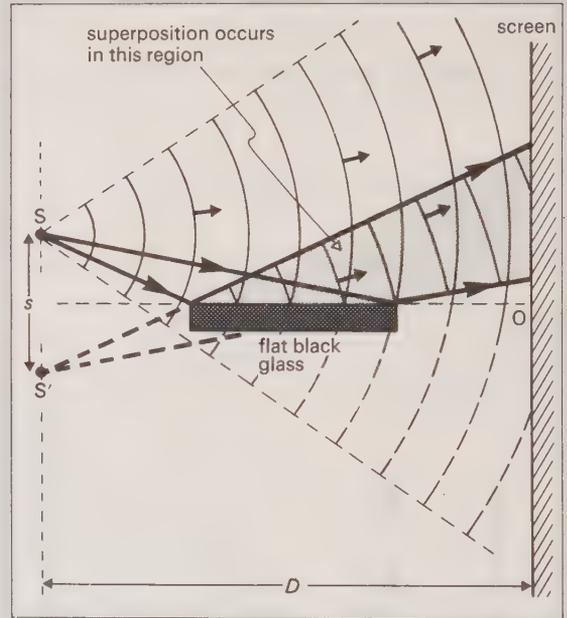


Fig. 38.10. Lloyd's mirror arrangement.

(B) Fresnel's biprism

The two sources are S₁ and S₂ (both being virtual images of the source slit S₀) formed by refraction in the two halves of the double prism (fig. 38.11).

The fringes (which are brighter than those from double slits) are observed using a travelling eyepiece in the plane of the screen.

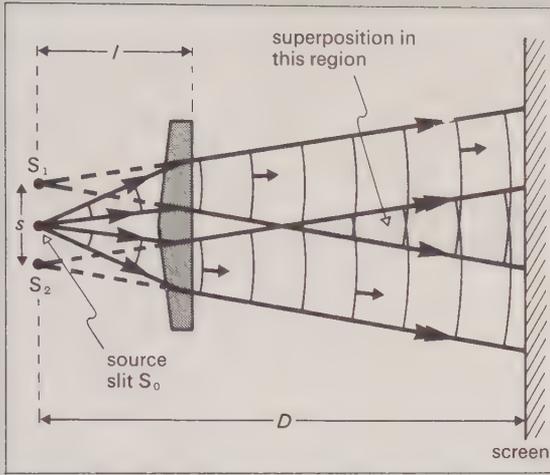


Fig. 38.11. Fresnel's biprism.

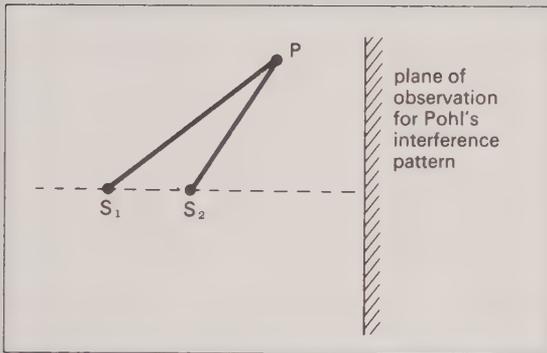


Fig. 38.12. Discussion of locus of a particular fringe.

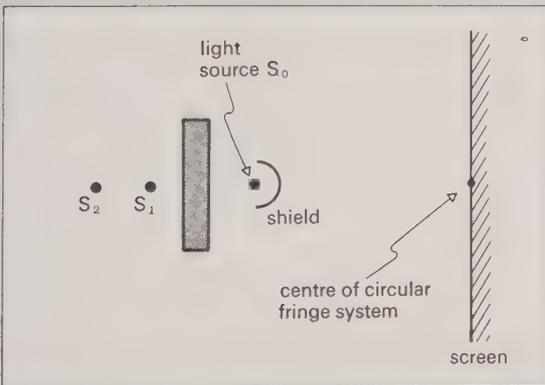


Fig. 38.13. The scheme of Pohl's interferometer.

Measurement of s

(1) Introduce a converging lens between prism and screen. Focus a real image of the two virtual sources. Note their separation s_1 , and the object and image distances a and b .

(2) Repeat with the lens in the conjugate position, and note the new measurements s_2 , b and a respectively.

From (1)
$$\frac{s}{s_1} = \frac{b}{a}$$

and from (2)
$$\frac{s}{s_2} = \frac{a}{b}$$

so
$$s = \sqrt{(s_1 s_2)}$$

Alternatively one can measure the angular separation θ of the virtual images (using parallel light from a spectrometer), and then apply $s = l\theta$ (fig. 38.11).

(C) Pohl's interferometer

In fig. 38.12 the point P lies on a particular fringe (bright or dark) if

$$(S_1P - S_2P) = \text{a constant}$$

The fringe system consists of a series of surfaces each being a hyperboloid of revolution about S_1S_2 as axis. A plane placed perpendicular to S_1S_2 intersects these surfaces in a series of concentric circles.

In Pohl's interferometer the two sources are the virtual images of a light source, formed by reflection at the two surfaces of a mica sheet about 10^{-5} m thick. A series of large concentric fringes can be produced on the screen of fig. 38.13.

38.6 WHITE LIGHT INTERFERENCE

In a two-source experiment like Young's, the fringe width $w \propto \lambda$.

When white light is used, different colours produce their own fringe systems which overlap (fig. 38.14, overleaf). At P the eye sees the overlapped fringes, which give general illumination, but analysis by a spectroscope would reveal that some colours had suffered destructive superposition.

The central white fringe is the only location at which all colours show zero optical path difference: it is invaluable for locating the centre of a fringe system, and so for measuring shifts of the whole system.

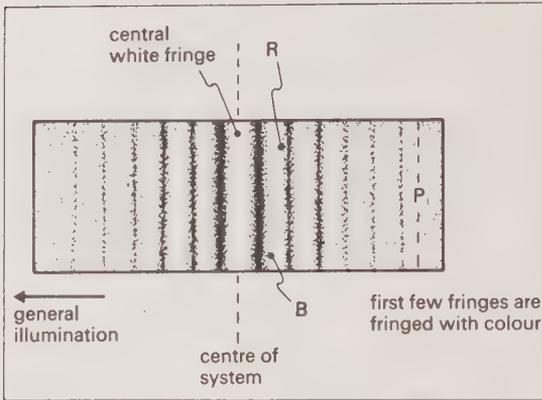


Fig. 38.14. Appearance at the centre of a Young's fringe system using white light.

Shifting a system of fringes

The central fringe of a ripple tank pattern is moved by making one source show a constant phase shift (say π rad) relative to the other. The same effect can be seen in a Young's two-slit experiment by introducing a different optical path length for one wavetrain while leaving the geometrical arrangement unchanged (fig. 38.15).

The optical path through the glass = nt
 The extra optical path introduced = $nt - t$
 = $(n - 1)t$.

This will cause the zero of the system to shift by

$$(n - 1)t \frac{D}{s}$$

since $d = \frac{pD}{s}$ (p. 295)

Measurement of this shift enables either n or t to be calculated.

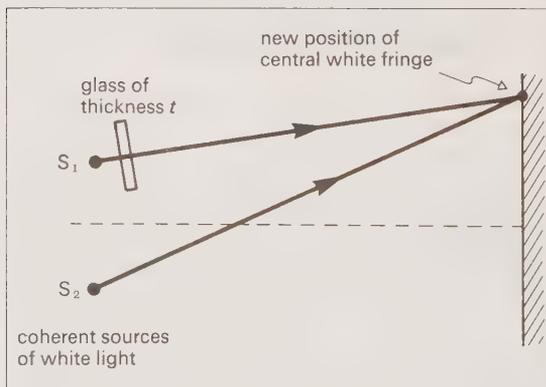


Fig. 38.15. Shifting a fringe system.

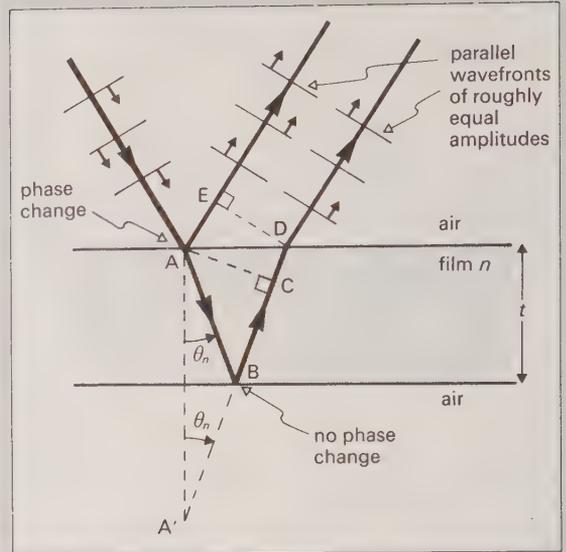


Fig. 38.16. Calculation of path difference for thin film interference using reflected light. (N.B. Light transmitted through the film has not been shown.)

► 38.7 PARALLEL-SIDED THIN FILMS

Fig. 38.16 shows an arbitrary ray of light incident on a film, and undergoing partial reflections and transmissions. We want to calculate the optical path difference for rays that travel

- (a) along the path AE in air, and
- (b) along the path ABCD through the film.

The parallel wavefronts formed by partial reflection will be coherent, and so will give interference fringes when brought together in the focal plane of the eye or some other converging lens.

The phase change of π rad at A is equivalent to an effective path difference $\lambda/2$.

The net optical path difference is then

$$n(AB + BC) + \frac{\lambda}{2} \quad \text{since } AE = n(CD)$$

By geometry

$$(AB + BC) = A'C = 2t \cos \theta_n$$

The net optical path difference is

$$2nt \cos \theta_n + \frac{\lambda}{2}$$

Constructive superposition occurs where

$$2nt \cos \theta_n + \frac{\lambda}{2} = m\lambda$$

and destructive superposition where

$$2nt \cos \theta_n + \frac{\lambda}{2} = (m + \frac{1}{2})\lambda$$

$$2nt \cos \theta_n = m\lambda$$

Observation of fringe systems

(A) Using a point source

In fig. 38.17 the eye is focussed to receive *parallel* rays reflected from the film surfaces. The image of the point source formed on the retina may be dark or bright, depending on θ_a (which controls θ_n).

Any interference pattern is only revealed *when the eye moves* so as to sample (in turn) a range of values for θ_a .

(B) Fringes using an extended source

If the source is extended, then an eye (or lens) at a given point receives rays leaving the film over a range of values θ_a (fig. 38.18). Because the film has parallel sides, $t = \text{constant}$. A particular dark fringe is seen along a direction where

$$2nt \cos \theta_n = m\lambda$$

so $\cos \theta_n = \text{constant}$

The fringes are called **fringes of equal inclination**: fig. 38.19 (overleaf) shows that they will appear as concentric circles about XX' as axis in the focal plane of the lens used to superpose the parallel rays. The fringes alter position as this lens is moved.

Miscellaneous effects

(A) Colours of thin films

Suppose that t is small, and that the incident light is white. Reinforcement occurs along directions satisfying

$$2nt \cos \theta_n + \frac{\lambda}{2} = m\lambda$$

When t is small one or two colours may reinforce along a direction in which others cancel. The film will then display brilliant colours, as does oil on water.

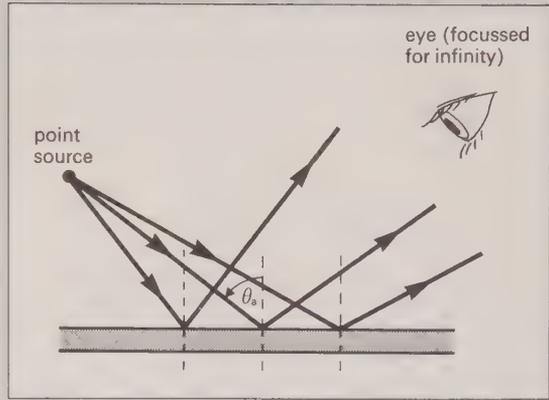


Fig. 38.17. Thin film effects using a point monochromatic source.

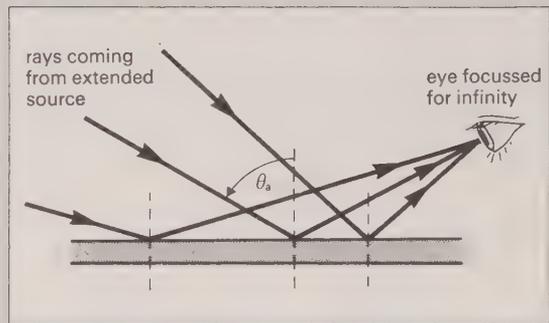


Fig. 38.18. Thin film effects using an extended monochromatic source.

(B) Thick films

When t is large, several values of λ satisfy the equation for a given value of θ_n . The film will appear to be generally illuminated.

(C) Blooming

Refer to fig. 38.20 (overleaf), which shows a coated lens.

Suppose $n_c t = \lambda/4$ where λ is the free-space wavelength for green light (chosen because it lies in the middle of the visible spectrum).

The reflected light undergoes destructive superposition because the net optical path difference is $\lambda/2$ for near normal incidence. The wasted reflected energy is reduced if the refractive index of the coating material is correctly chosen.

The lens appears purple, because the reflected light is deficient in green.

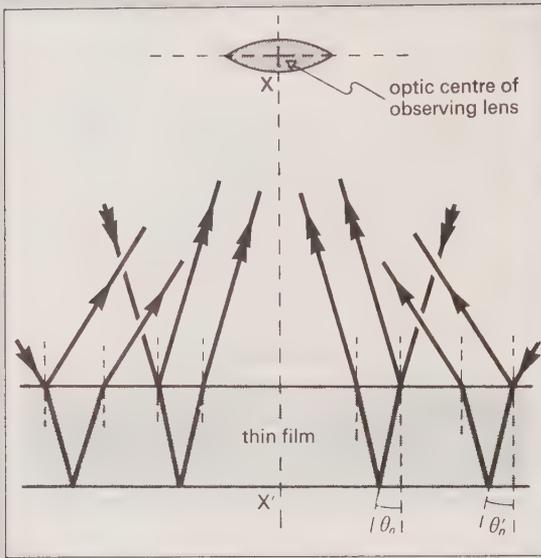


Fig. 38.19. Fringes of equal inclination appear as circles about XX' as axis.

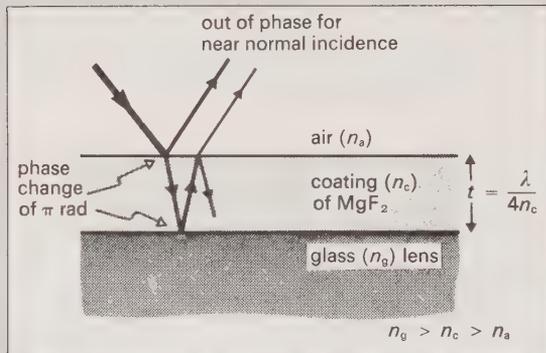


Fig. 38.20. Coating a lens can reduce wasted energy.

► 38.8 WEDGE FRINGES

In this section we again apply

$$2nt \cos \theta_n + \frac{\lambda}{2} = m\lambda$$

as the condition for brightness, but we consider *near-normal incidence*, so that

$$\theta_a = \theta_n = 0$$

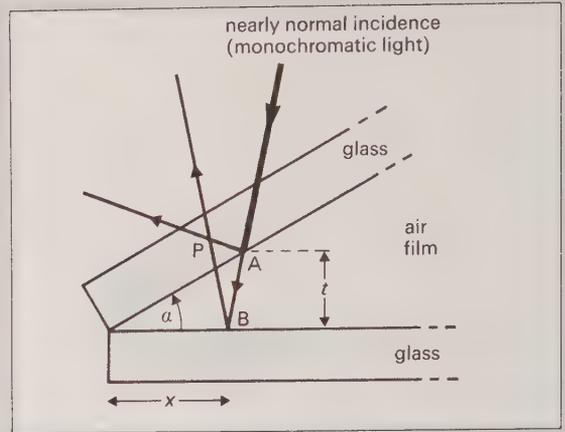


Fig. 38.21. Fringes of equal thickness formed in a thin air wedge ($\alpha \approx 10^{-4}$ rad, and is exaggerated).

A given fringe must then satisfy

$$t = \text{constant.}$$

The fringes are called **contour fringes**, or **fringes of equal thickness**.

In fig. 38.21, α is so small that refraction has been ignored. A phase change occurs at B, but not at A. The net optical path difference between rays superposed at P is

$$2t + \frac{\lambda}{2}$$

For bright fringes at P

$$2t + \frac{\lambda}{2} = m\lambda$$

in which

$$t = \alpha x$$

so

$$x = (m - \frac{1}{2}) \frac{\lambda}{2\alpha}$$

The fringe separation is given by

$$x_{m+1} - x_m = \frac{\lambda}{2\alpha}$$

A draining soap film

If the film drains uniformly, it will show horizontal bands in monochromatic light. Where $t \ll \lambda$ (near the top of the film), destructive superposition occurs for reflected light (because of the phase change). The film appears dark. By transmitted light it would appear bright.

Location of the fringes

Using a point source

The two surfaces each produce reflected wavefronts: their superposition could occur anywhere above the wedge, so the fringes would be *non-localized*.

Using an extended source

Each point of the source produces its own non-localized fringe system. The only place where the systems coincide for all such point sources is close to the plane of the wedge: *it is here that the fringes become localized*.

Use of wedge fringes

Using an experimental arrangement similar to that of fig. 38.21, one sees a series of parallel equally spaced straight fringes.

(a) Measurement of their separation allows α and t to be calculated. This method is adapted in Fizeau's apparatus for measuring distance changes of about 10^{-7} m.

(b) The regularity of the fringe system is a useful guide for the manufacture of an **optically flat** surface.

► 38.9 NEWTON'S RINGS

Newton's rings represent a system of contour fringes with radial symmetry (fig. 38.22).

Following section 38.7, the path difference for an air film thickness t_m is

$$2t_m + \frac{\lambda}{2}$$

For the m th dark fringe

$$2t_m + \frac{\lambda}{2} = (m + \frac{1}{2})\lambda$$

or

$$2t_m = m\lambda$$

From the intersecting chords of fig. 38.22(b),

$$r_m^2 \approx (2R - t_m)t_m$$

so

$$r_m^2 \approx 2Rt_m \quad \text{since } t_m \ll R$$

Then dark fringes occur where

$$\frac{r_m^2}{R} = m\lambda$$

When m is constant, r_m will be constant: the fringe system consists of *unequally spaced concentric circles*.

Since

$$2r_m \frac{dr_m}{dm} = \lambda R$$

$$\frac{dr_m}{dm} = \frac{\lambda R}{2r_m}$$

As r_m increases, the circles are closer together.

Notes

(a) For fringes viewed by *reflected* light, the *centre* of the fringe system is *dark* provided there

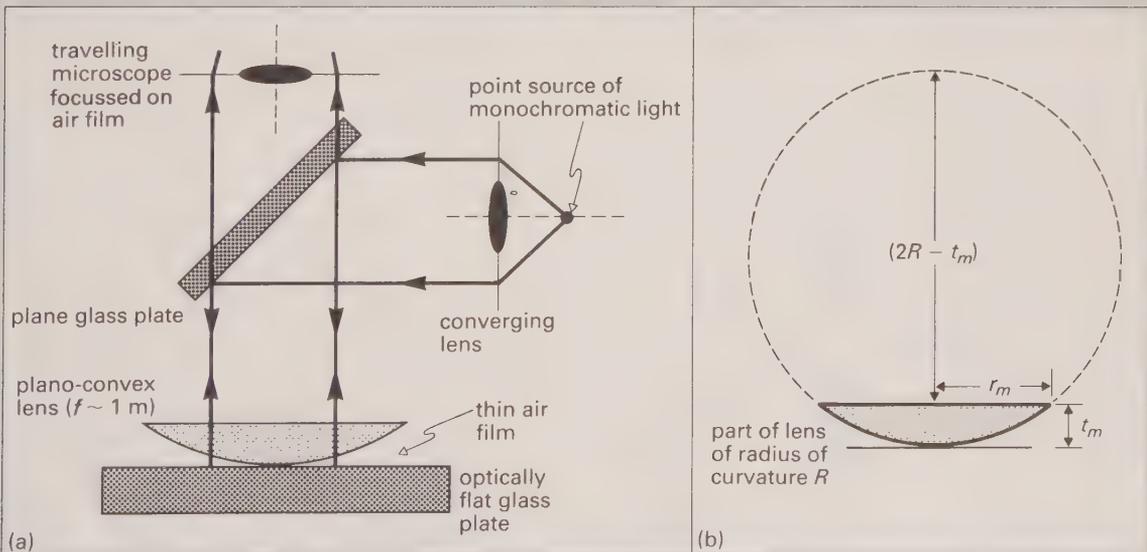


Fig. 38.22. Newton's rings: (a) experimental arrangement, (b) geometry.

is true contact between lens and plate. (A dust speck of thickness $\lambda/4$ could cause a central bright fringe.)

(b) Fringes formed by *transmitted light* are complementary (bright and dark exchange positions): the central fringe is bright. There is less contrast than in (a).

Measurement of λ

Suppose the m th ring has diameter d_m . Then

$$\frac{r_m^2}{R} = m\lambda$$

can be written $d_m^2 = (4R\lambda)m$

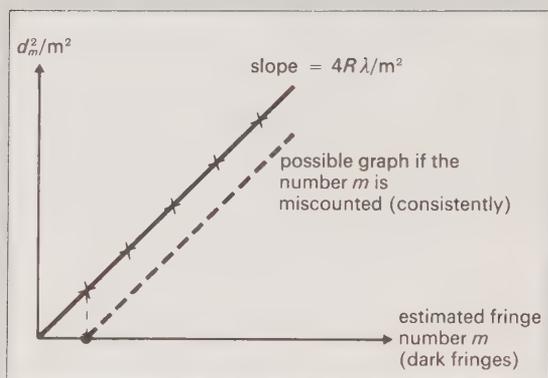


Fig. 38.23. Calculation of λ from Newton's rings measurements.

A graph (fig. 38.23) which plots d_m^2 against m has a gradient $(4R\lambda)$. The graph goes through the origin only if there is perfect contact between the lens and plate. Nevertheless the *slope* of the graph enables λ to be found, and the graph *will* be a straight line if the fringe number m is counted *consistently*.

d_m is measured by travelling microscope.

R is measured by a separate experiment.

► 38.10 SUMMARY OF THIN FILM EFFECTS

$2nt \cos \theta_n + (\lambda/2) = m\lambda$ is the condition for brightness.

Films of constant thickness	Films of varying thickness
Fringes are formed where $\theta_n = \text{const.}$, and are called fringes of equal inclination.	Fringes are formed where $t = \text{const.}$, and are called contour fringes.
θ_n varies from fringe to fringe.	In this book we assume near-normal incidence, i.e. $\theta_n = 0$.
Examples Colour of oil films Blooming of lenses	Examples Wedge fringes (e.g. a draining soap film) Newton's rings

39 DIFFRACTION

39.1	THE PHENOMENON OF DIFFRACTION	303
39.2	THE DIFFRACTION GRATING	304
39.3	GRATING SPECTRA	305
▶ 39.4	FRESNEL DIFFRACTION: THE NATURE OF SHADOWS	306

▶ 39.5	FRAUNHÖFER DIFFRACTION AT A SINGLE SLIT	309
▶ 39.6	RESOLVING POWER	310
▶ 39.7	X-RAY DIFFRACTION	311

39.1 THE PHENOMENON OF DIFFRACTION

If, when using *Huygens' construction* (p. 106) we assume the secondary wavelets to be effective only over the surface of tangency, then we predict *rectilinear propagation*. By introducing a factor $(1 + \cos \theta)$, where θ is the angle between the direction of wave travel and some other direction, we recognize that the amplitude is finite along other directions. We are thereby led to the possibility (within one medium) of a *change in direction* of the wave energy.

Diffraction is the result of superposing (light) waves from coherent sources on the same wavefront after the wavefront has been distorted by some obstacle.

By implication we superpose an *infinite* number of *infinitesimal* wavelets.

Typical diffracting obstacles include

- (a) lens edges and mirrors,
- (b) stops in optical instruments,
- (c) the source slit of a spectrometer.

These objects will be assumed to be *passive*. (An object which absorbs and re-radiates incident energy is said to **scatter** the light.)

The effects of diffraction become more obvious when the linear dimensions of the obstacle approach the wavelength of the wave motion. Most diffraction effects can only be observed with *small sources*. Each point on a large source produces its own diffraction pattern, but when these patterns overlap, only general illumination can be seen.

The principle of diffraction calculations

In fig. 39.1 the amplitude at P is found by superposing wavelets originating in the gap. It differs from what would be observed in the absence of the obstacle

- (a) because the obstacle removes some secondary wavelets from the wavefront, and
- (b) because the secondary wavelets originate at differing distances from P: this
 - (i) affects their relative amplitudes, and
 - (ii) controls their phase relationships.

In practice this calculation may become complex mathematically.

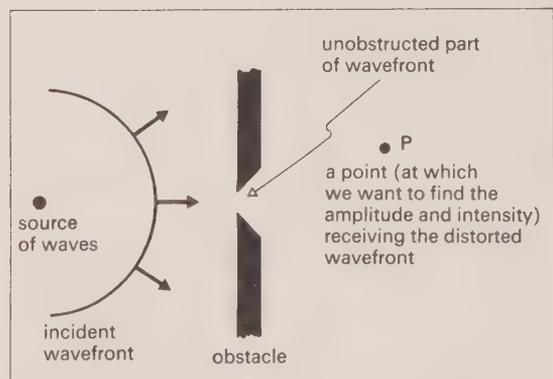


Fig. 39.1. Scheme for calculating the effects of a diffracting obstacle.

Classes of diffraction

Fresnel diffraction (fig. 39.2(a)) concerns *non-plane wavefronts*, so that the source S and/or the point P are to be a finite distance from the diffracting obstacle. Experimentally this is very simple, but analysis proves to be very complex.

Fraunhofer diffraction (fig. 39.2(b)) deals with *wavefronts* that are *plane* on arrival, and an effective viewing distance of *infinity*. Experimentally this is achieved by using lenses (fig. 39.2(c)). Mathematically the problem is simplified because we handle parallel rays.

Notes

(a) It follows that *Fraunhofer* diffraction is an important special case of *Fresnel* diffraction.

(b) When doing quantitative work on *Young's experiment* (p. 294), we assume the screen to be relatively distant, i.e. that we have *Fraunhofer* conditions.

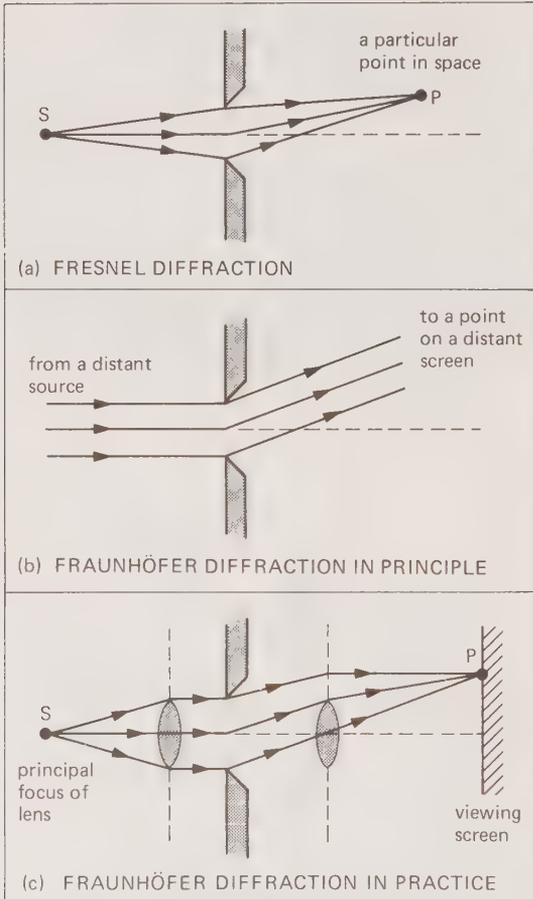


Fig. 39.2. A way of classifying diffraction.

39.2 THE DIFFRACTION GRATING

An optical transmission diffraction grating consists of a large number of fine equidistant parallel lines ruled onto a transparent plate (such as glass) so that light can pass between the lines, but not through them.

If each gap is sufficiently narrow (relative to λ), *diffraction* causes it to act as a centre of secondary disturbance. Nearly semicircular wavefronts (in section) are superposed on the far side of the grating (fig. 39.3).

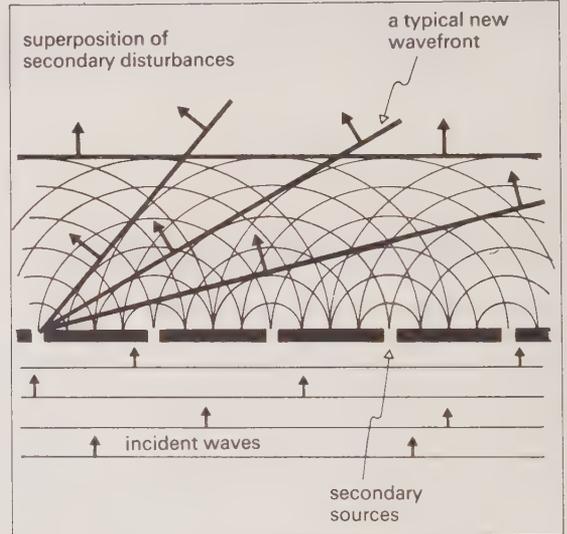


Fig. 39.3. Diffraction at the slits of an optical diffraction grating.

Along a few specified *directions* (rather than at some points) these secondary wavelets combine to produce a continuous wavefront. Along others their superposition is destructive, and no wavefront exists.

Fig. 39.4 shows rays of path difference p being brought to a point P by a lens. (All rays have the same optical path through the lens, so their phase relationships are unaltered.)

The superposition will be *constructive* (and result in brightness) when

$$p = m\lambda$$

where m is an integer, i.e. when

$$s \sin \theta_m = m\lambda$$

where s , the separation of the centres of neighbouring gaps, is called the **grating spacing**.

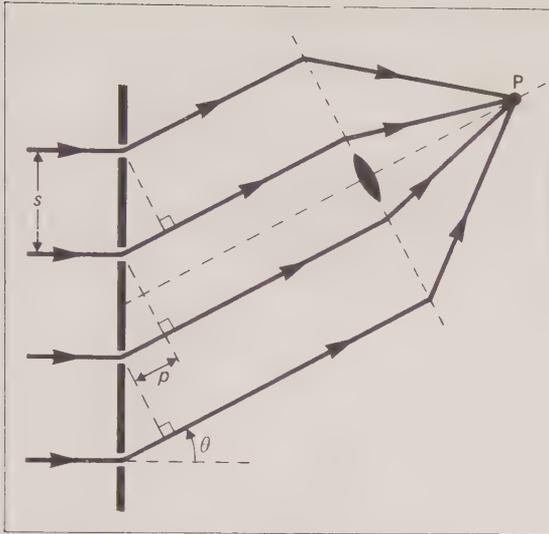


Fig. 39.4. Superposition of secondary wavelets brought together in the focal plane of a lens.

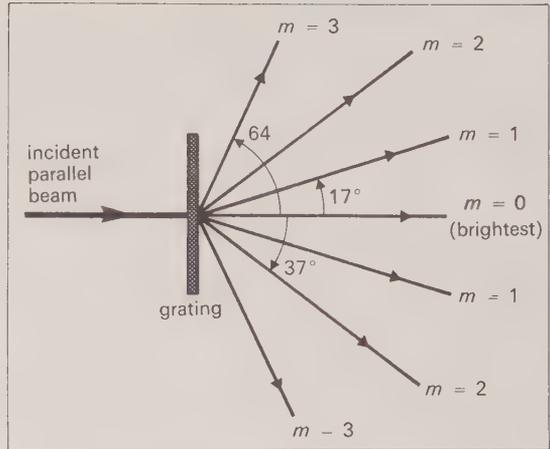


Fig. 39.5. Directions of reinforcement for a grating of $s = 2 \times 10^{-6}$ m using light of $\lambda = 6 \times 10^{-7}$ m.

Example

Suppose a grating has 5×10^5 lines per metre, so that $s = 2 \times 10^{-6}$ m, and that monochromatic light of wavelength 6×10^{-7} m is used. Then when

$$\begin{aligned} m = 0, & \quad \theta_0 = 0^\circ \\ m = 1, & \quad \theta_1 = 17^\circ \\ m = 2, & \quad \theta_2 = 37^\circ \\ m = 3, & \quad \theta_3 = 64^\circ \end{aligned}$$

but $m = 4$ predicts $\sin \theta_4 = 1.2$

so there is no value θ_4 or above.

This grating would show 7 directions of brightness (fig. 39.5). Along all other directions the secondary wavelets undergo destructive superposition, and effectively there would be darkness.

If white light is used, different wavelengths reinforce along different directions (fig. 39.6, overleaf), and so a spectrum is formed in the focal plane of the focussing lens.

When $m = 0$, all colours reinforce along $\theta_0 = 0^\circ$. The **zero-order spectrum**, or central maximum, will always be white when a white source is used. The **order** of a spectrum is the value of m .

Example

Using a grating of spacing $s = 2 \times 10^{-6}$ m, and taking

$$\begin{aligned} \lambda_V &= 4 \times 10^{-7} \text{ m} \\ \lambda_R &= 7.5 \times 10^{-7} \text{ m}, \end{aligned}$$

we have

$$\begin{aligned} (\theta_1)_V &= 12^\circ \\ (\theta_1)_R &= 22^\circ \end{aligned}$$

The angular dispersion of the first-order spectrum is

$$(22^\circ - 12^\circ) = 10^\circ$$

The second-order spectrum ranges from 24° to 49° (a dispersion of 25°). The angular dispersion increases with the order m (see fig. 39.7, overleaf).

Spectra of different orders can overlap. For this grating $(\theta_3)_V = 37^\circ$, and so the third-order overlaps the second by 12° . The part of the fourth order that is present overlaps the third.

39.3 GRATING SPECTRA

Since

$$\begin{aligned} \sin \theta_m &= \frac{m\lambda}{s} \\ \sin \theta &\propto \lambda \end{aligned}$$

for a given value of m .

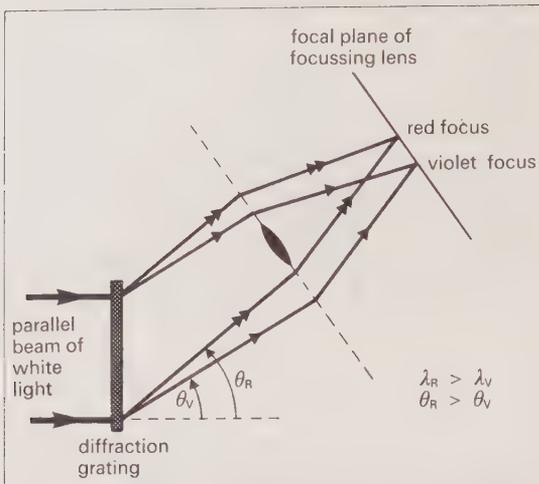


Fig. 39.6. A diffraction grating separates colours according to wavelength.

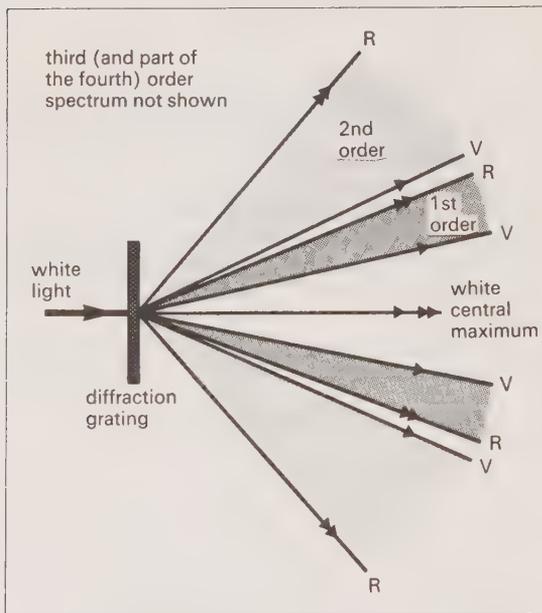


Fig. 39.7. Schematic representation of spectra from a grating.

Comparison of prism and grating spectra

The various wavelengths within an emission spectrum can be observed when the grating is mounted on a spectrometer table, as in fig. 39.8. An absorption spectrum can be studied by placing a sample of the gas to be investigated between the light source and the diffraction grating.

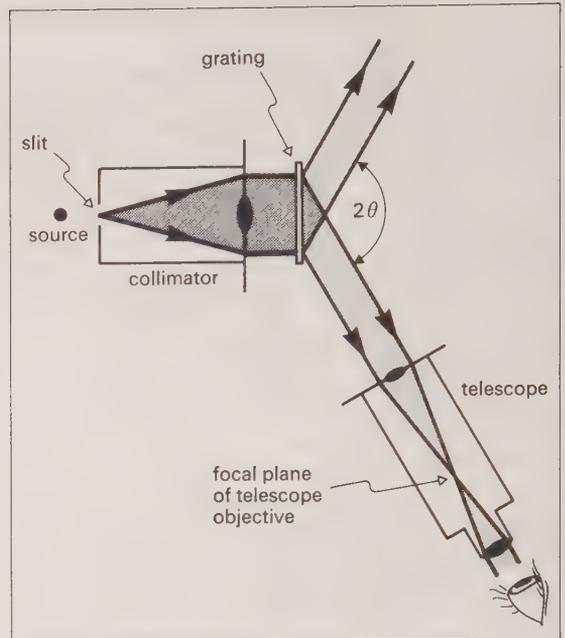


Fig. 39.8. A grating spectrometer.

(a) Since the energy from the slit is distributed over several spectra, their intensity is less than that of a prism spectrum. (Nevertheless it is possible to shape a groove profile so as to concentrate the light in a particular order of spectrum.)

(b) The wavelengths are distributed fairly evenly through the grating spectrum, whereas in the prism spectrum wavelengths are compressed at the red end, and extended at the violet.

(c) The **resolving power** of a spectrometer is its ability to distinguish very close wavelengths. It is generally greater for a grating than for a prism. It does not depend separately upon the size and spacing of the grating ruling, but it increases with the total number of rulings (i.e. *ceteris paribus* it increases with the overall width of the grating). It also increases with the order of the spectrum.

(d) Gratings can be made *reflecting* by ruling them on metal. If the metal is shaped as a concave mirror, lenses are not required, and it is possible to investigate wavelengths otherwise absorbed by glass.

► 39.4 FRESNEL DIFFRACTION: THE NATURE OF SHADOWS

These *Fresnel* patterns are fringed images of the obstructing obstacles. The angular relationships involved are *dependent on wavelength*, so that if

white light is used, a detailed analysis shows that the shadows have coloured edges.

(A) The straight edge (fig. 39.9)

The viewing screen shows a diffraction fringe system when viewed through a magnifying glass (fig. 39.11(a), overleaf). The intensity falls swiftly within the geometrical shadow. The actual location of a fringe is determined by an angular relationship, i.e. by *relative* distances.

(B) An opaque disc (fig. 39.11(b))

The exact pattern depends on the separation of the viewing screen from the disc, but there is always a *bright spot at the centre of the geometrical shadow*. The disc's shadow is surrounded by a complex of circular fringes of varying brightness.

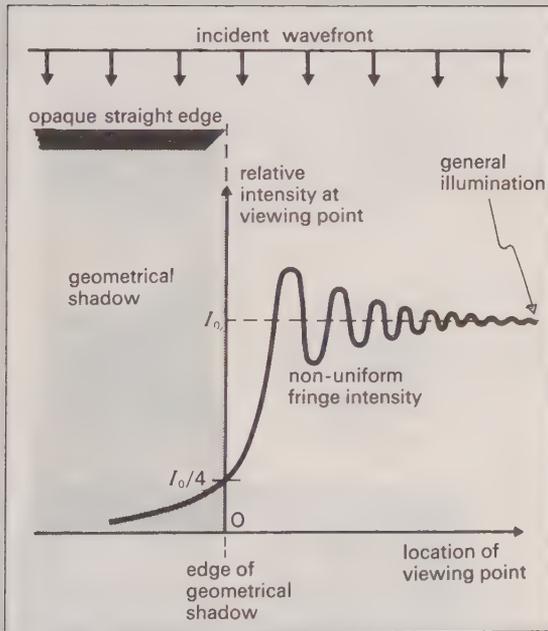


Fig. 39.9. Fresnel diffraction at a straight edge, showing the arrangement and the result.

(C) The circular aperture (fig. 39.10)

The general pattern (fig. 39.11(c)) is similar to (B), but the central spot is sometimes dark (depending on the viewing distance). At *large* distances we approach *Fraunhofer* patterns, and the central spot will be *bright* (fig. 39.14, p. 309).

(D) The long rectangular slit (fig. 39.11(d), (e), (f))

The diffraction effects are emphasized as the width a approaches λ . When $a \gg \lambda$, we have approximate rectilinear propagation. The *shadow* patterns of fig. 39.11 should be compared to the diffracted *wave* patterns shown on p. 117, in which details of fringes are not shown.

(E) The rectangular aperture (fig. 39.11(g))

Note that the pattern is extended along the direction of smaller aperture width ($a < b$).

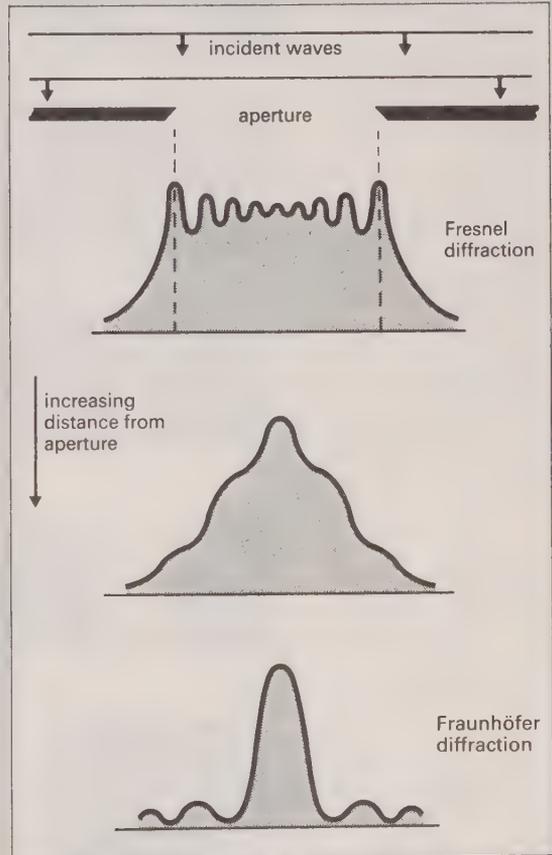


Fig. 39.10. Fresnel and Fraunhofer diffraction patterns for a circular aperture.

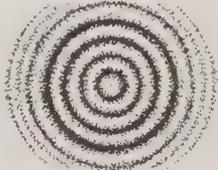
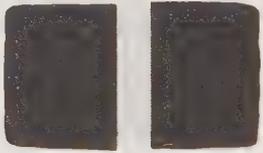
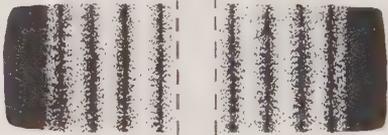
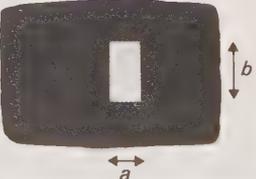
NATURE OF OBSTACLE	GEOMETRICAL SHADOW (BASED ON RECTILINEAR PROPAGATION)	DIFFRACTION PATTERN
(a) straight edge		
(b) opaque disc		 central white spot
(c) circular aperture		 central black spot (in this instance)
(d) long slit $a \sim \lambda$		energy reaches all parts of geometrical shadow aperture acts like a secondary source
(e) long slit $a \sim 3\lambda$		 geometrical shadow illuminated by bands of brightness
(f) long slit $a \sim 5\lambda$		approaching rectilinear propagation  geometrical shadow nearly in darkness
(g) rectangular aperture		

Fig. 39.11. Diffraction patterns compared with geometrical shadows.

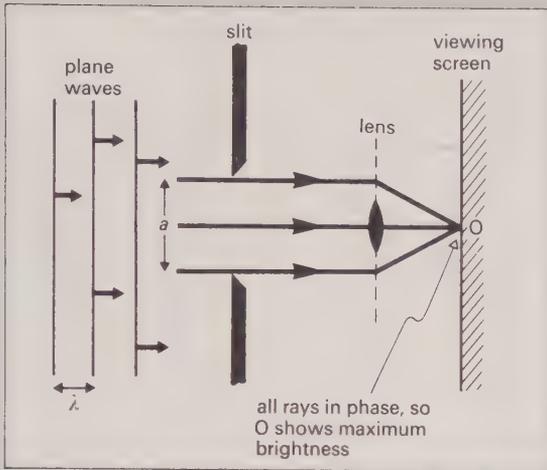


Fig. 39.12. Arrangement for showing Fraunhofer diffraction at a single slit.

► 39.5 FRAUNHÖFER DIFFRACTION AT A SINGLE SLIT

In fig. 39.12 we want to investigate the diffraction pattern on the screen. The point O will show that there is always a central maximum.

When locating the *direction* in which we expect to find the first minimum (fig. 39.13) we use the fact that the lens does not change the relative phases of the different rays that it brings to a point on the screen.

If $BD = \lambda/2$, then the two rays shown are out of

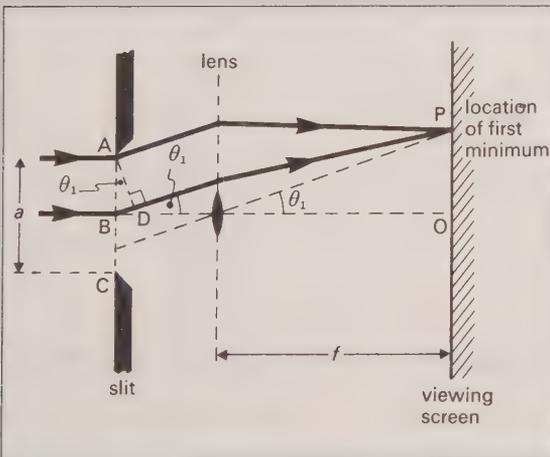


Fig. 39.13. Path differences calculated for pairs of rays diffracted through the slits.

phase, and their superposition at P gives zero amplitude.

Every ray that originates between A and B has a corresponding ray originating a distance $a/2$ below, between B and C, so that the *resultant* amplitude (and intensity) at P is zero.

θ_1 can be found from

$$\frac{BD}{AB} = \sin \theta_1$$

$$\frac{\lambda/2}{a/2} = \sin \theta_1$$

$$a \sin \theta_1 = \lambda$$

for the first *minimum*.

Note that this is an *angular* relationship. The distance OP can be found from

$$\frac{OP}{f} = \tan \theta_1$$

This argument can be extended to show that the angular displacements θ_m of the *minima* are given by

$$a \sin \theta_m = m\lambda$$

where $m = \pm 1, 2, 3$, etc. ($m = 0$ gives the central *maximum*.)

Maxima occur roughly halfway between the minima. Full analysis enables the relative intensities of the different maxima to be calculated. The result is given in fig. 39.14. For example,

- (i) when $a/\lambda \sim 5$, then $\theta_1 = 12^\circ$,
- (ii) when $a/\lambda \sim 10$, then $\theta_1 = 6^\circ$.

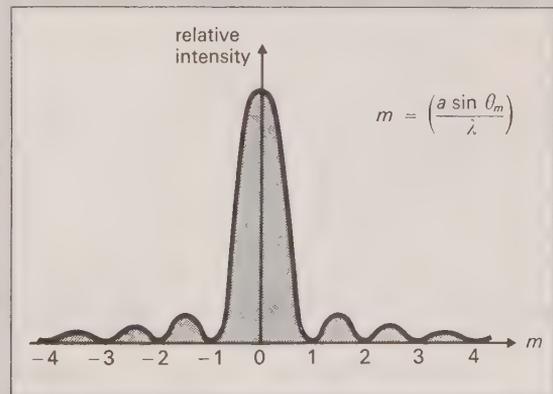


Fig. 39.14. Relative intensities of the different maxima. (Scale approximate only.)

Notes

(a) As a increases, so θ_1 decreases. Then the conditions approach more nearly those assumed in geometrical optics.

(b) As a decreases, so θ_1 increases. When $a = \lambda$, $\theta_1 = 90^\circ$; this means that the central maximum spreads fully into the geometrical shadow. We generally assume this to hold

(i) for the source-slit and twin-slits of a *Young's* experiment, and

(ii) for the diffraction grating.

(c) If θ_1 is small, we may approximate and write $\theta_1 \approx \lambda/a$ for the first minimum.

(d) The central bright fringe has several times the intensity and twice the width of any other.

The relation of this pattern to the *Fresnel* pattern is shown on p. 307.

► 39.6 RESOLVING POWER

The **angular resolving power** of an obstacle (which term includes a slit or an aperture) is the reciprocal of the minimum angle subtended by the directions of travel of incident waves coming from distant point sources such that we can distinguish their diffraction patterns. (See fig. 39.15.)

Rayleigh's criterion suggests that we agree to call the patterns distinguishable when the first minimum of one falls on the central maximum of the other (or better).

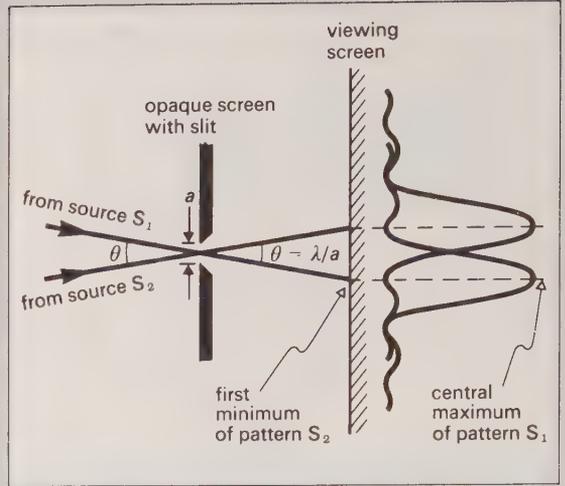


Fig. 39.15. Diffraction patterns distinguished according to Rayleigh's criterion.

Using this criterion, we find the following.

(a) The resolving power of a slit is found from

$$\theta = \frac{\lambda}{a}$$

We cannot distinguish the images of S_1 and S_2 if they subtend an angle less than λ/a . θ can be regarded as the minimum angle of resolution.

(b) The resolving power of a circular aperture can be calculated from

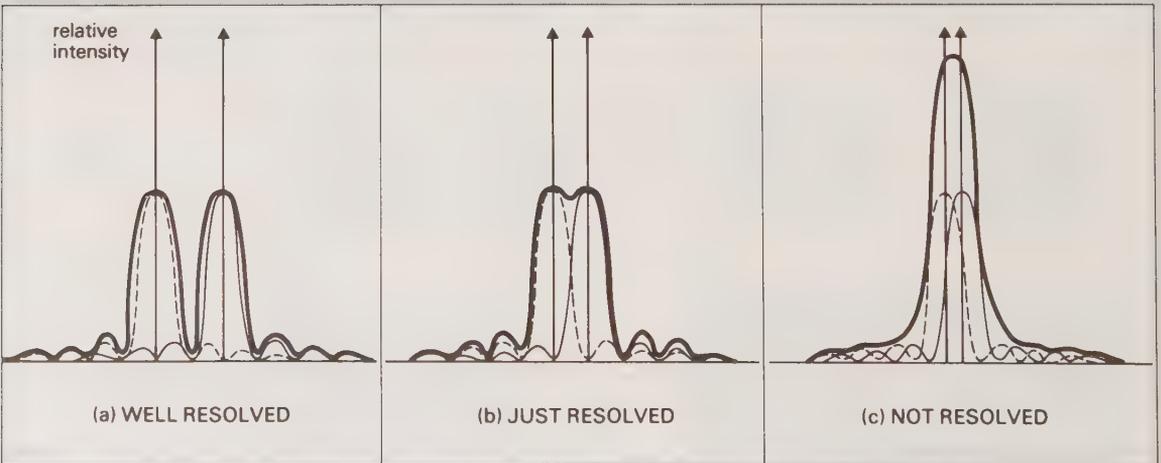


Fig. 39.16. Resolution of diffraction images of two slit sources.

$$\theta \approx \sin \theta = 1.22 \frac{\lambda}{D}$$

where D is the aperture diameter. The result is difficult to prove mathematically, but it is important when we consider optimum lens and mirror diameters for optical instruments such as telescopes (p. 280).

(c) The resolving power of the *eye* (when the pupil is small) is such that θ is about 10^{-4} rad of arc. This controls the minimum separation of points that are to be distinguishable: we can just resolve points 0.1 mm apart when they are 250 mm from the eye. (This resolving power is matched by the structure of the retina, which is made of discrete rods and cones.)

Fig. 39.16 shows different degrees of resolution

of diffraction images from two slit sources—the intensities from the separate patterns combine to produce the resultant pattern in each case.

► 39.7 X-RAY DIFFRACTION

See p. 507 for the production and properties of X-rays.

Von Laue's experiment (1912)

The practical arrangement is described in fig. 39.17. The experiment established the wave nature of X-rays. At this time it confirmed that they could not be matter (earlier than *de Broglie*) and must be electromagnetic radiation. The very small grating

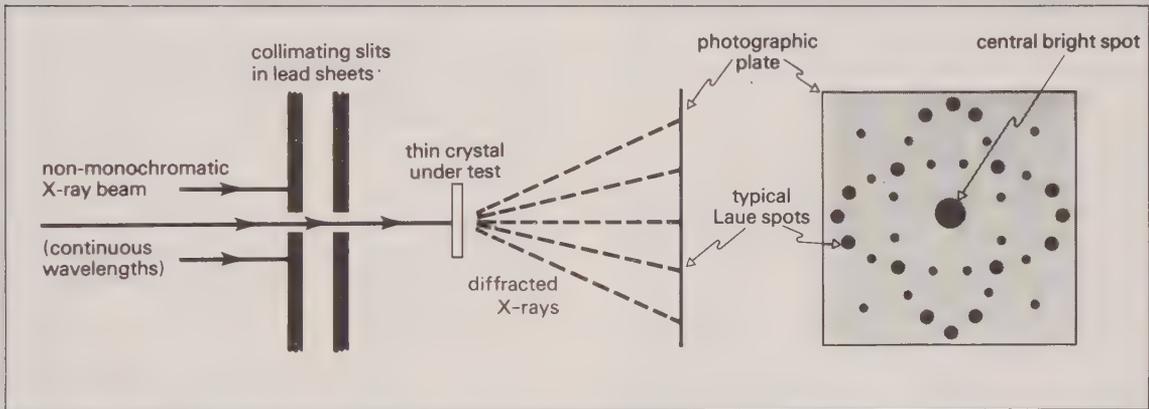


Fig. 39.17. Von Laue X-ray diffraction.

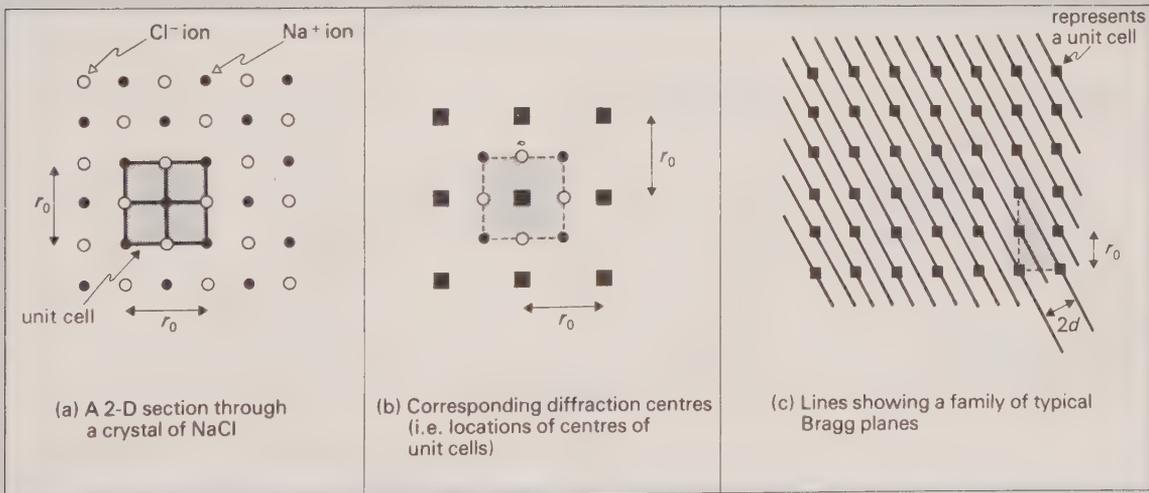


Fig. 39.18. Unit cells and reflecting planes for a crystal of NaCl.

spacing demanded by the very small wavelength of the radiation could not then be ruled mechanically. The crystal's periodic three-dimensional spacing of ions acts as a 3-D transmission diffraction grating. Thus the many diffracting centres produce constructive superposition in certain sharply-defined directions. The angles of diffraction will be large if the wavelength is comparable with the repetition length r_0 .

Interpretation of the transmission pattern is not simple. Positions and intensities of *Laue spots* enable atomic arrangements within the crystal to be deduced. The fundamental repetitive diffracting unit is called the **unit cell**—for the NaCl crystal it has four sodium ions and four chloride ions associated with it.

(a) Directions of diffracted X-ray beams indicate basic symmetry of unit cells in the crystal.

(b) Intensities show the distribution of electrons within the unit cell—it is the electrons rather than the nuclei that diffract the X-rays.

Unit cells and reflecting planes

Fig. 39.18 (p. 311) shows a two-dimensional representation of an NaCl crystal, its diffraction centres and a possible family of planes passing through these centres.

When X-radiation falls on a crystal, each unit cell acts as a diffracting centre. Superposition of the scattered secondary wavelets from those unit cells lying in a particular family of planes produces an apparently reflected beam which can be interpreted quite easily. The perpendicular spacing of the planes is closely related to the unit cell dimension—in fig. 39.18(c) the spacing $d = r_0/\sqrt{5}$.

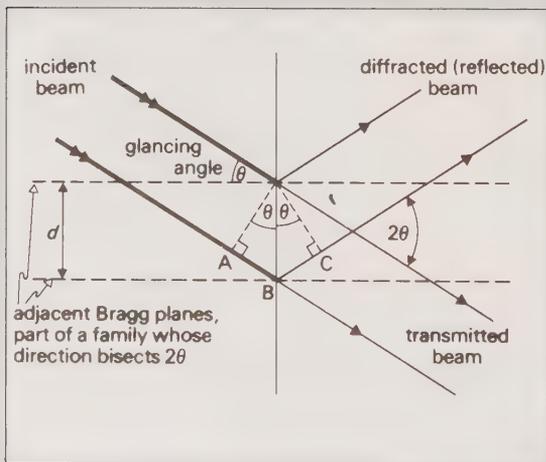


Fig. 39.19. Derivation of the Bragg equation.

The Bragg equation

(a) A particular surface plane always acts like a mirror, giving a weak reflected beam for any angle of incidence.

(b) An entire family of planes will produce a reflected beam by constructive superposition only when the path difference from adjacent planes obeys the proper condition.

In fig. 39.19, the path difference for adjacent planes is $AB + BC = 2d \sin \theta$. For constructive superposition this must equal $m\lambda$.

$$2d \sin \theta = m\lambda$$

W. L. Bragg's Law

$m = 1, 2, 3, \text{ etc.}$, the order of diffraction.

(a) For monochromatic X-rays incident at arbitrary θ , no (reflected) diffracted beam appears, and the beam is totally transmitted, but

(b) for beams of a mixture of X-ray wavelengths, diffracted beams appear when

$$\lambda = 2d \sin \theta / m$$

is satisfied.

Bragg's law can be applied to the expression for the *de Broglie* wavelength discussed in electron diffraction (p. 288).

$$2d \sin \theta = m\lambda = \frac{mh}{\sqrt{2Vem_e}}$$

This equation can be used to confirm *de Broglie's* theory.

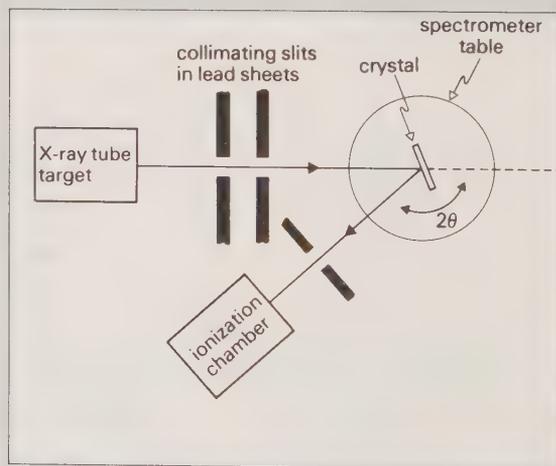


Fig. 39.20. The X-ray spectrometer.

The X-ray spectrometer

The practical arrangement is shown in fig. 39.20. The wavelength can be measured using a crystal of known unit cell dimensions. The value of λ can then be used to measure d and r_0 , and so to analyse the nature of the atomic pattern in crystals, and the structure of complex substances. The ion-

ization chamber and crystal are rotated to keep the angle of incidence equal to the angle of 'reflection'. The intensity of the diffracted X-rays is deduced from the ionization current.

In the school laboratory analogous X-ray diffraction experiments can be carried out using 30 mm electromagnetic waves and an arrangement of polystyrene spheres or tiles.

40 POLARIZATION

40.1	LIGHT AS A TRANSVERSE WAVE MOTION	314
40.2	MALUS'S EXPERIMENT AND BREWSTER'S LAW	315
40.3	PRODUCTION OF PLANE-POLARIZED LIGHT	316

▶ 40.4	SCATTERING AND POLARIZATION	317
40.5	DETECTION OF PLANE-POLARIZED LIGHT	318
40.6	USES OF POLARIZATION	318

40.1 LIGHT AS A TRANSVERSE WAVE MOTION

The term **transverse** was defined on p. 96. The ideas of wave polarization were discussed on p. 102, and polarization in optics was mentioned on p. 283.

Fig. 37.1 showed a propagating electromagnetic wave in which the plane containing the electric vector is fixed: it is therefore **plane-polarized**. For historical reasons (p. 316) the **plane of polarization** was chosen to be that *perpendicular* to the electric vector.

The **plane of vibration** is that which contains the *electric* vector and the direction of propagation. We define it relative to the electric field, since it is E that enables a force to be exerted on a *stationary* charge, and thus which is primarily responsible for the interaction of electromagnetic waves and matter.

Light is said to be **unpolarized** if the plane of vibration and the magnitude of E continually and randomly change.

Examples

(a) Many radio waves, radar waves and the microwaves we use for laboratory demonstrations are plane-polarized as a result of the nature of the source (fig. 40.1(a)).

(b) Most visible light, being emitted *randomly* by atoms and molecules which are only a small part of a light source, is unpolarized (fig. 40.1(b)).

Fig. 40.1(a) could be demonstrated by rotating a detector of microwaves through an angle of 90° . If it previously registered a maximum signal, it would now show zero (and vice versa).

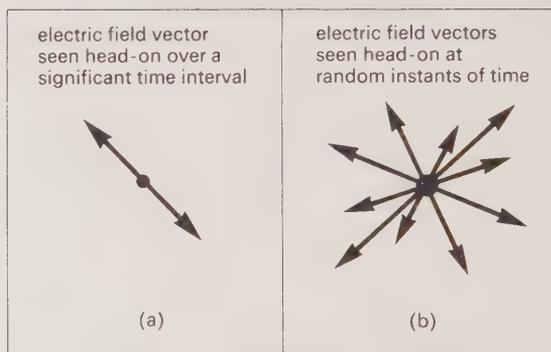


Fig. 40.1. Polarization in electromagnetic radiation: (a) polarized radiation, (b) unpolarized radiation.

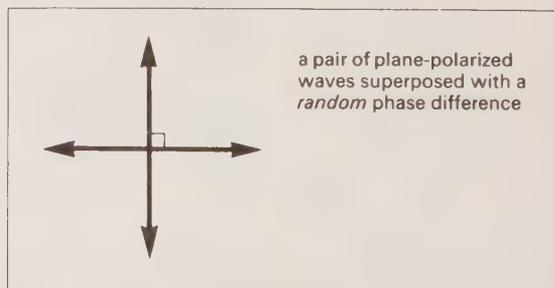


Fig. 40.2. An equivalent representation of fig. 40.1(b).

Fig. 40.2 shows an alternative way of representing fig. 40.1(b). At any given instant of time, the particular resultant of fig. 40.1(b) could be resolved into its resolved parts along the directions shown.

40.2 MALUS'S EXPERIMENT AND BREWSTER'S LAW

Malus's experiment (1809)

In fig. 40.3, an observer rotating the plate AB about an axis CD views the parallel beam reflected from the plate.

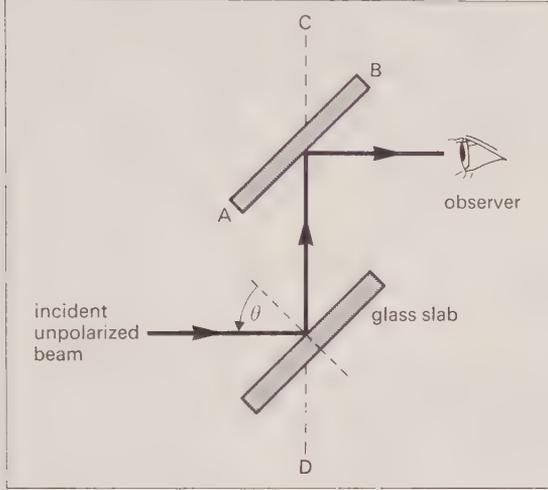


Fig. 40.3. Malus's experiment.

- (a) When the beam is reflected to left or right (in the plane of the paper) the viewed intensity is a maximum.
- (b) When the beam is reflected into or out of the paper normally, the intensity is a minimum.
- (c) When θ takes a value θ_p , the **polarizing angle**, the minimum intensity becomes zero.

We deduce that when $\theta = \theta_p$, the light travelling from the lower to the upper plate has become completely plane-polarized by reflection.

Explanation

Fig. 40.4 shows the electric field radiated from a dipole oscillator.

- (a) E reaches a maximum along the *transverse* dipole axis.
- (b) E is zero along the longitudinal dipole axis.

We conclude that since intensity $\propto E^2$, no energy is radiated along the direction of oscillation.

When a linearly-polarized ray (of correct orientation) is incident at the polarizing angle (fig. 40.5) there is *no reflected ray*: the oscillators set into motion by the incident ray cannot send energy along the direction of the usual reflected ray if it would be perpendicular to the refracted ray.

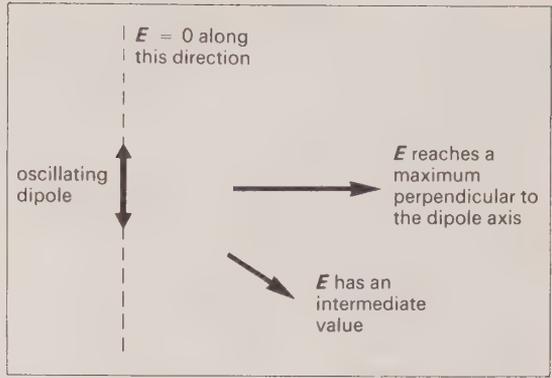


Fig. 40.4. Two-dimensional diagram of radiation from an oscillating electric dipole.

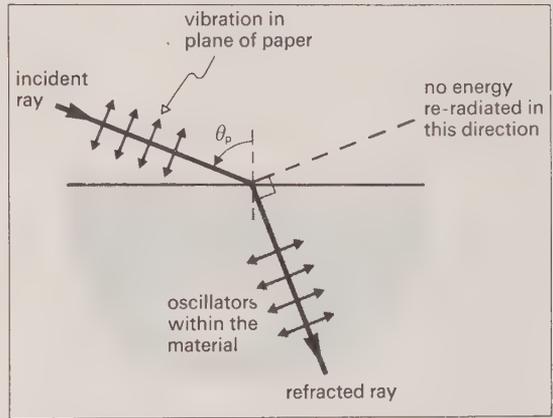


Fig. 40.5. When the angle of incidence equals θ_p the refracted ray is perpendicular to the usual direction of the reflected ray.

Brewster's Law (1812)

Fig. 40.6 (overleaf) shows unpolarized light incident at the polarizing angle. Since the reflected ray is completely plane-polarized

$$\theta_1 + \theta_2 = 90^\circ$$

By Snell's Law

$$\begin{aligned} n_1 \sin \theta_1 &= n_2 \sin \theta_2 \\ &= n_2 \cos \theta_1 \end{aligned}$$

where $\theta_1 = \theta_p$.

$$\therefore \frac{n_2}{n_1} = \tan \theta_p$$

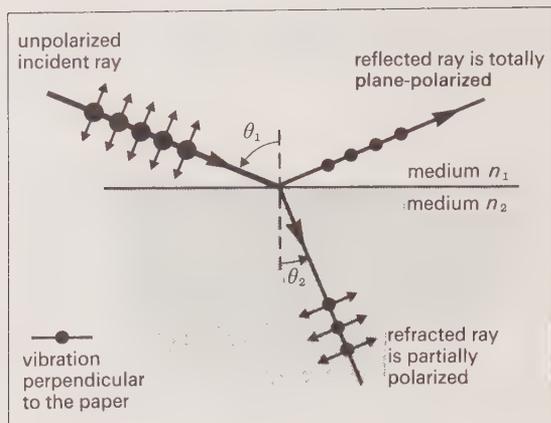


Fig. 40.6. Derivation of Brewster's Law.

If medium 1 is air or vacuum

$$n_2 = \tan \theta_p \quad \text{Brewster's Law}$$

Notes

- (a) Brewster discovered the law experimentally.
- (b) The reflected beam is *always* partially plane-polarized ($\theta_1 \neq 0$).
- (c) For glass of $n = 1.5$, $\theta_p = 57^\circ$.

Since θ_p depends upon n , we need to use a parallel *monochromatic* beam for accurate measurements.

(d) The plane containing the incident and reflected (plane-polarized) rays of fig. 40.6 was chosen *historically* to be called the **plane of polarization**.

40.3 PRODUCTION OF PLANE-POLARIZED LIGHT

Fig. 40.7 illustrates the function of a **polarizer**.

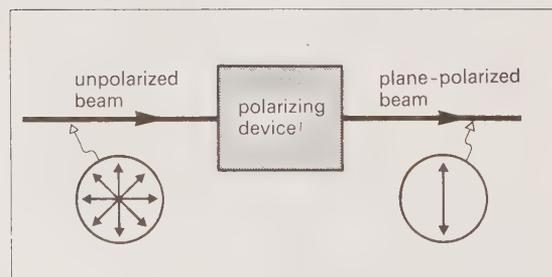


Fig. 40.7. Function of a polarizer.

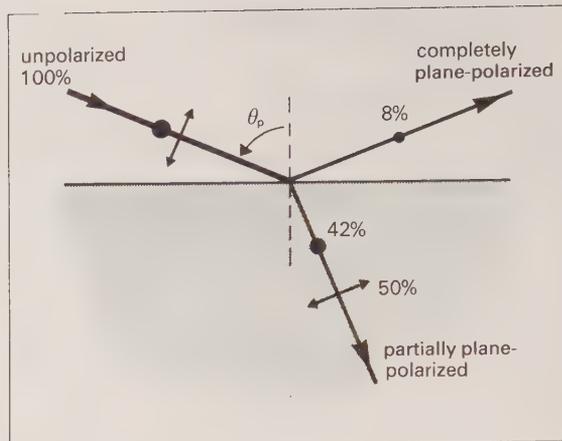


Fig. 40.8. Proportion of energy taken by reflected and refracted rays.

(A) Using reflection

Fig. 40.8 shows the possible distribution of energy from the incident beam.

(B) Using normal refraction

Fig. 40.8 illustrates how a transmitted ray is *partially* plane-polarized. Fig. 40.9 shows how *successive* refractions can produce a transmitted ray that is *nearly* plane-polarized.

(C) Using selective absorption

The microwave analogy

Fig. 40.10 shows how completely polarized microwaves are absorbed by conducting wires *parallel* to their plane of vibration, but transmitted by wires perpendicular to that plane. During absorption the microwave energy is used to set the wire's electric charge into oscillation along the length of the wire.

Dichroism

Materials which transmit one component of polarization but attenuate the mutually perpendicular component exhibit **dichroism**. *Polaroid* is a commercial material which contains long molecules whose axes have been aligned during manufacture. They have a similar effect on light waves to that of conducting wires on microwaves (figs. 40.10 and 40.11). *Tourmaline* is a crystal which occurs naturally, and which has the same effect.

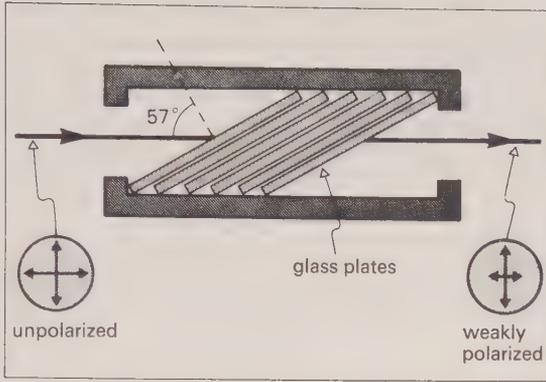


Fig. 40.9. An attempt to produce plane-polarized light by successive refractions.

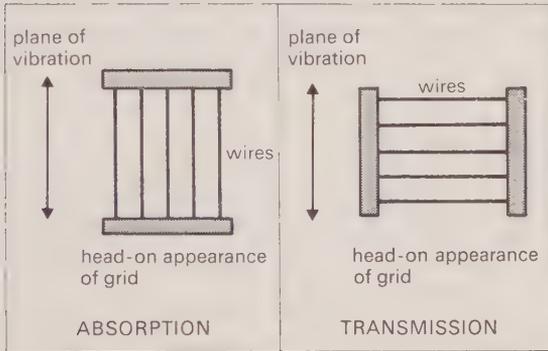


Fig. 40.10. Demonstration of selective absorption of polarized microwaves by parallel wires.

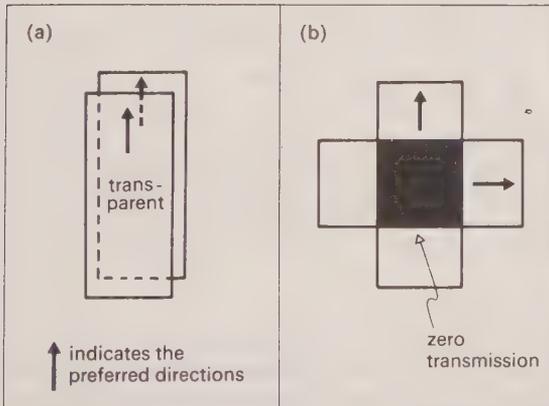


Fig. 40.11. Polaroid sheets with preferred directions (a) parallel, allowing transmission, and (b) crossed at 90° causing extinction of unpolarized light.

(D) Using double refraction

Many crystals show **double refraction** or **birefringence**: an unpolarized ray which enters the crystal is split into an **ordinary ray**, and an **extraordinary ray** (which does not obey *Snell's Law*). Each ray is plane-polarized, their planes of vibration being perpendicular.

A calcite crystal (which is cut correctly and then reconstructed) can be used to separate the ordinary ray from the extraordinary, and the plane-polarized extraordinary ray is then transmitted alone. This is the principle of the **Nicol prism**, a device which has the same effect as a *Polaroid* sheet (quinine iodosulphate crystals embedded in a plastic material).

► 40.4 SCATTERING AND POLARIZATION

Electromagnetic radiation energy is absorbed by an atom or molecule when the E -field of the wave acts on the electrons causing them to become forced oscillating electric dipoles with their axes along the direction of the *field*. The re-radiation of the electromagnetic wave in all directions is known as **scattering**, and results in a reduction of intensity from the initial radiation direction (p. 247). Fig. 40.4 shows that, if the incident radiation is linearly *polarized*, the intensity of the scattered wave is a maximum along the transverse dipole axis, and that it is zero along the longitudinal axis where the scattering angle is 90° . If the incident radiation is *unpolarized*, the dipole will oscillate in numerous different directions. The re-radiation is

- (a) completely polarized when scattered through 90° ,
- (b) completely unpolarized in the direction of the incident wave, and
- (c) partially polarized in all other directions.

The natural oscillation frequencies within a typical atom or molecule are in the u.v. region of the electromagnetic spectrum. The greater the frequency of waves in the visible region the closer they become to the u.v. resonant frequency, and hence the larger the amplitude of the forced oscillations. Blue light is therefore scattered more effectively than red light. This contributes to the blue colour of the sky and to the redness of the Sun at sunrise and sunset. If the Earth were in a vacuum, the sky would appear black as it did to the cosmonauts on the Moon. Smoke or dust particles and water droplets also cause scattering, as does a colloidal suspension of particles in a liquid (**Tyndall effect**).

40.5 DETECTION OF PLANE-POLARIZED LIGHT

An **analyser** is a device which can detect polarized light. It has the same effect on the light as a polarizer and is constructed in the same way.

Consider (fig. 40.12) plane-polarized light incident on an analyser. The component of E_0 that is transmitted is $E_0 \cos \theta$. Since intensity

$$I \propto E^2 \quad (\text{p. 104})$$

the transmitted intensity

$$I \propto E_0^2 \cos^2 \theta$$

or

$$I = I_0 \cos^2 \theta \quad \text{Malus's Law (1809)}$$

This equation should be related to *Malus's* experiment on p. 315.

If an analyser is rotated in a plane-polarized beam, the observer sees two maxima and two minima (zero) for each rotation.

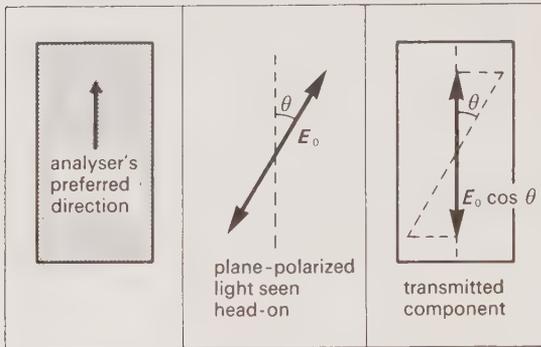


Fig. 40.12. Analysis of plane-polarized light.

40.6 USES OF POLARIZATION

(A) Academic

(a) Polarization is the only wave property of those on p. 99 which can be used to distinguish *transverse* from longitudinal waves. *Arago* and *Fresnel* (1815) had shown that the two beams emerging from a doubly-refracting crystal, although otherwise coherent, did not give interference effects. *Young* (1817) suggested that this

resulted from a lack of symmetry: this is only possible if the quantity associated with the periodic fluctuations is a *transverse vector*.

(b) Polarized light is very useful for general experiments on the *coherency* of different light beams. For example, two beams of plane-polarized light which were originally derived from perpendicular components of unpolarized light, and which are later rotated into the same plane, *cannot* be made to exhibit typical interference fringes.

(c) The phenomenon can be used to show that the *electric vector* is largely responsible for the interaction of electromagnetic waves and matter.

(B) Applications

Reduction of glare

Sunglasses and camera filters are made of polarizing materials to reduce the glare of light polarized by reflection from shiny surfaces (such as water).

Solar compass

The blue light scattered from gas molecules in the sky is partly plane-polarized. In polar regions, where a magnetic compass becomes useless, one uses a solar compass, which analyses the polarization of scattered sunlight.

Bees have eyes which are polarization-sensitive: these enable them to orientate themselves.

Optical activity

Many crystals (such as quartz) and some liquids (especially sugar solutions) have the ability to cause rotation of the plane of vibration of plane-polarized light.

The angle of rotation α for solutions depends on

- (a) the concentration, and
- (b) the length of solution through which light passes.

A **polarimeter** is a device which measures α , from which the concentration of the solution can be found. **Optical activity** is the name given to the phenomenon.

Photoelasticity

Glass and some plastic materials become doubly refracting only while stressed. If polarized white light is then passed through them, and analysed, bright coloured lines show the existence of strains. In engineering work plastic models of structures are constructed, and weaknesses are exposed in this way.

Kerr cell

This is a light shutter with no mechanical parts, whose working is based on the **Kerr electro-optic effect** (1875). (A material which is usually **isotropic** may become anisotropic when subjected to a strong electric field.)

Comparison of light intensities

This can be done by using

$$I = I_0 \cos^2 \theta \quad (\text{Malus's Law})$$

if one measures θ .

Shareed Al.

VIII SOUND WAVES

41 SOUND WAVES

41.1 CLASSIFICATION OF SOUND WAVES	322	41.7 SPEED OF SOUND BY A DIMENSIONAL ARGUMENT	327
41.2 DISPLACEMENT, VELOCITY, EXCESS PRESSURE AND DENSITY	322	41.8 VARIATIONS IN THE SPEED OF SOUND	328
41.3 FACTS ABOUT SOUND WAVES	323	41.9 VIBRATION OF AIR COLUMNS	329
41.4 NOISE AND MUSICAL NOTES	325	41.10 VIBRATION OF STRINGS	331
41.5 GENERAL PROPERTIES OF SOUND WAVES	325	41.11 MEASUREMENT OF c, f AND λ	331
41.6 PROPERTIES BASED ON SUPERPOSITION	326	41.12 MEASUREMENT OF c IN A METAL ROD	332
		41.13 SOUND RECORDING AND REPRODUCTION	333

41.1 CLASSIFICATION OF SOUND WAVES

Refer to the wave classification on p. 98. Sound waves

- (a) (i) can be *periodic*, e.g. those emitted by a tuning fork, or
- (ii) *non-periodic*, e.g. the shock wave emitted by an aircraft which breaks the sound barrier
- (b) (i) can be *progressive*, e.g. tuning fork waves, or
- (ii) *stationary*, e.g. the sound waves inside an organ pipe
- (c) are *longitudinal*, and
- (d) the *disturbance* is a variation of the pressure and density of the medium through which the wave passes. The disturbance is thus a *scalar* quantity.

(d) above indicates that a *medium* is necessary for the transmission of the waves, which (unlike electromagnetic waves) cannot travel through a vacuum. This medium can be solid, liquid or gas.

The *source* of a sound wave is an object which vibrates so as to send out waves through the medium (frequently air) in which it is immersed.

41.2 DISPLACEMENT, VELOCITY, EXCESS PRESSURE AND DENSITY

When a longitudinal wave disturbs a particle of the medium through which it passes, an oscillation is superimposed on the particle's previous movement. Gas molecules are already in random motion, but we are only concerned here with the oscillation.

If the wave is sinusoidal, the variation *with time* of the displacement (y) and velocity (v) of a typical particle are given by fig. 41.1.

Notes

- (a) These variations differ in phase by $\pi/2$ rad.
- (b) Particles at other locations show the same variations, but with systematic phase shifts relative to their near neighbours.

The displacement of a particle from its usual position causes a variation in the *number of molecules per unit volume*. Recalling (p. 199) that

$$p = \frac{1}{3} n m \bar{c}^2$$

$$= \frac{1}{3} \rho \bar{c}^2$$

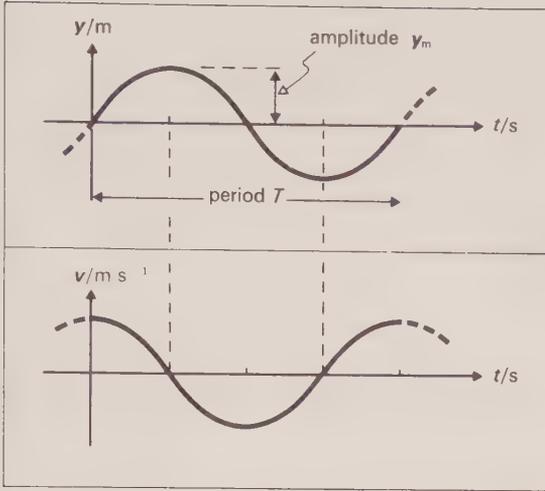


Fig. 41.1. Displacement–time and velocity–time curves for a particle disturbed by a sinusoidal wave.

we deduce that there are also variations in

- (a) pressure, and
- (b) density, relative to their usual values.

The disturbance consists of a train of **compressions** or **condensations** (of higher pressure and density than usual) and **rarefactions** (lower pressure and density). Fig. 41.2 relates particle displacement and excess pressure (or excess medium density) to usual particle position.

	At a compression	When excess pressure and density are zero	At a rarefaction
then displacement y is	zero	$\pm y_m$	zero
then particle velocity \dot{y} is	maximum and positive	zero	maximum and negative

Notes

(a) Graphs such as those in fig. 41.2 necessarily represent a *longitudinal* displacement as though it were transverse. They cannot be a pictorial representation of reality.

(b) A **waveform** is a graph showing either

(i) pressure variation plotted against time at a given location (as one would display on an oscilloscope), or

(ii) pressure variation plotted against position at an instant of time.

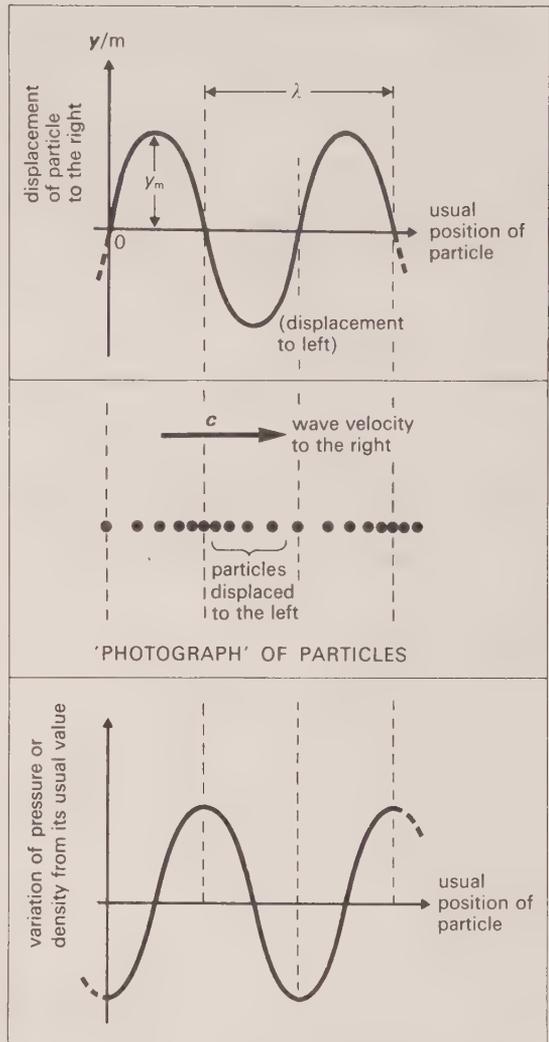


Fig. 41.2. Zero particle displacement corresponds to maximum variation in pressure and density. The graphs correspond to one instant of time.

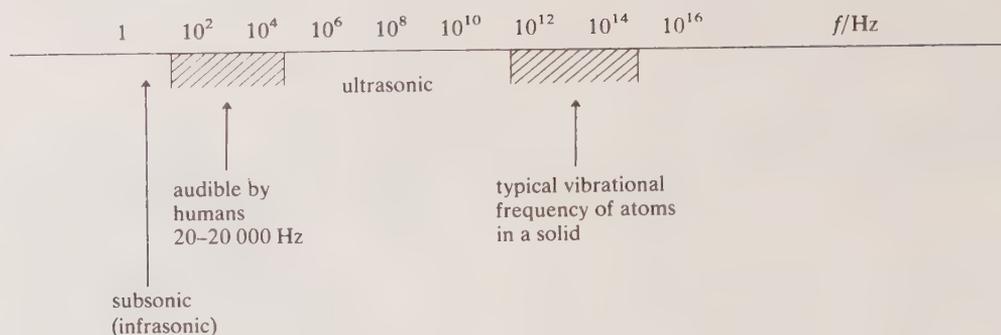
(c) It can be shown that the size of the pressure variation is proportional to the size of the particle displacement: therefore fig. 41.2 is also a waveform.

41.3 FACTS ABOUT SOUND WAVES

Sound waves are longitudinal mechanical waves audible to humans, and are generated by vibrating strings, air columns and surfaces.

An earthquake produces examples of **subsonic** (or **infrasonic**) waves. **Ultrasonic** waves of frequencies as high as 6×10^8 Hz (with a wavelength

Spectrum



Note that the numbers are presented on a logarithmic scale.

in air of about 5×10^{-7} m) can be produced by using the piezoelectric effect in a quartz crystal.

Bats produce waves of wavelength about 20 to 30 mm, a range beyond adult human hearing.

Some uses of ultrasonics

(a) Low intensity

- (i) echo sounding,
- (ii) finding flaws in materials,
- (iii) locating tumours in the body.

(b) High intensity

- (i) destruction of living cells suspended in liquids,
- (ii) removal of grease and dirt.

Speeds

Substance	Speed $c/10^2 \text{ m s}^{-1}$
granite (at 293 K)	60
fresh water (at 298 K)	15
air (at 273 K)	3.3

Amplitudes (for $f = 10^3$ Hz)

	Air molecule displacement amplitude y_m/m	Variation in air pressure $\Delta p/\text{Pa}$
for faintest sound that can be heard	10^{-11}	2×10^{-5}
for loudest sound that can be tolerated	10^{-5}	30

Intensity

The intensity I of a wave is defined by

$$I = \frac{P}{A} = \frac{E}{At} \quad (\text{p. 104})$$

$$I \left[\frac{\text{W}}{\text{m}^2} \right] = \frac{P[\text{W}]}{A[\text{m}^2]}$$

Examples ($f \approx 10^3$ Hz)

Loudness of sound	$I/\text{W m}^{-2}$
threshold of hearing	10^{-12}
a person talking	10^{-6}
threshold of feeling	$10^0 = 1$

The intensity decreases away from the source

(a) because the energy is distributed through a larger volume (for a point source and isotropic medium the intensity will obey an *inverse square law*), and

(b) because of **attenuation**: the wave energy is gradually converted by an imperfectly elastic medium into random energy of the medium's molecules (its internal or thermal energy).

The *power* of a source is the rate at which it converts other forms of energy into sound-wave energy. The power of a full orchestra is ~ 70 W.

The human ear

The assessment of **loudness** is controlled by both

- (a) the *wave intensity*, which depends on
 - (i) (amplitude)²
 - (ii) (frequency)²
 - (iii) speed, and
- (b) the *sensitivity* of the hearer to the particular frequency being sounded.

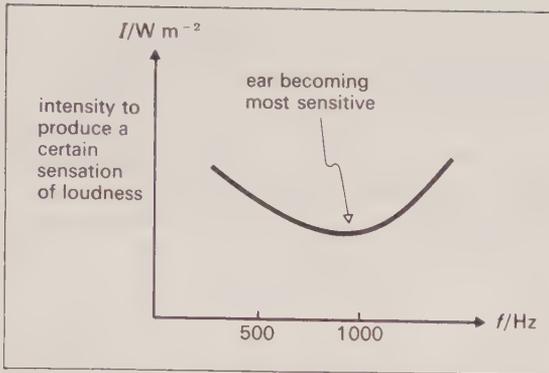


Fig. 41.3. The ear's sensitivity depends on frequency.

The ear responds to pressure variations, rather than air molecule displacement. It can be shown that

$$I = \frac{(\text{pressure amplitude})^2}{2(\text{air density})(\text{wave speed})}$$

$$I = \frac{\Delta p_{\max}^2}{2\rho c}$$

The frequency does not appear in this expression, so there will often be an advantage in measuring *pressure* rather than *displacement* amplitude.

Fig. 41.3 indicates in a general way the variation of sensitivity of the human ear.

41.4 NOISE AND MUSICAL NOTES

A **noise** is an audible sound which is not harmonious to the ear: it is heard when the ear-drum is disturbed by a wave whose waveform is *non-periodic*.

A **musical note** is produced when the sound is pleasant: it is heard when the waveform is *periodic*. The air molecules which transmit the energy undergo a periodic motion, which is the result of superposing a relatively small number of simple harmonic motions.

Musical notes are described conveniently by three properties or characteristics:

(a) **Loudness**, which was discussed in 41.3.

(b) **Pitch**. Pitch is a subjective property of a musical note that is determined principally by the *frequency* of vibration of the air molecules, which in turn is fixed by the frequency of vibration of the source. It determines the position of a particular note in the musical scale, being analogous to colour in the visible spectrum.

(c) **Quality**. The quality of a note is controlled by the detailed behaviour of the air molecules, which is determined by the exact mode of vibration of the source.

Two notes of the same amplitude and frequency from different instruments will show *different waveforms* on an oscilloscope, and are said to have different quality or **timbre** (fig. 41.4).

(See also *overtones* and *harmonics* on p. 330.)

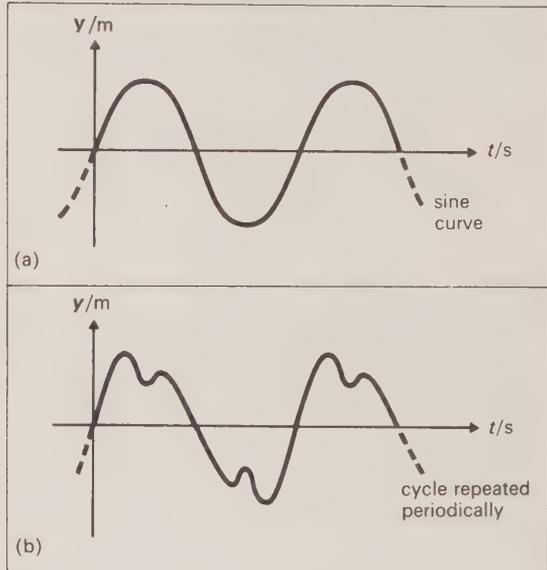


Fig. 41.4. Waveform (a) has a sinusoidal shape and sounds characterless beside the sound from (b), whose source is not vibrating with s.h.m.

41.5 GENERAL PROPERTIES OF SOUND WAVES

Being a longitudinal wave motion, sound has all the properties of waves discussed on p. 99 except polarization.

(A) Reflection

Reflection occurs such that the law of reflection is obeyed: this enables images (**echoes**) to be formed. *Periodic* reflection of a noise from a set of railings results in a musical note.

(B) Refraction

When sound energy is incident on a boundary at which the wave speed changes, we observe

(a) part of the energy carried back into the first medium by *partial reflection*,

(b) part of the energy *transmitted* into the second medium (unless there is total reflection).

The transmitted energy shows

(i) a change of direction of travel, and

(ii) a change of wavelength λ , since the frequency (being determined by the source) does not alter.

For a given pair of media we can define a **relative refractive index** by

$${}_1n_2 = \frac{c_1}{c_2}$$

and according to *Snell's Law* (p. 100)

$${}_1n_2 = \frac{c_1}{c_2} = \frac{\sin \theta_1}{\sin \theta_2} = \text{constant}$$

for waves of a given frequency.

Ripple tank analogies exist. A balloon filled with a relatively dense gas would become a *converging* agent (p. 102); one filled with a gas of low density would be a *diverging* agent. Simple experiments on focussing are complicated by diffraction effects (since when $f \sim 300$ Hz (say), then $\lambda \sim 1$ m).

The effect of temperature change on wave speed is discussed on p. 328.

(C) Total internal reflection

This is a general wave property, and is not confined to light waves.

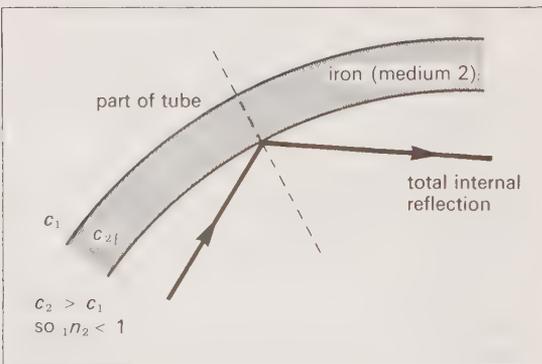


Fig. 41.5. Total internal reflection of sound energy.

- If (1) medium 2 is rarer than medium 1, and
- (2) θ_1 exceeds the critical angle θ_c

then (apart from local effects) no energy is transmitted into medium 2. (We can find θ_c from

$$\theta_c = \arcsin {}_1n_2 = \operatorname{arccosec} {}_2n_1.)$$

Examples

(a) The Whispering Gallery of St. Paul's Cathedral.

(b) Ships' speaking tubes (fig. 41.5).

(D) Dispersion

For waves whose frequency lies in the *audible range*, the wave speed in air is independent of wavelength, i.e. dispersion does *not* occur. (The different harmonics of a cymbal clash are not separated in time for an observer 100 m distant.) Dispersion does occur at ultrasonic frequencies.

(E) Speed

The speed c of sound in air is about 330 m s^{-1} . Sources of sound often have speeds v such that (v/c) is relatively large. Hence the *Doppler effect* (p. 119) is frequently observed.

Reverberation

The energy given out by a sound source in a large hall reaches the listener

- (a) directly,
- (b) by regular reflection (echoes), and
- (c) by diffuse reflection.

When each route produces sounds of appreciable intensity, a noise emitted at a given instant by the source takes some time to die away at the location of the listener. The phenomenon is called **reverberation**, and is of importance in the acoustics of buildings.

41.6 PROPERTIES BASED ON SUPERPOSITION

(A) Interference

A sounding tuning fork shows an interference pattern. If the fork is held with the shaft vertical, and rotated once about a vertical axis close to the ear, the loudness of the sound shows four maxima, and four minima.

Interference experiments are often complicated by unwanted reflections, but a *Young's* experiment can be carried out using two identical speakers activated by the same oscillator. Unless a pure note is used, the fringe pattern will be analogous to that produced by white light (p. 297).

Quincke's method (p. 332) gives a simple demonstration of superposition in sound.

(B) Beats

(Refer to the explanations on p. 114.)

When two notes of comparable frequency are sounded together, the *intensity* or loudness at a given location varies in time with a frequency

$$f_{\text{beat}} = (f_1 \sim f_2)$$

Suppose f_1 is known, and f_{beat} is measured, then f_2 can be calculated from

$$f_2 = f_1 \pm f_{\text{beat}}$$

To discover whether

$$f_1 < f_2 \quad \text{or} \quad f_1 > f_2$$

one varies either in a *known* direction (e.g. loading a tuning fork decreases f), and observes whether f_{beat} increases or decreases. This method is used for *tuning*.

Beats will also be heard when a wave is superposed on its own reflection from a *moving* reflector. According to the *Doppler effect* the reflected wave has a changed frequency (p. 119).

(C) Stationary waves

(Refer to the full account on p. 110.)

Superposition between waves emitted by identical loudspeakers (energized by the same alternating current) set up facing each other would result in a stationary wave pattern. There would be no net energy transfer along the line joining them.

In practice it is easier to superpose a wave on its reflection from a *still* reflector. It is most important to distinguish

(a) *displacement nodes* and *pressure antinodes* which occur together, from

(b) *displacement antinodes* and *pressure nodes*, which also occur together.

The adjacent node-antinode separation for each is always $\lambda/4$, but it is safest to specify which type is considered. (See p. 330.)

(D) Diffraction

Diffraction is discussed fully in chapter 39. Results obtained there apply fully to sound waves, but of course the scale of apparatus is modified to take into account the greater wavelength involved.

Diffraction effects are important in the design of foghorns and directional loudspeakers.

41.7 SPEED OF SOUND BY A DIMENSIONAL ARGUMENT

Suppose the wave speed c through a gas is governed by

- (a) an elasticity factor, its bulk modulus K ,
- (b) an inertia factor, its undisturbed density ρ , and
- (c) the wavelength λ of the waves.

Provided no relevant factor has been omitted (of which we cannot be sure), then

$$c = (\text{dimensionless constant}) K^x \rho^y \lambda^z$$

Equating dimensions

$$[LT^{-1}] = [MLT^{-2} \cdot L^{-2}]^x [ML^{-3}]^y [L]^z$$

Equating separately the dimensions of

$$T, \text{ we have } -1 = -2x$$

$$M, \quad 0 = x + y$$

$$L, \quad 1 = -x - 3y + z$$

from which

$$x = \frac{1}{2}, \quad y = -\frac{1}{2} \quad \text{and} \quad z = 0$$

Thus

$$c \propto \sqrt{\frac{K}{\rho}}$$

The dimensionless constant is shown by full analysis to be 1.

$$c = \sqrt{\frac{K}{\rho}}$$

According to this argument the speed does not depend on the wavelength, unless this controls the value of K .

If we assume that conditions are

(a) *isothermal* (p. 208), as *Newton* did, we can show

$$K_{\text{iso}} = p$$

(b) *adiabatic* (p. 209), as *Laplace* did, we can show

$$K_{\text{ad}} = \gamma p$$

where

$$\gamma = \frac{C_{p,m}}{C_{v,m}} \quad (\text{p. 209})$$

Experiment indicates that $K = \gamma p$, i.e. that sound waves in the audible range are propagated *under adiabatic conditions*.

The explanation is complex, but we suppose that at audible frequencies there is insufficient time for the wave energy to flow as heat from a compression to a relatively distant rarefaction, because the temperature gradient is too small. We use

$$c = \sqrt{\frac{\gamma p}{\rho}}$$

Speed of sound through bars

An expression for the speed of any mechanical wave through an elastic medium will be of the form

$$c = \sqrt{\frac{\text{elasticity factor}}{\text{inertia factor}}}$$

For example, the speed of longitudinal waves in a solid rod, such as that used in the dust-tube experiment, is given by $c = \sqrt{E/\rho}$, where E is the material's *Young* modulus and ρ its density.

41.8 VARIATIONS IN THE SPEED OF SOUND

For an amount μ [mol] of an ideal gas

$$pV = \mu RT$$

If we put M_m = molar mass, and M_g = mass of a particular gas sample, then

$$\rho = \frac{M_g}{V} = \frac{\mu M_m}{V} = \frac{p M_m}{RT}$$

Substituting in $c = \sqrt{\frac{\gamma p}{\rho}}$

we have

$$c = \sqrt{\frac{\gamma RT}{M_m}}$$

(A) Pressure

Changes in pressure are proportional to changes in density, and do not affect γ , R or M_m . At a fixed temperature, the speed of sound is *independent of pressure*.

(B) Temperature

Provided the temperature change does not alter either γ or M_m , we have shown

$$c \propto \sqrt{T}, \text{ i.e. } \frac{c_1}{c_2} = \sqrt{\frac{T_1}{T_2}}$$

This is an important factor in the tuning of wind instruments.

(C) Nature of gas

If the nature of the transmitting gas changes, then so may γ and M_m . Thus

(a) γ decreases as the gas atomicity increases (p. 211): this decreases c , but

(b) a low value of M_m corresponds to a low gas density, and a higher value for c .

Variations in atmospheric conditions

(A) Effect of humidity

An increase in the proportion of (triatomic) water vapour molecules decreases γ slightly and M_m considerably. Since the latter effect predominates, the wave speed in humid air slightly exceeds that in dry air.

(B) Effect of wind

The wind velocity u is added vectorially to that of the sound: since u usually increases with height, the wavefront is distorted according to fig. 41.6.

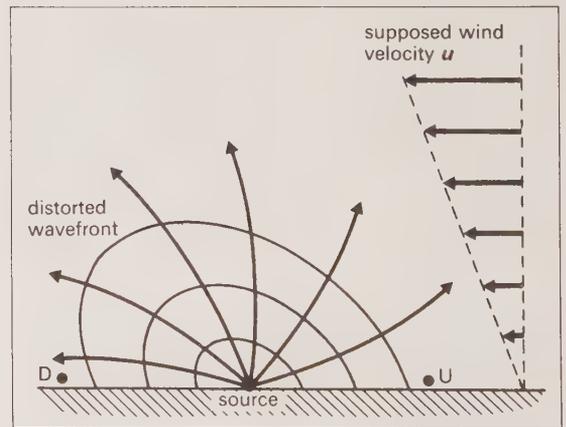


Fig. 41.6. Wind makes a source easier to hear downwind (at D), but more difficult upwind (at U).

(C) Audibility and temperature gradient

Fig. 41.7 (opposite) is explained by

(a) the speed of sound in warm air exceeding that in cold air, and

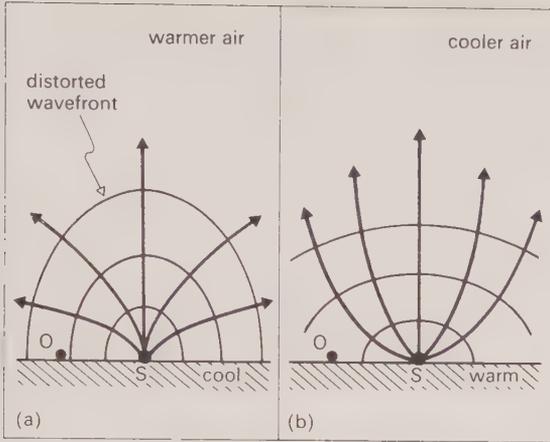


Fig. 41.7. In (a) (perhaps a summer evening) refraction causes distant sounds to be heard distinctly, whereas in (b) (a summer morning) refraction has taken the sound away from the Earth.

(b) air temperature being largely controlled by that of the ground, or an open expanse of water, rather than by direct radiation from the Sun.

41.9 VIBRATION OF AIR COLUMNS

Principles

- (a) If the air inside a pipe is disturbed by a source of sound such as a vibrating reed, waves travel in both directions away from the source.
- (b) The conditions at the ends of the pipe (the boundary conditions) will determine the phase of any reflected waves.

(c) Reflected waves superimposed on incident waves will then establish a *stationary wave pattern* according to section 14.2 (p. 110).

(d) *Resonance* will occur only for those waves whose wavelengths are correctly matched to the length of the pipe. On reflection, such waves cause stimulated emission of sound from the source, thereby increasing their intensity.

Boundary conditions

- (a) At a closed end,
 - (i) the displacement amplitude must be zero
 - (ii) the pressure amplitude is therefore a maximum
 - (iii) the reflected pressure wave is in phase with the incident wave.
- (b) At an open end (supposing the tube to be narrow relative to the wavelength),
 - (i) the pressure amplitude must be zero
 - (ii) the displacement amplitude is therefore a maximum
 - (iii) the reflected wave is effectively π rad out of phase with the incident wave.

A small length of air outside a tube is also set into vibration, so the effective length of a vibrating air column exceeds the length of the pipe by an **end-correction**. The correction (about $0.6 \times$ (pipe radius) for each open end) is illustrated on p. 333.

Example of stationary wave formation

In fig. 41.8 stationary waves result from superposition of reflected waves on the incident progressive waves.

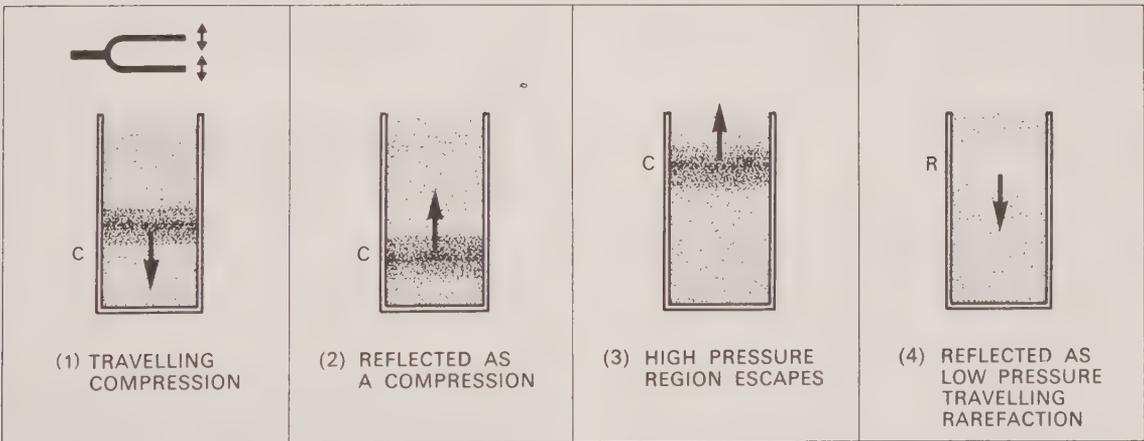


Fig. 41.8. Stationary wave formation in a closed pipe.

Resonance would occur if the reflected rarefaction in (4) entered the pipe at the same instant as a rarefaction from the tuning fork (constructive superposition).

Fig. 41.9 shows how longitudinal molecule displacement y and excess pressure Δp are related to molecular location x .

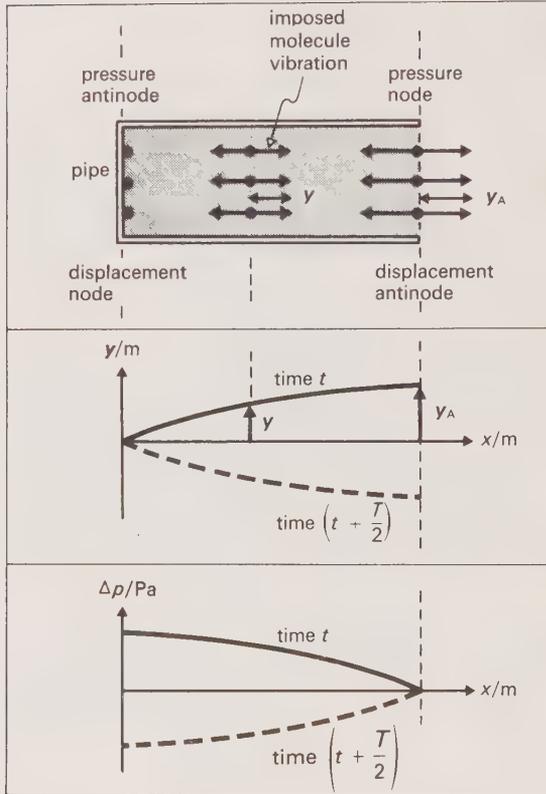


Fig. 41.9. Summary of stationary wave conditions inside a pipe.

Modes of vibration

Definitions

(a) The **fundamental** (mode of vibration) is that whose frequency f_0 is the lowest obtainable from the pipe.

(b) An **harmonic** is a note whose frequency is a whole number multiple of f_0 .

(c) An **overtone** is a note whose frequency is actually obtainable from a particular pipe.

For the diagrams that follow, N and A represent the locations of *displacement* nodes and antinodes. The distance AN = $\lambda/4$.

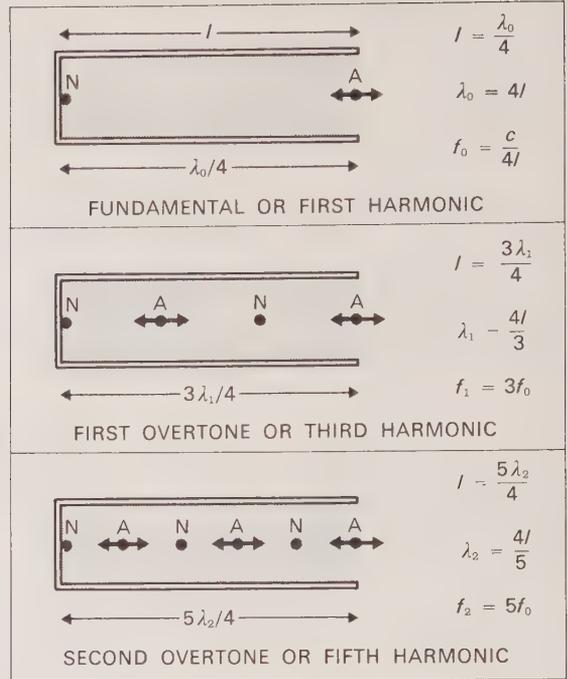


Fig. 41.10. First three modes of vibration of a closed pipe (a pipe stopped at one end).

(A) Closed (stopped) pipes

Fig. 41.10 demonstrates that the first overtone is the third harmonic, the second overtone is the fifth, etc.

A closed pipe gives odd harmonics only.

The overtone frequencies are

$$f_m = (2m + 1)f_0$$

where m is the number of the overtone.

(B) Open pipes

Fig. 41.11 demonstrates that open pipes give both even and odd harmonics, the m th overtone having a frequency

$$f_m = (m + 1)f_0$$

Notes

(a) For a given length of pipe, the fundamental of the closed pipe has half the frequency of that of the open.

(b) The presence of extra overtones in the note from an open pipe gives it a richer tone than that from a closed pipe. The higher overtones are promoted by hard blowing in a musical instrument.

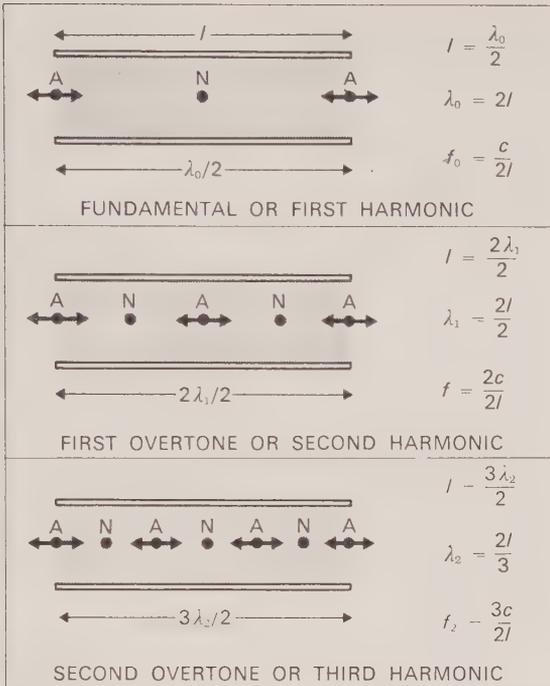


Fig. 41.11. First three modes of vibration of an open pipe.

The location of nodes and antinodes can be demonstrated by

(a) *jumping sand* on taut silk: the sand is buffeted by moving air molecules at *displacement antinodes*,

(b) *manometric capsule*: a rubber membrane responds actively to the pressure variations at *pressure antinodes*, and can cause a flame to flicker.

41.10 VIBRATION OF STRINGS

Principles and boundary conditions

(a) A transverse wave travels at speed c along a string mass per unit length μ and tension T , where

$$c = \sqrt{\frac{T}{\mu}}$$

Remembering that $[\mu] = [ML^{-1}]$, the reader should verify this equation using dimensional analysis.

(b) A stretched string is necessarily fixed at each end: the ends are therefore *displacement nodes*.

(c) When the string is set vibrating at frequency f , a *stationary wave* pattern of wavelength λ is established, where

$$\lambda = \frac{1}{f} \sqrt{\frac{T}{\mu}}$$

(d) *Resonance* occurs for those waves whose wavelengths are correctly matched to the length of the string. This is illustrated by *Melde's experiment* (fig. 41.12).

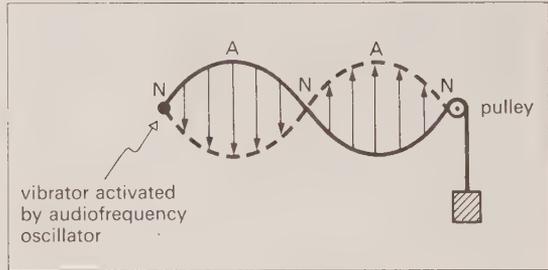


Fig. 41.12. Melde's experiment: the string vibrates with a large amplitude only when the frequency is correctly adjusted.

Notes

(i) The point to which the vibrator is attached is (surprisingly) nearly a displacement *node*.

(ii) The relative phases of the adjacent string segments are clearly revealed if it is illuminated by a flashing xenon lamp (a *stroboscopic device*).

Modes of vibration

Fig. 41.13 (overleaf) illustrates that a stretched string produces both odd and even harmonics.

The overtone frequencies are given by

$$f_m = (m + 1)f_0 = \frac{m + 1}{2l} \sqrt{\frac{T}{\mu}}$$

where m is the number of the overtone.

A string can vibrate in several modes at the same time. Their frequencies and relative intensities are determined by the original disturbance, and control the *quality* of the note emitted (as a travelling wave) from the string.

41.11 MEASUREMENT OF c , f AND λ

Since $c = f\lambda$, if we measure (or assume) any two, the value of the third can be calculated. The reader should refer to a textbook on sound for details.

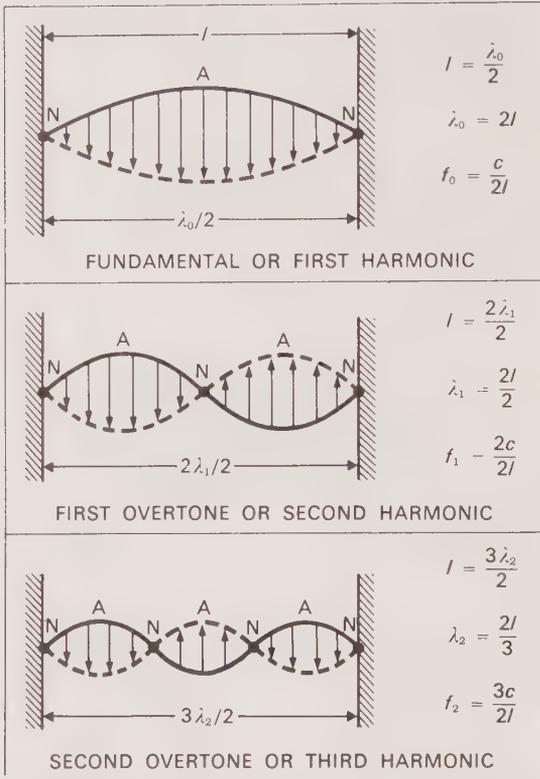


Fig. 41.13. First three modes of vibration of a stretched string.

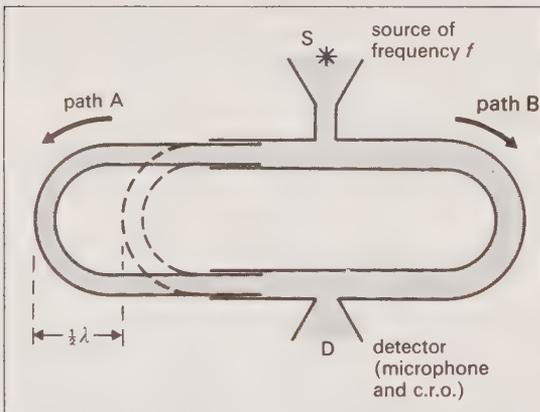


Fig. 41.14. Measurement of c using Quincke's tube.

(A) Speed c

- (a) Time sound over a known distance.
- (b) Quincke's tube method (fig. 41.14).

The sound from S has two possible paths (A and B) to the detector D. If the path difference is an odd number of half wavelengths, compressions

from A meet rarefactions from B at D. The sliding tube is moved to find two consecutive positions of minimum intensity. The distance between these positions is $\lambda/2$.

(c) Hebb's method is very similar, and gives a more precise result.

(B) Frequency f

(a) The frequency of a tuning fork can be measured by the *rotating disc method*, in which the prong causes a sinusoidal mark on a disc rotating at known speed.

(b) The frequency of a vibrating object can be found by viewing it through an adjustable *stroboscope*.

(c) Two frequencies which are close can be related by their *beat frequency*.

The known frequency could come from

- (i) a standard fork,
- (ii) a variable audiofrequency oscillator and loudspeaker,
- (iii) a siren, or
- (iv) a stretched wire (the *sonometer*).

(C) Wavelength λ

(a) *Kundt's dust tube*: illustrated by fig. 41.15. Standing waves in a gas are revealed by the periodic pattern made by dust in an enclosed tube. The dust may be agitated either by rubbing a brass rod which is clamped at its mid-point with one end projecting into the dust tube, or by means of a loudspeaker fitted over one end of the tube. The experiment is useful for comparing the speeds of sound in different gases.

(b) *The resonance tube*: illustrated by fig. 41.16. If l_0 and l_1 are the lengths of the tube for the first two positions of resonance and δ is the end-correction, then

$$\frac{\lambda}{4} = l_0 + \delta \quad \text{and} \quad \frac{3\lambda}{4} = l_1 + \delta$$

$$\therefore (l_1 - l_0) = x = \frac{\lambda}{2}$$

Measurement of x therefore eliminates the end-correction. If the tuning fork frequency f is known, the value of c can be calculated.

41.12 MEASUREMENT OF c IN A METAL ROD

The metal rod in fig. 41.17 is given a sharp tap with the hammer head. The time of contact t can be determined from the frequency setting of the

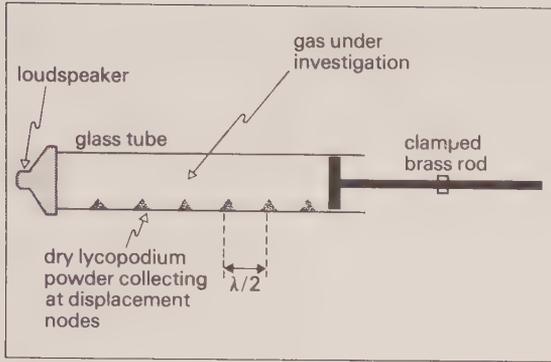


Fig. 41.15. Kundt's dust tube. (The loudspeaker and brass rod are alternatives.)

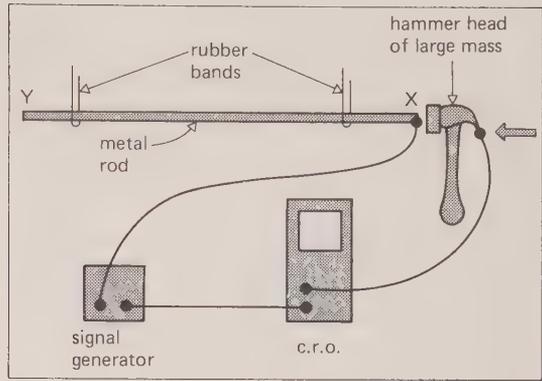


Fig. 41.17. Apparatus for measurement of c in a metal rod.

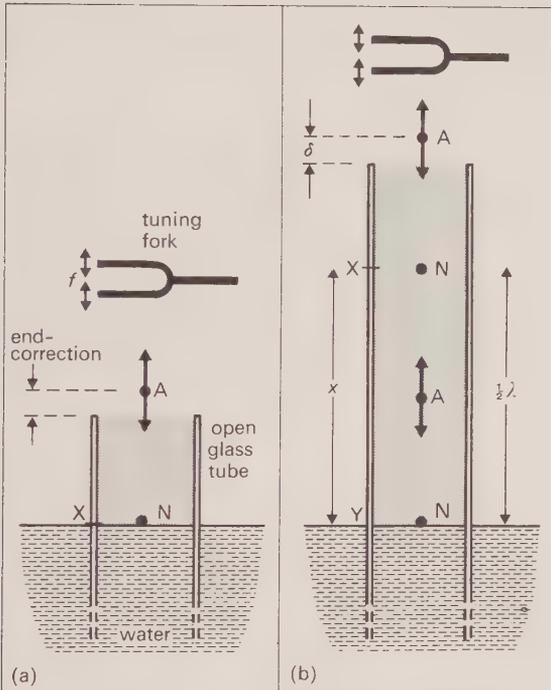


Fig. 41.16. Finding λ by resonance tube. (a) First resonance (fundamental), (b) second resonance (first overtone of the longer tube.)

Young modulus could be determined by using $c = \sqrt{E/\rho}$ where ρ is the density of the material.

41.13 SOUND RECORDING AND REPRODUCTION

A **microphone** converts sound energy into electrical energy. In most types, a thin diaphragm is used which vibrates due to the pressure changes of the incident sound waves. The diaphragm is attached to a device which produces alternating currents, and therefore electrical energy, when it moves. For high-quality reproduction, the microphone design must ensure that resonant effects are avoided. A **loudspeaker** converts electrical energy back into sound energy. It has a construction similar to a microphone but is designed to deal with greater power so that the sound is audible over an extended region. In most loudspeakers, the a.c. passes through a small cylindrical coil which is fixed to the centre of a conical diaphragm of stiff paper and which can move in an annular gap across which there is a strong magnetic field. The coil and diaphragm vibrate, producing sound waves with the same frequency as the a.c. The sound waves produced at the front of the cone are in antiphase with those produced at the back—a compression at the front coincides with a rarefaction at the back. If these two waves were to arrive together at the listener's ear, superposition would lead to a signal of low intensity. To minimize this possibility, and to improve the low frequency response, the cone is set in a large baffle (usually the cabinet) designed to prevent sound from travelling directly from front to back. The baffle design must also ensure that there are no unwanted resonance effects.

signal generator and the number of oscillations that appear on the c.r.o. screen. A compression pulse travels from X and is reflected at Y as a rarefaction pulse which breaks the contact between hammer and rod when it returns to X. The value of c is given by $2XY/t$. The experiment can be repeated for rods of different materials. The

Discs

The amplified alternating currents from the microphone produce mechanical vibrations of a magnetic cutter which travels in a spiral groove on the surface of a plastic disc. This results in an undulating trace on the sides of the groove which corresponds to the sound being recorded. In stereophonic recording, two coils at right angles cause the cutter to move in two planes simultaneously. The disc is electroplated with nickel, giving a *negative master* of the original record when the plastic is removed. A *positive mother* copy is produced from this and, after chromium plating, is used to give tough negatives for the production of plastic copies. When the record is played, the stylus follows the wavy trace and its vibrations are changed into currents which are amplified and passed to the loudspeaker. In a compact disc, an hour of sound is recorded digitally as a series of minute pits ($\sim 1 \mu\text{m}$) on a plastic disc of diameter 0.12 m. A narrow, intense beam from a semiconductor diode laser is focussed onto the pits and the reflections are converted into electrical signals by means of a photodiode. The compact disc never wears down, produces virtually no noise and has a significantly higher dynamic range than other discs.

Tapes

The microphone currents produce an alternating magnetic field in the region of the small gap of an electromagnet called the recording head. This re-

sults in a pattern of permanent magnets being superimposed on the specially coated plastic tape. The louder the sound being recorded, the greater the current and the stronger the magnet formed on the tape. The frequency of the sound (and the current) and the speed of the tape determine the length of the magnet. For high-quality recording, unreliable short magnets are avoided by using high tape speeds. With a double-track tape, half the tape width is used for recording before being turned over so that the other half tape width can be used. When the tape runs past the playback head, the varying magnetization on the tape induces small currents which are amplified and passed to the loudspeaker. *Monophonic* sound reproduction uses a single audiofrequency channel, and one or more loudspeakers in parallel at the output. *Stereophonic* sound reproduction uses two channels, and therefore two microphones, recording heads and loudspeakers. Each channel has an equal proportion of the tape width.

Film soundtrack

At one side of the picture frames of a cine film there is a sound track which has width variations superimposed on a narrow slit. These variations correspond to the original sound that reached the recording microphone. In reproduction of the sound, a beam of light is focussed onto the sound track and is amplitude-modulated by the width variations. This beam then reaches a photocell and produces a varying current that can be amplified to operate a loudspeaker.

IX ELECTROSTATICS

42	CHARGE AND COULOMB'S LAW	336
43	THE ELECTRIC FIELD	339
44	GAUSS'S LAW	344
45	ELECTRIC POTENTIAL	349
46	CAPACITORS	358
47	DIELECTRIC MATERIALS	365
	SUMMARY OF ELECTROSTATICS	370

42 CHARGE AND COULOMB'S LAW

42.1	CONDUCTORS AND INSULATORS	336
42.2	COULOMB'S LAW (1785)	336
42.3	THE PERMITTIVITY CONSTANT	337

42.4	CHARGE AND MATTER	337
42.5	CHARGE QUANTIZATION AND CONSERVATION	337
42.6	MEASUREMENT OF CHARGE	338

42.1 CONDUCTORS AND INSULATORS

Elementary experiments on rubbed bodies indicate that

- (a) there are two, and only two, kinds of electric charge,
- (b) unlike charges attract each other, but
- (c) like charges repel.

The charge associated with rubbed *ebonite* is *negative* by choice. This choice determines that the charge of an *electron* is also *negative*.

A charge placed on an **insulator** (or **dielectric** material) is confined to the region in which it was placed. An insulator has no *charge-carriers* that are free to migrate within the boundaries of the body.

A charge placed on a **conductor** may be allowed to spread over the whole surface of the body, because the body has charge-carriers which are free to migrate. The **Hall effect** (p. 403) enables us to find the *sign* of the charge-carriers.

(a) In *metals*, which are good conductors, the charges are carried by **free** or **conduction electrons** (negative).

(b) In *electrolytes*, which are intermediate conductors, the charge-carriers are ions and have both signs, whereas

(c) in *semiconductors* (p. 379) the charge-carriers may appear to be positive, or negative, or both.

42.2 COULOMB'S LAW (1785)

The law was formulated from experiments using the *torsion balance* of fig. 42.1.

Coulomb twisted the torsion head to maintain equilibrium: the angle of twist θ is a measure of the (electric) force F which acts on *each* (*Newton's Law III*).

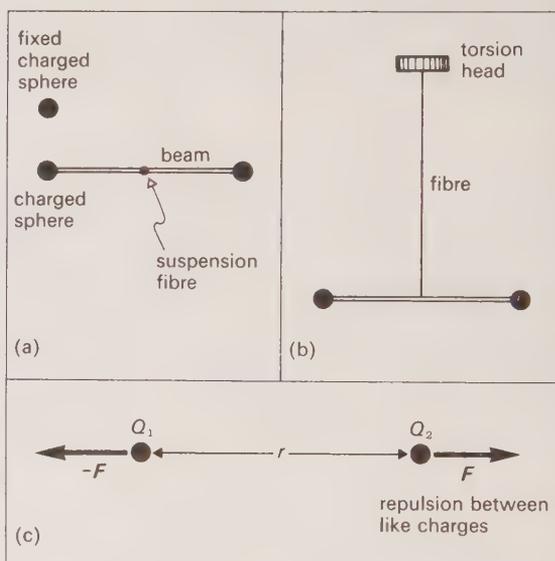


Fig. 42.1. *Coulomb's torsion balance experiment: (a) plan view, (b) elevation, (c) symbols.*

He varied

- (a) the charge separation r , and showed

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

- (b) the charge magnitudes Q_1 and Q_2 to show

$$F \propto Q_1 Q_2$$

He concluded that for *point charges*

$$F \propto \frac{Q_1 Q_2}{r^2}$$

The electric force between static electric charges is sometimes referred to as the **Coulomb force**.

This force

(a) obeys the *superposition principle*—a third charge will not affect the forces exerted by two charges on each other,

(b) is a *conservative* force (p. 46), and

(c) is a **central** force—it acts along the line joining the two point charges.

Verification of Coulomb's Law

(a) The *torsion balance* is accurate to a few per cent.

(b) The field inside a hollow charged conductor can only be zero for a strict inverse square law. *Cavendish* showed (to about 2%) that this was so for a sphere.

(c) *Plimpton* and *Lawton* repeated *Cavendish's* experiment to show an inverse n th law,

where $n = 2 \pm 2 \times 10^{-9}$

(d) *Rutherford's* scattering experiments with α -particles and the nucleus (a very close approach to a point charge) show that the law holds down to at least 10^{-15} m (p. 512).

42.3 THE PERMITTIVITY CONSTANT

Coulomb's Law can be written for a vacuum as

$$F = k \frac{Q_1 Q_2}{r^2}$$

in which k is a constant. We already have units for

(a) F [N] where $1 \text{ N} = 1 \text{ kg m s}^{-2}$

(b) r [m], and

(c) Q [coulomb, symbol C] (p. 372)

so k must be *found by experiment*. (Cf. the value of G on p. 169.)

$$k = \frac{Fr^2}{Q_1 Q_2} \approx 9.0 \times 10^9 \frac{\text{N m}^2}{\text{C}^2}$$

In the **rationalized system** of units it is convenient to write

$$k = \frac{1}{4\pi\epsilon_0}$$

where ϵ_0 is called the **permittivity constant**, the **permittivity of free space**, or sometimes simply the **electric constant**. It follows that

$$\epsilon_0 = 8.9 \times 10^{-12} \frac{\text{C}^2}{\text{N m}^2}$$

which is more usually written

$$8.9 \times 10^{-12} \text{ F m}^{-1} \quad (\text{p. 358})$$

ϵ_0 has the dimensions $[\text{M}^{-1}\text{L}^{-3}\text{T}^4\text{I}^2]$.

Thus for a vacuum we write **Coulomb's Law** as

$$F = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{Q_1 Q_2}{r^2}$$

(See p. 347 for a comment on rationalization.)

In chapters 42 to 46, we assume charges to be placed *in air*, in which case k has about the same value as it does in a vacuum.

42.4 CHARGE AND MATTER

Matter can be considered as made up of electrons, protons and neutrons (p. 126). Consider a hydrogen atom made up of an electron encircling a proton. Using experimental values for G and ϵ_0 we calculate the forces they exert on each other as follows:

$$(a) F_{\text{elec}} = \frac{1}{4\pi\epsilon_0} \cdot \frac{Q_1 Q_2}{r^2} \approx 8 \times 10^{-8} \text{ N, and}$$

$$(b) F_{\text{grav}} = G \cdot \frac{m_1 m_2}{r^2} \approx 4 \times 10^{-47} \text{ N.}$$

Thus $\frac{F_{\text{elec}}}{F_{\text{grav}}} \sim 10^{39}$

This demonstrates that *Coulomb forces* (not gravitational forces) are responsible for binding

(i) electrons to nuclei to form atoms,

(ii) atoms to atoms to form molecules and

(iii) molecules to molecules to form solids and liquids.

(The nucleus has a net positive charge—it is *not* the *Coulomb forces* that bind it together.)

42.5 CHARGE QUANTIZATION AND CONSERVATION

(A) Quantization

The charge of an electron is $-e$, where

$$e = 1.6 \times 10^{-19} \text{ coulomb}$$

No particle has been observed to carry a fraction of this charge.

Any charge Q that we observe in the laboratory is given by

$$Q = Ne$$

where N is a positive or negative integer, and e is the *fundamental electric charge* (or *quantum of charge*.)

Because e is so small and N consequently so large, in the usual laboratory experiments Q behaves as though it were continuous—the 'graininess' is not apparent.

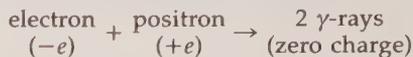
(B) Conservation

There are no known exceptions to the law that *charge is conserved*.

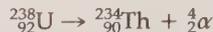
Examples

(a) When silk rubs glass, electrons are transferred from glass to silk, but the *pair* of them remain neutral. They constitute an *isolated* system.

(b) During an annihilation process, (p. 511)



(c) During a radioactive transformation, such as



the initial charge = $92e$, and

final charge = $(90 + 2)e$.

In an isolated system, the net charge is constant.

42.6 MEASUREMENT OF CHARGE

Electrostatic charges are conveniently measured in the laboratory using a *d.c. amplifier* (described in 46.9). A capacitor is connected across the input, and the small charge (typically a few nanocoulombs) is put onto a conductor connected to the upper input terminal (fig. 46.14). This causes a small current to flow in the input. It is amplified by a factor of about 10^{10} , enabling the output current to operate a milliammeter.

43 THE ELECTRIC FIELD

43.1	CONCEPT OF THE ELECTRIC FIELD	339
43.2	ELECTRIC FIELD STRENGTH E	339
43.3	ELECTRIC FIELD LINES	340
43.4	CALCULATION OF E AT A POINT	340

43.5	THE ELECTRIC DIPOLE	341
43.6	A POINT CHARGE IN AN ELECTRIC FIELD	341
43.7	AN ELECTRIC DIPOLE IN AN ELECTRIC FIELD	342

43.1 CONCEPT OF THE ELECTRIC FIELD

An **electric field** is said to exist at a point if a force of electrical origin is exerted on a test charge placed at that point.

It is convenient to consider the *interaction* in two parts, as shown in fig. 43.1.

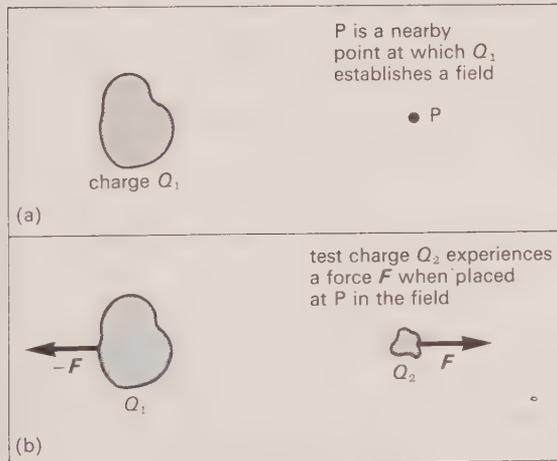


Fig. 43.1. The concept of electric field: (a) the cause of the field, and (b) the effect of the field.

The problems are

- calculating the field set up by a certain distribution of charges, and
- calculating the force experienced by a charge when placed in the field.

Analogous fields include

- the gravitational field (vector g),
- the magnetic field (vector B),

(iii) fluid flow, which can be described by a vector v , the velocity of flow at a particular point (p. 74).

The corresponding vector here is given the symbol E .

43.2 ELECTRIC FIELD STRENGTH E

E is called **electric field strength**, the *electric field intensity*, the *electric vector* or simply the electric field.

Suppose we place a test charge Q_0 at a point in an electric field where it experiences a force F .

The **electric field strength** at that point is defined by the equation

$$E = \frac{F}{Q_0}$$

$$E \left[\frac{\text{N}}{\text{C}} \right] = \frac{F[\text{N}]}{Q_0[\text{C}]}$$

Notes

(a) E represents a *vector* quantity whose direction is that of the force that would be experienced by a *positive* test charge.

(b) The unit for E following from the definition is N C^{-1} , but later it will be seen that a *volt metre*⁻¹ is an equivalent unit which is frequently used (p. 351).

(c) The magnitude of Q_0 must be small enough not to affect the distribution of the charges that are responsible for E —that is, Q_0 *must be very small* if it is not to change the value of E that it is measuring.

- (d) E is numerically equal to the force acting on unit charge placed at a point—but note (c).
- (e) The dimensions of E are $[MLT^{-3}I^{-1}]$.

43.3 ELECTRIC FIELD LINES

An **electric field line** is a convenient fictional concept developed to aid the visualization of an electric field. The field *line* is imaginary, but it represents a real *field*.

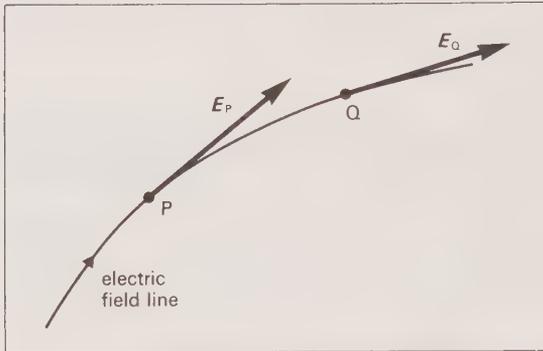


Fig. 43.2. An electric field line.

Field lines (fig. 43.2) are drawn such that

- (a) the *tangent* to a field line at a point gives the *direction* of E at that point, and
- (b) the *number* of field lines drawn per unit cross-sectional area is proportional to the *magnitude* E .

Field lines can be demonstrated (in strong fields) by grass seeds, which behave as iron filings do in a magnetic field.

Properties of electric field lines

- (a) They begin from positive charges, and end on equal negative charges. (Note that lines of B (p. 400) do not end, but form closed loops.)
- (b) They cannot cross (since the direction of E at a point is unique).
- (c) Where the lines are
 - (i) close together, the field is strong,
 - (ii) far apart, the field is weak, and
 - (iii) parallel and equally-spaced, the field is uniform.
- (d) As one travels *along* a field line in the direction of the arrow, the electric potential (p. 349) is *decreasing*.

- (e) They cut equipotential surfaces (such as those of conductors) at right angles (p. 351).

Although a field line shows the *initial* direction of movement of a resting positive charge, it does *not* show the path of a charged particle in an electric field. (Think of the trajectory of a projectile in the Earth's gravitational field.)

The reader should be able to sketch the field lines that represent the fields of the following charge distributions:

- (a) a negatively charged sphere,
- (b) an infinite sheet of positive charge, *uniformly distributed*,
- (c) two equal like charges, and
- (d) two equal unlike charges.

43.4 CALCULATION OF E AT A POINT

(A) Due to a point charge

Consider a point distance r from a point charge Q (and strictly in a vacuum).

By *Coulomb's Law*

$$F = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{QQ_0}{r^2}$$

gives the force F on a test charge Q_0 .

By the definition of E

$$E = \frac{F}{Q_0} = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{Q}{r^2}$$

N.B. Radial field lines ONLY

gives the magnitude E . Its direction is radial.

Note that although this is a useful result, it is for the *special case* of a field established by a *point charge*, and it takes no account of any charges that may be induced on nearby conductors.

(B) Due to several point charges

Suppose we have several point charges Q_1 , etc. Then we can

- (a) evaluate E_1 , etc., as above, and
- (b) find $E = \Sigma E_1$

by using the principle of superposition and *vectorial* addition. There is an example in the next section.

If the charge distribution is effectively continuous (e.g. throughout a volume) calculus may be necessary.

43.5 THE ELECTRIC DIPOLE

A pair of charges $+Q$ and $-Q$ separated by a distance $2a$ constitute an **electric dipole** of electric dipole moment $p = 2aQ$. (See also p. 342.)

$$E_P = E_1 + E_2$$

which is to be found by vector addition.

In magnitude

$$E_1 = E_2 = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{Q}{(a^2 + x^2)}$$

$$E_P = 2E_1 \cos \theta$$

$$= \frac{2}{4\pi\epsilon_0} \cdot \frac{Q}{(a^2 + x^2)} \cdot \frac{a}{(a^2 + x^2)^{1/2}}$$

$$= \frac{1}{4\pi\epsilon_0} \cdot \frac{2aQ}{(a^2 + x^2)^{3/2}}$$

Suppose $x \gg a$, so that $r \approx x$, then putting $p = 2aQ$

$$E_P = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{p}{r^3}$$

where P is a point along the equatorial axis of the dipole distance r away. The direction of E_P is given in fig. 43.3.

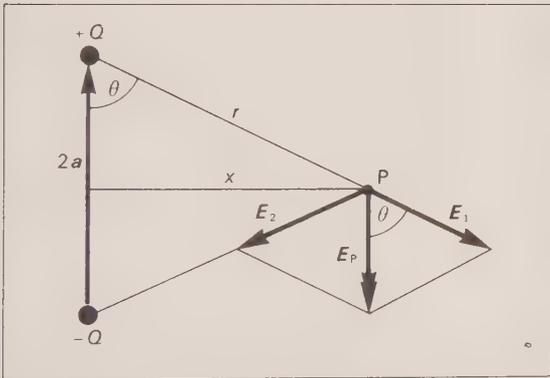


Fig. 43.3. The field set up by an electric dipole.

43.6 A POINT CHARGE IN AN ELECTRIC FIELD

(A) Charge released from rest in a field

The particle in fig. 43.4 does not experience forces resulting from its own field (unless its charge changes the charge distribution of the plates). The mass m carrying the charge Q is accelerated as is a mass in a gravitational field.

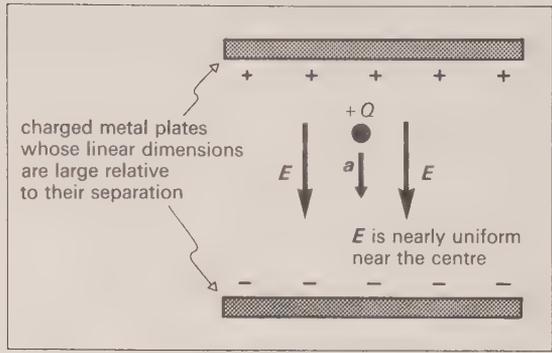


Fig. 43.4. A charged point mass in an electric field

$$a = \frac{F}{m} = \frac{EQ}{m} \text{ and is uniform.}$$

The k.e. acquired by the particle in moving distance y

$$= \text{work done by electric force}$$

$$= QEy$$

(Compare this with p.e. = mgh for a gravitational field.)

Application: Millikan's experiment (p. 472), in which an oil drop acquires a terminal velocity.

(B) Charge moving through a field

Suppose an electron (fig. 43.5) in the beam has an initial horizontal velocity u . Its subsequent

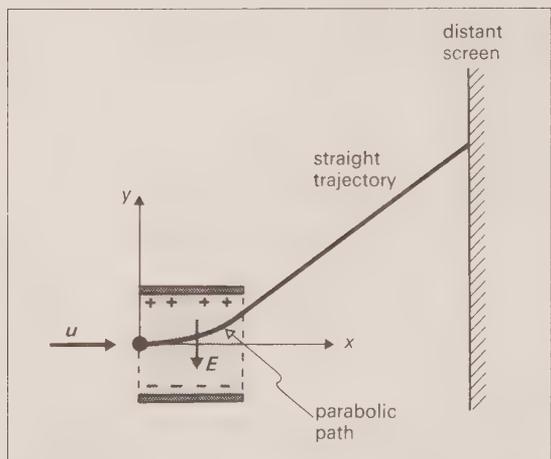


Fig. 43.5. An electron beam being deflected by an electric field.

upward acceleration while between the plates has a size (Ee/m_e) , and is uniform if the electric field lines are parallel.

Thus using $s = ut + \frac{1}{2}at^2$

horizontally $x = ut$

vertically $y = \frac{\frac{1}{2}(Ee)t^2}{m_e}$

Eliminating t , $y = \left(\frac{eE}{2m_e u^2}\right)x^2$

which shows that the path between the plates is parabolic (since it is of the form $y = kx^2$ (p. 25)).

Applications

- (a) Deflection experiments to measure e/m_e (p. 470)
- (b) The cathode ray oscilloscope (p. 473).

43.7 AN ELECTRIC DIPOLE IN AN ELECTRIC FIELD

Refer to fig. 43.3 for the symbols a and Q .

We define the **electric dipole moment** by the equation

$$p = 2aQ$$

$$p[\text{C m}] = 2a[\text{m}]Q[\text{C}]$$

The concept of an electric dipole is an important one which is discussed further on p. 366.

(A) Torque acting on a dipole

Refer to fig. 43.6.

Using $E = \frac{F}{Q_0}$ we have $F = EQ$

The *net force* on the dipole is zero.

The torque of the couple

$$\begin{aligned} T &= F(\text{arm of couple}) \\ &= EQ2a \sin \theta \\ &= (2aQ)E \sin \theta \\ &= pE \sin \theta \end{aligned}$$

Notes

- (a) The torque tends to align p and E .
- (b) p is a vector whose direction is shown in fig. 43.6.
- (c) $T = pE \sin \theta$ could have been used to define electric dipole moment. (See p. 407.)

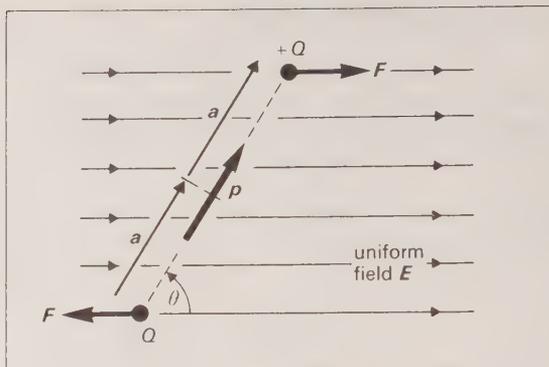


Fig. 43.6. An electric dipole experiences a torque in a uniform field but zero resultant force.

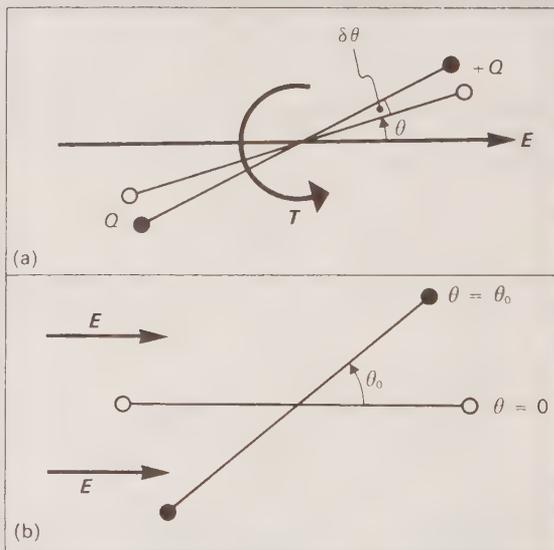


Fig. 43.7. Calculation of the work done in rotating a dipole: (a) the increment, and (b) the net change.

► (B) Energy and dipole orientation

Suppose an external agent exerts a torque of magnitude T while θ increases to $(\theta + \delta\theta)$ in fig. 43.7.

We assume the axis of T to be parallel to that of $\delta\theta$. The work done by the external agent

$$\begin{aligned} \delta W &= T \delta\theta \\ &= (pE \sin \theta) \delta\theta \end{aligned} \quad (\text{p. 60})$$

The work done in rotating the dipole from $\theta = 0$ to $\theta = \theta_0$

$$\begin{aligned} W &= \int \delta W \\ &= \int_0^{\theta_0} pE \sin \theta \, d\theta \\ &= pE(1 - \cos \theta_0) \end{aligned}$$

Notes

(a) The maximum amount of work has been

done when $\theta_0 = \pi$ rad: the dipole and E are then *anti-parallel*.

(b) The work done is stored as potential energy in the electric field set up by the system (the dipole and charges responsible for the external field).

(c) If we arbitrarily choose a zero of potential energy E_p at $\theta = \pi/2$ rad, we would have

$$E_p = -pE \cos \theta$$

Compare the result with that for the magnetic dipole (p. 407).

44 GAUSS'S LAW

44.1	ELECTRIC FIELD FLUX ψ_E	344
44.2	GAUSS'S LAW	345
44.3	COULOMB'S LAW DEDUCED FROM GAUSS'S LAW	345
44.4	THE NET CHARGE MIGRATES TO THE SURFACE OF AN INSULATED CHARGED CONDUCTOR	345
44.5	E INSIDE AN INSULATED HOLLOW CHARGED CONDUCTOR	346

44.6	THE MAGNITUDE OF AN INDUCED CHARGE	346
44.7	FARADAY'S ICE-PAIL EXPERIMENT	347
44.8	E DUE TO A SPHERICALLY SYMMETRIC CHARGE DISTRIBUTION	347
44.9	E DUE TO A LINE OF CHARGE	347
44.10	E DUE TO AN INFINITE CONDUCTING SHEET	348

44.1 ELECTRIC FIELD FLUX ψ_E

Suppose an electric field E makes an angle θ with the outward normal to a small area ΔA , as shown in fig. 44.1. Since ΔA is small, then E has effectively the same value at each point on the surface.

The electric field flux ψ_E through the area ΔA is defined by the equation

$$\psi_E = E \cos \theta \Delta A$$

$$\psi_E \left[\frac{\text{N m}^2}{\text{C}} \right] = E \left[\frac{\text{N}}{\text{C}} \right] \cdot \Delta A [\text{m}^2] \cdot \cos \theta$$

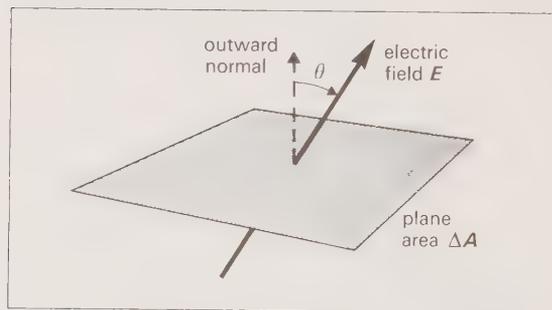


Fig. 44.1. Definition of electric field flux ψ_E .

Notes

(a) We can visualize ψ_E as being proportional to the number of electric field lines that are drawn through the area ΔA .

(b) When E is uniform, we can calculate ψ_E from

$$\psi_E = \left(\begin{array}{l} \text{resolved part of } E \\ \text{parallel to the} \\ \text{normal to an area} \end{array} \right) \times (\text{the area})$$

(c) A closed surface can frequently be divided conveniently into several elementary areas such as ΔA . The total flux ψ_E through the closed surface can be found by the scalar addition of the flux through the elementary areas. (See the examples of sections 44.9 and 44.10.) ψ_E represents a scalar quantity.

(d) For a closed surface the value of ψ_E over a specified area is taken to be *positive* if the electric field lines point *outwards* ($\theta < 90^\circ$), and *negative* if they point *inwards*.

►(e) If E is non-uniform, or A non-planar, then

$$\psi_E = \int E \cos \theta \, dA$$

►(f) For a closed surface, the flux emerging is written

$$\psi_E = \oint E \cos \theta \, dA$$

This means that the integral is to be taken over the whole surface.

$$(g) [\psi_E] = [ML^3T^{-3}I^{-1}]$$

44.2 GAUSS'S LAW

Gauss's Law states that for any closed surface

$$\epsilon_0 \psi_E = \Sigma Q$$

where ΣQ represents the algebraic sum of the charges enclosed by the surface.

(This means that if a surface encloses a pair of equal but opposite charges, the net charge is zero.)

The actual location of a charge inside a surface is not relevant. A charge outside the surface makes no net contribution to ψ_E .

In the next section we show that *Gauss's Law* is an alternative statement of *Coulomb's Law*: both are fundamental laws in electromagnetic theory. *Gauss's Law* can be proved by assuming an *inverse square law*, and is often simpler to apply than *Coulomb's Law*.

Experimental verification of Gauss's Law

It is shown below that *Gauss's Law* predicts that the charge on an insulated conductor moves to its outer surface. Several workers have devised methods to show that this is in fact so (p. 347). Their work is an experimental verification of the law.

The usefulness of Gauss's Law

(a) If we have a situation in which the charge distribution has sufficient symmetry for us to be able to calculate ψ_E , then we may be able to evaluate E at various points on a closed surface.

(b) If we know E at points on a closed surface, then calculation of ψ_E will enable us to find ΣQ . If in any situation we find that ψ_E is positive, this means that there is a net positive charge within that surface.

The technique of using the law

- (1) Draw the situation.
- (2) Imagine a closed surface to be constructed which is appropriate to the symmetry. (This imaginary surface is frequently called a **Gaussian surface**.)
- (3) Apply the law.

In sections 44.3 to 44.10 this is done

- (a) to deduce *Coulomb's Law*,
- (b) to investigate the charges of insulated conductors, and
- (c) to investigate the field established by various charge distributions.

44.3 COULOMB'S LAW DEDUCED FROM GAUSS'S LAW

We want to find the force F acting on a point charge Q_2 placed in a vacuum a distance r from a point charge Q_1 (fig. 44.2).

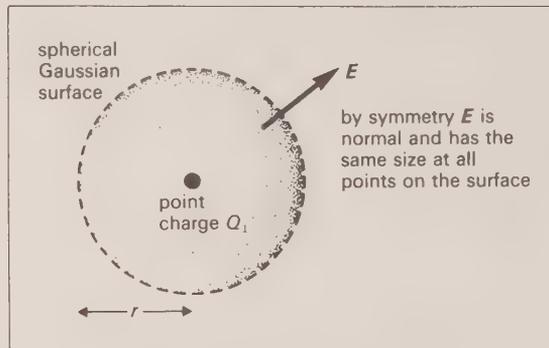


Fig. 44.2. Deduction of Coulomb's Law.

By *Gauss's Law*

$$\epsilon_0 \psi_E = Q_1$$

$$\text{in which } \psi_E = EA \cos \theta \\ = E4\pi r^2$$

$$\text{so } \epsilon_0 E4\pi r^2 = Q_1$$

$$E = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{Q_1}{r^2}$$

For the test charge Q_2 placed at the surface

$$F = EQ_2 = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{Q_1 Q_2}{r^2}$$

which is *Coulomb's Law*.

44.4 THE NET CHARGE MIGRATES TO THE SURFACE OF AN INSULATED CHARGED CONDUCTOR

Any charge distributes itself over a conductor very quickly until electrostatic equilibrium has

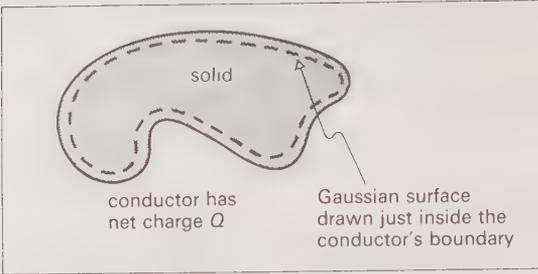


Fig. 44.3. An insulated charged conductor of arbitrary shape.

been reached. Then there is no systematic drift of the mobile charge-carriers. Since they are not accelerated, the average resultant force acting on them is zero: the field *within the material* of the conductor must be zero at every point.

At every point on the *Gaussian surface* of fig. 44.3

$$E = 0, \text{ and so } \psi_E = 0$$

By *Gauss's Law* $\epsilon_0 \psi_E = \Sigma Q$

in which $\psi_E = 0$

so $\Sigma Q = 0$

There is no net charge inside the *Gaussian surface*.

Any net charge placed on an insulated conductor resides entirely on its outer surface.

(See *Faraday's ice-pail experiment*, opposite.)

44.5 E INSIDE AN INSULATED HOLLOW CHARGED CONDUCTOR

Since the hollow conductor of fig. 44.4 is empty, the charge Q inside the *Gaussian surface* is zero. Using

$$\epsilon_0 \psi_E = \Sigma Q$$

$$\psi_E = 0$$

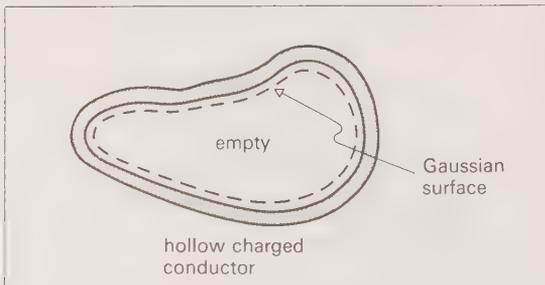


Fig. 44.4. An empty hollow charged conductor.

If there were a field at *any point* on the *Gaussian surface*, it would have the same direction (out of or into the surface) at *every point*. Thus $\psi_E = 0$ here requires $E = 0$ (but not in general).

The field must be zero at every point inside an empty hollow charged conductor.

44.6 THE MAGNITUDE OF AN INDUCED CHARGE

Suppose the conductor of fig. 44.5 *totally* encloses the charge $+Q$, and that there is an induced charge Q' . At every point on the *Gaussian surface*

$$E = 0$$

By *Gauss's Law*

$$\epsilon_0 \psi_E = \Sigma (\text{enclosed charge})$$

in which $\psi_E = 0$

so $\Sigma (\text{enclosed charge}) = Q + Q' = 0$

$$\therefore Q = -Q'$$

The induced charge is equal in magnitude but opposite in sign.

(See *Faraday's ice-pail experiment*, opposite.)

If the conductor were originally electrically neutral, it would now carry a charge $+Q$ on its outer surface (charge conservation).

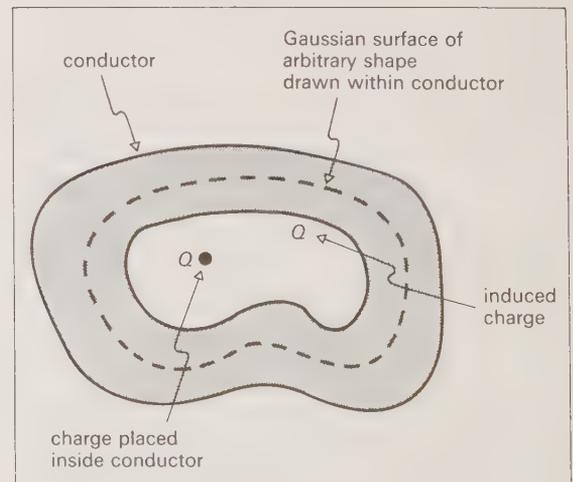


Fig. 44.5. Calculation of an induced charge.

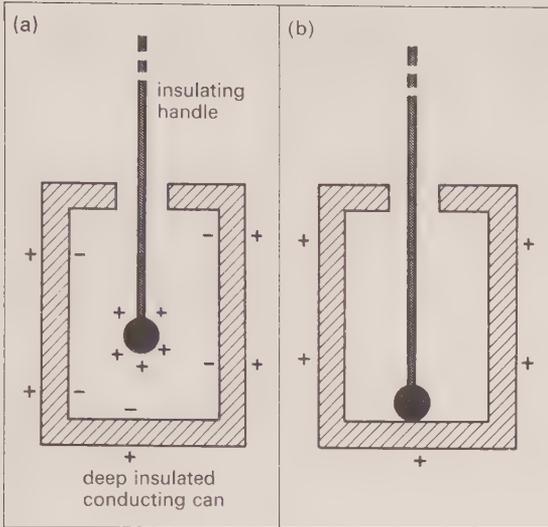


Fig. 44.6. Faraday's ice-pail experiment: (a) the induced charges, (b) charge transfer on contact.

44.7 FARADAY'S ICE-PAIL EXPERIMENT

This illustrates the ideas of sections 44.4 and 44.6. Refer to fig. 44.6.

In (a), provided the can is deep (so as to enclose the sphere), the original charge and the positive and negative induced charges all have the same magnitude.

In (b) neutralization has occurred: the net effect is that the ball's entire charge is transferred to the outside of the can.

This idea is used in the Van de Graaff generator (45.9).

44.8 E DUE TO A SPHERICALLY SYMMETRIC CHARGE DISTRIBUTION

Suppose the charge density ρ [C m⁻³] in any small volume is a function of r only (but not of direction).

By choice the Gaussian surface of fig. 44.7 is concentric with the spherical charge distribution.

By Gauss's Law

$$\epsilon_0 \psi_E = Q$$

in which

$$\psi_E = E4\pi r^2$$

so

$$\epsilon_0 E4\pi r^2 = Q$$

$$E = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{Q}{r^2}$$

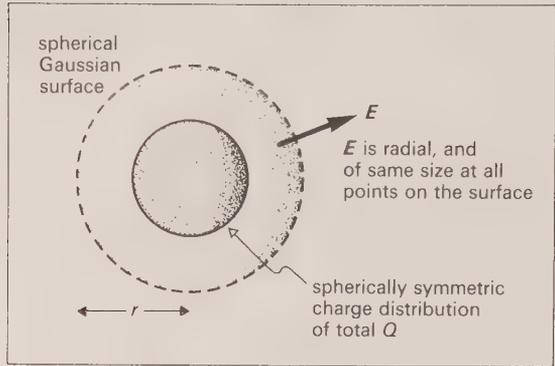


Fig. 44.7. A spherically symmetric charge distribution

At points outside the charge distribution, the charge behaves as though it were all placed at the centre.

A spherical shell makes no contribution to the value of E inside the shell. (Refer to the diagram on p. 172.)

We use these results in gravitation, where frequently they are derived by integration from the inverse square law (p. 169).

44.9 E DUE TO A LINE OF CHARGE

In fig. 44.8, by the choice of Gaussian surface, the flux emerging from it is made up from

(a) zero at the plane ends (where E is parallel to the surface), and

(b) $E2\pi rh$ over the curved surface (where E is everywhere perpendicular to the surface).

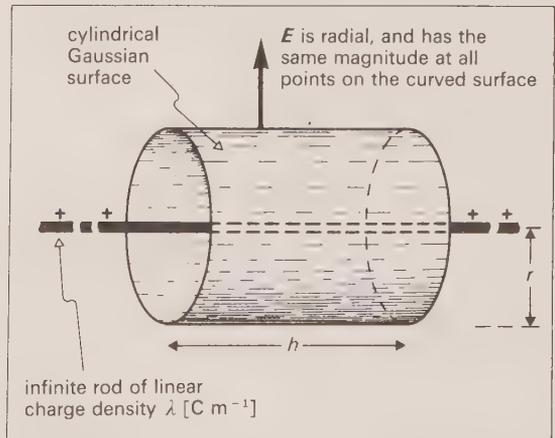


Fig. 44.8. E due to a line of charge.

$$\begin{aligned} \varepsilon_0 \psi_E &= Q \\ \text{becomes } \varepsilon_0 E 2\pi r h &= \lambda h \\ \text{so } E &= \frac{\lambda}{2\pi \varepsilon_0 r} \end{aligned}$$

and is radial in direction.

This result is used on p. 354.

Rationalization (mentioned on p. 337) is a convenient (but arbitrary) procedure. As a result of it situations involving

- (i) cylindrical symmetry (such as this one) give equations with factors of 2π ,
- (ii) spherical symmetry (e.g. p. 345) show factors of 4π , and those with
- (iii) plane symmetry (next section) involve equations without π .

44.10 E DUE TO AN INFINITE CONDUCTING SHEET

In fig. 44.9, by the choice of *Gaussian surface*, the flux emerging from the closed surface is

- (a) zero over the left plane end (since $E = 0$ at every point),
- (b) zero over the curved surface, and
- (c) EA over the right plane end.

$$\begin{aligned} \varepsilon_0 \psi_E &= Q \\ \text{becomes } \varepsilon_0 EA &= \sigma A \end{aligned}$$

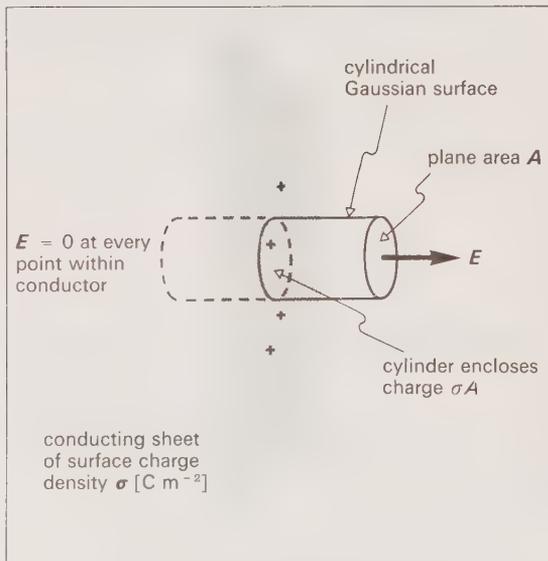


Fig. 44.9. An infinite conducting sheet.

so
$$E = \frac{\sigma}{\varepsilon_0}$$

and is normal to the infinite sheet.

This result applies

- (i) at all distances from an *infinite* conducting sheet, and
- (ii) to points *close* to a *finite* sheet (but not near the edges where the field lines are not parallel).

If the sheet is made of *insulating* material the charge is not *necessarily* on the surface, and this must be taken into account.

E close to the surface of a charged conductor

The previous result gives the value of E *very close* to the surface of *any charged conductor* whose surface charge density is σ .

$$E = \sigma / \varepsilon_0$$

$$E \left[\frac{\text{N}}{\text{C}} \right] = \frac{\sigma \left[\frac{\text{C}}{\text{m}^2} \right]}{\varepsilon_0 \left[\frac{\text{C}^2}{\text{N m}^2} \right]}$$

The direction of E at the surface of a *conductor* is *perpendicular to the surface*: if it were not, its resolved part along the surface would accelerate charges along the surface. This cannot happen if a conductor is to be in electrostatic equilibrium.

Field strength between parallel charged plates

The field between the plates

$$E = \sigma / \varepsilon_0$$

This can be verified by considering the *Gaussian surface* shown (fig. 44.10).

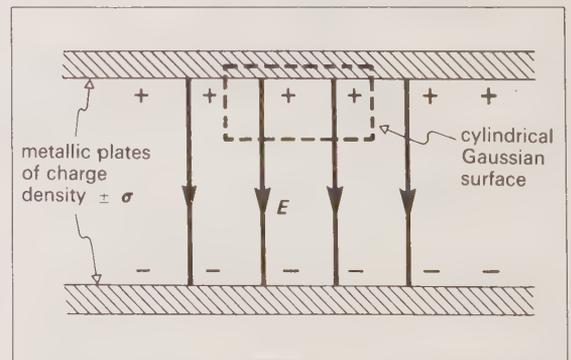


Fig. 44.10. Parallel charged plates.

45 ELECTRIC POTENTIAL

45.1	ELECTRIC POTENTIAL ENERGY, POTENTIAL AND POTENTIAL DIFFERENCE	349
45.2	RELATIONSHIP BETWEEN POTENTIAL AND FIELD	350
45.3	THE GRAVITATIONAL ANALOGY	352
45.4	POTENTIAL DUE TO A POINT CHARGE	353
45.5	A CONDUCTING MATERIAL IN AN ELECTRIC FIELD	353

45.6	FIELD AND POTENTIAL DUE TO AN INSULATED CONDUCTING CHARGED SPHERE	353
▶ 45.7	POTENTIAL DIFFERENCES CLOSE TO A LINE OF CHARGE	354
45.8	CHARGE DISTRIBUTION OVER A CONDUCTOR	355
45.9	THE VAN DE GRAEFF GENERATOR	355
45.10	THE GOLD LEAF ELECTROSCOPE	356
▶ 45.11	THE ATTRACTED DISC ELECTROMETER	356

45.1 ELECTRIC POTENTIAL ENERGY, POTENTIAL AND POTENTIAL DIFFERENCE

Electric potential energy and potential

To move charge B from position (1) to (2) in fig. 45.1 requires an external agent to do work, because B is in the field established by A. The work done by the agent increases the **electric potential energy** of the system.

The increase of electric p.e. depends on

- the magnitude of the charge, and
- the locations in the field of positions (1) and (2).

We say that different points in the field have different **electric potentials**.

Compare

- electric potential*, which is a property of a point in an electric field, even in the absence of a test charge, with

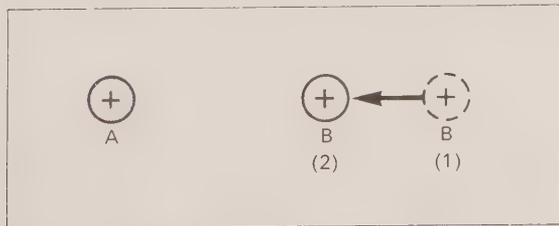


Fig. 45.1. To illustrate electric p.e.

- electric potential energy*, which depends on both the position and size of a charge.

Electric potential V and field strength E lead to equivalent (but alternative) ways of describing a given electric field. Their relationship is discussed on p. 350.

Electric potential V

Suppose an external agent does work W in bringing a test charge Q_0 from infinity to a particular point in an electric field.

Then we define the **electric potential** V at that point by the equation

$$V = \frac{W}{Q_0}$$

$$V \left[\frac{\text{J}}{\text{C}} \right] = \frac{W[\text{J}]}{Q_0[\text{C}]}$$

The unit of potential, a *joule coulomb*⁻¹, occurs so often it is convenient to call it a **volt (V)**.

Notes

- The potential at a point is numerically equal to the work done in bringing unit positive charge from infinity to the point (provided the field is not disturbed by the presence of such a large charge).

(b) $[V] = [ML^2T^{-3}I^{-1}]$.

(c) If the external agent does *negative work* (p. 43), then either

(i) it moves a negative charge to a point of positive potential, or

(ii) it moves a positive charge to a point of negative potential.

The agent then exerts a *restraining force* on the charge that it moves.

(d) Electric potential is a *scalar quantity*.

Potential difference V_{ab}

We are nearly always concerned with the *difference* of electrical potential between two points in a field. Suppose an external agent does work W_{ab} in taking a test charge Q_0 from a point A to a point B.

Then we calculate the **potential difference** V_{ab} between the points A and B by the equation

$$V_{ab} = V_b - V_a = \frac{W_{ab}}{Q_0}$$

$$V_{ab}[\text{V}] = \frac{W_{ab}[\text{J}]}{Q_0[\text{C}]}$$

$V_b > V_a$ if an external agent does *positive work* when moving a *positive* test charge.

The zero of potential

(a) The *practical zero* is that of the Earth (see p. 359).

(b) The *theoretical zero*, according to the definition of V , is that of a point at infinity.

Because we are nearly always concerned with *differences* of potential, the zero is not of fundamental importance.

The uniqueness of potential

In electrostatics the electric field is *conservative* (p. 46). In fig. 45.2 the work W_{ab} has the same value whatever path is taken in moving the test charge. So

$$V_{ab} = V_b - V_a = \frac{W_{ab}}{Q_0}$$

has the same value, and V_b and V_a are unique for the points A and B.

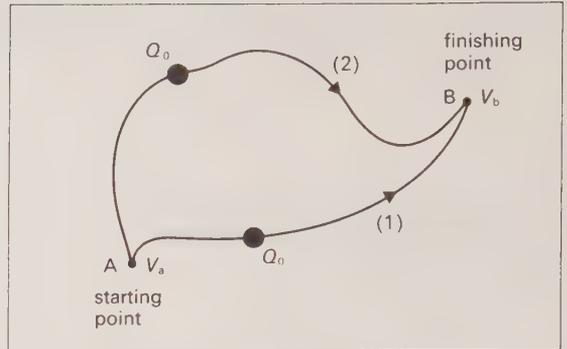


Fig. 45.2. Movement of a test charge along alternative paths.

The electronvolt

When a charged particle is accelerated by an electric field, its increased (mechanical) k.e. is accounted for by a reduction in (electrical) p.e.

Suppose an electron ($Q = -1.6 \times 10^{-19}$ C) moves through a p.d. of 1 volt: then the change of electrical potential energy is found from

$$\begin{aligned} W &= VQ && \text{(def. of } V) \\ &= (1 \text{ J C}^{-1}) \times (-1.6 \times 10^{-19} \text{ C}) \\ &= -1.6 \times 10^{-19} \text{ J} \end{aligned}$$

This quantity of energy is convenient for much work in atomic and molecular physics: it is called an **electronvolt** (eV). (Strictly it is not a *unit*, and does not form part of the SI.)

45.2 RELATIONSHIP BETWEEN POTENTIAL AND FIELD

We now have two different ways of describing the same field. We can either

- (a) (i) give values of E ,
- (ii) calculate the *force* on a charge at a point and
- (iii) draw electric field lines, or
- (b) (i) give values of V ,
- (ii) calculate the *potential energy* of a charge at a point, and
- (iii) draw equipotential lines or surfaces.

Equipotential surfaces are surfaces on which the potential is the same at all points: this means that no work has to be done in moving a test charge between any two points.

It follows that equipotential surfaces and electric field lines are perpendicular to each other at

any crossing point, since there can be no component of the E -field parallel to the equipotential surface. In electrostatics the surface of a conductor is an equipotential surface. (This is not so when a current flows.)

Equipotential surfaces due to a point charge are a family of concentric spheres (p. 353) with E -field lines directed radially inwards or outwards depending on whether the charge is negative or positive. Fig. 45.3 shows equipotential surfaces and E -field lines for some charge distributions. Note that these surfaces are three-dimensional.

Field strength and potential gradient

Refer to fig. 45.4 (overleaf).

The *electric force* on a positive test charge Q_0 at A or B is radially outward, and of magnitude EQ_0 . x is measured in the same direction as E . The work

δW done by *an external agent* in moving the charge from A to B is positive, being given by

$$\delta W = -EQ_0 \delta x$$

(In a move from A to B δx is negative.)

$$E = -\frac{1}{Q_0} \frac{\delta W}{\delta x}$$

$$= -\frac{dV}{dx} \quad \left(V = \frac{W}{Q_0} \right)$$

Let A and B be very close, so $\delta x \rightarrow 0$. Then

$$E = -\frac{dV}{dx}$$

$$E \left[\frac{\text{volt}}{\text{metre}} \right] = -\frac{\delta V[\text{volt}]}{\delta x[\text{m}]}$$

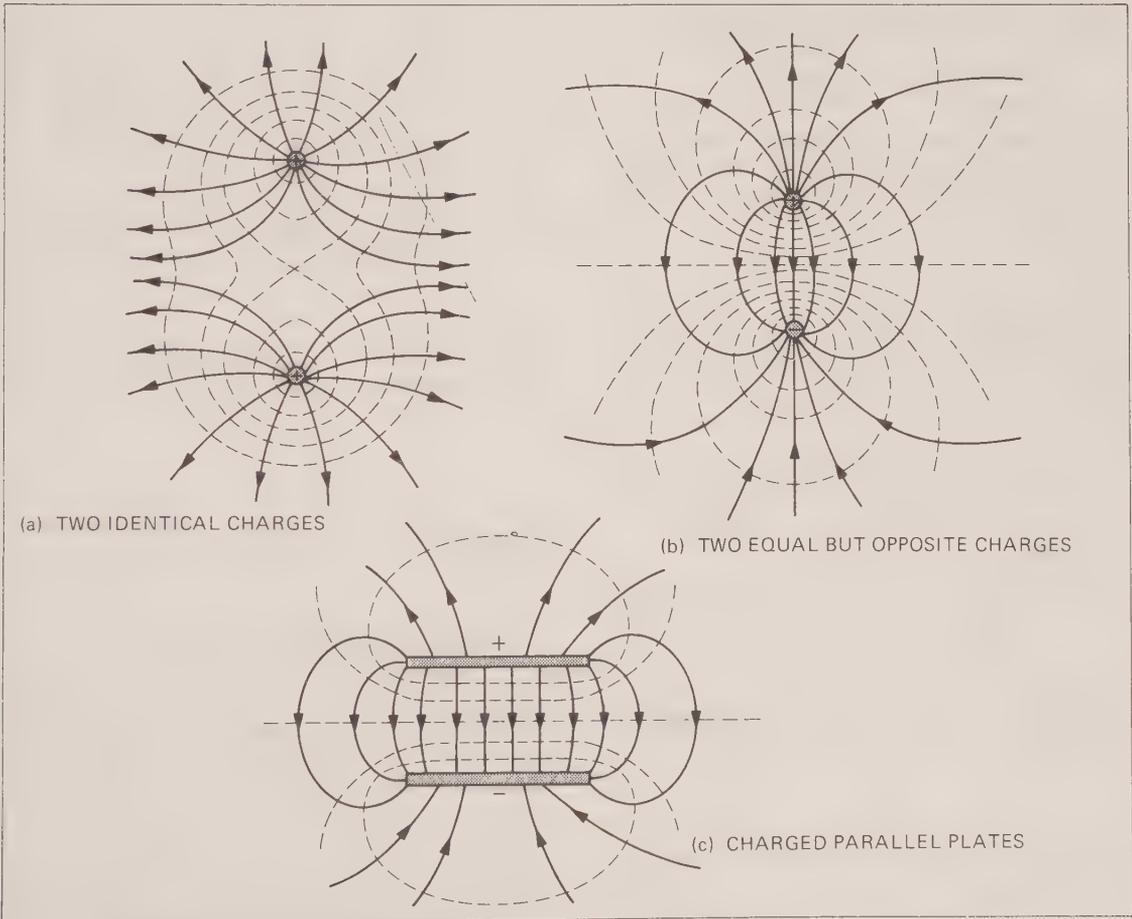


Fig. 45.3. Equipotential surfaces (dotted) and E -field lines.

Notes

(a) The volt metre⁻¹ is alternative to N C⁻¹ as a unit for *E*.

$$1 \frac{\text{V}}{\text{m}} = 1 \frac{\text{J}}{\text{C}} \frac{1}{\text{m}} = 1 \frac{\text{N}}{\text{C}}$$

(b) *dV/dx* represents the **potential gradient** (in the *x*-direction) at a point—this is another term for the resolved part of the electric field strength in that direction. Potential gradient is a *vector* quantity.

(c) The negative sign indicates that although the vectors potential gradient and electric field strength are along the same line, they are oppositely directed.

(d) For a *uniform* field

$$E = -\frac{dV}{dx}$$

can be written $\Delta V = -E\Delta x$

The sign of the potential change is usually inserted by inspection, and we often write

$$\left(\begin{array}{c} \text{change of} \\ \text{potential} \end{array} \right) = \left(\begin{array}{c} \text{uniform} \\ \text{field} \end{array} \right) (\text{distance})$$

(e) It follows from $E = -dV/dx$ that where *E* is zero, *V* is the same at all points. Thus all points inside an empty hollow conductor are at the same electric potential (that of the conductor).

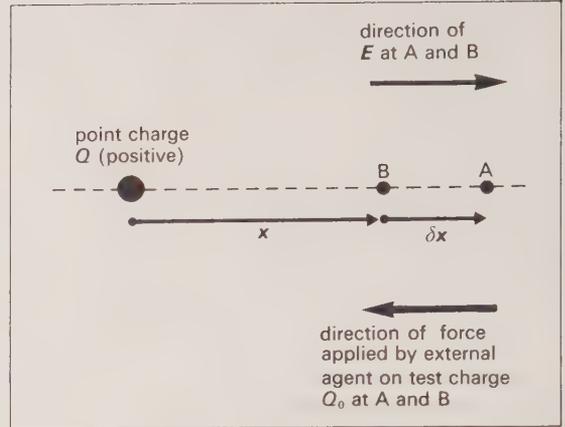


Fig. 45.4. To relate *E* and potential gradient.

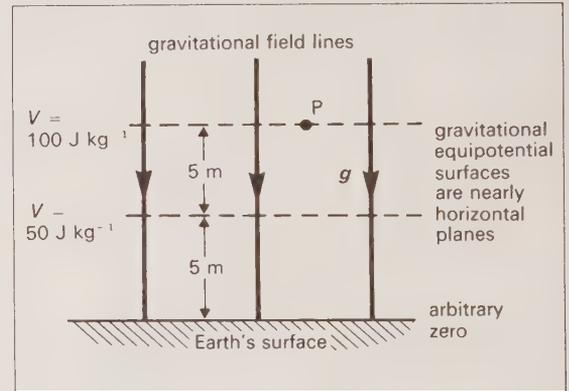


Fig. 45.5. Representation of an approximately uniform gravitational field.

45.3 THE GRAVITATIONAL ANALOGY

Electrostatics	Gravitation
Coulomb's Law $F = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{Q_1 Q_2}{r^2}$	Newton's Law $F = \frac{Gm_1 m_2}{r^2}$
Electric field vector $E = \frac{F}{Q_0}$	Gravitational field vector $g = \frac{F}{m_0}$
Electric potential $V = \frac{W}{Q_0}$	Gravitational potential $V = \frac{W}{m_0}$

Although the gravitational field vector *g* is usually quoted in m s⁻², the N kg⁻¹ is an equivalent unit. In fig. 45.5 we assume it takes the value

$$g = 10 \text{ N kg}^{-1} = 10 \text{ m s}^{-2}$$

vertically downwards at all points.

Example of a gravitational field

If the Earth had a spherically symmetric mass distribution, all points on a sphere concentric with its centre would have the same gravitational potential. (The resolved part of *g* along the tangent to the sphere would be zero.) This is nearly true in practice.

In fig. 45.5, a mass 5 kg placed at P (10 m above the Earth's surface) has 500 J more gravitational potential energy than it would have at the surface.

The gravitational potential difference between P and the surface is

$$\frac{500 \text{ J}}{5 \text{ kg}} = 100 \text{ J kg}^{-1}$$

This could have been found from

$$\begin{aligned} \left(\begin{array}{l} \text{difference of} \\ \text{potential} \end{array} \right) &= (\text{field})(\text{distance}) \\ &= (10 \text{ N kg}^{-1}) \times (10 \text{ m}) \\ &= 100 \text{ J kg}^{-1} \end{aligned}$$

A similar electric field

At all points in fig. 45.6

$$E = 5 \text{ N C}^{-1} = 5 \text{ V m}^{-1} \text{ vertically downwards.}$$

45.4 POTENTIAL DUE TO A POINT CHARGE

► Suppose that in fig. 45.4 the external agent brings the test charge Q_0 from infinity through a vacuum to a point distance r from a point charge Q .

$$\delta W = -EQ_0 \delta x \quad (\text{p. 351})$$

The work W done by the agent

$$\begin{aligned} W &= -\int_{\infty}^r EQ_0 dx \\ &= -\frac{QQ_0}{4\pi\epsilon_0} \int_{\infty}^r \frac{dx}{x^2} \\ &= \frac{QQ_0}{4\pi\epsilon_0} \cdot \frac{1}{r} \end{aligned}$$

Since $V = W/Q_0$, the potential at the point P is

$$V_P = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{Q}{r}$$

N.B. Radial field lines ONLY

Notes

This equation

(a) enables us to calculate the potential due to a point charge, but does not constitute a definition of potential,

(b) demonstrates that the equipotential surfaces associated with a point charge are spheres concentric with the charge, but

(c) takes no account of the potential resulting from any charges induced on nearby conductors.

Potential due to many point charges

Potential is a scalar. If several charges contribute to the potential at a point, the net potential can be found by algebraic addition without regard to direction. This addition may require the use of calculus.

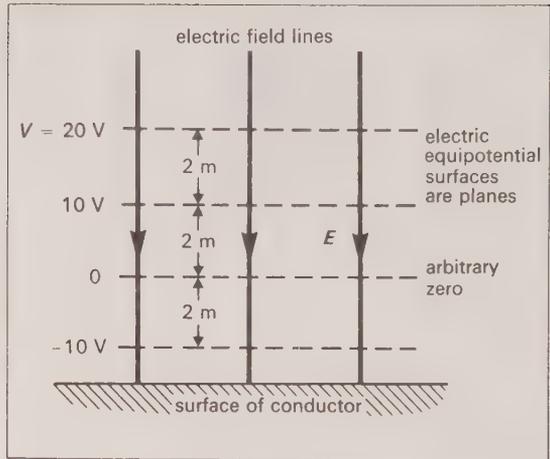


Fig. 45.6. Representation of a uniform electric field.

45.5 A CONDUCTING MATERIAL IN AN ELECTRIC FIELD

Fig. 45.7 (overleaf) shows

(a) a pair of oppositely charged plates which establish a uniform field between them,

(b) a $V-x$ and (c) an $E-x$ graph, both before and after a conducting plate is inserted as shown.

The insertion of the conductor is accompanied by the flow of an induced charge. Opposite sides of the plate acquire opposite charges so that E inside the plate is again zero at equilibrium. The plate's net charge remains zero. Note the effect of the plate on the values of V and E in the air gap.

45.6 FIELD AND POTENTIAL DUE TO AN INSULATED CONDUCTING CHARGED SPHERE

(A) Field

Suppose we have a sphere of radius a .

We have shown (p. 347 and p. 346)

$$E = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{Q}{r^2} \quad \text{for points outside}$$

and $E = 0$ for points inside

(B) Potential

If the sphere is isolated, then (from 45.4)

$$V = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{Q}{r} \quad \text{for points outside}$$

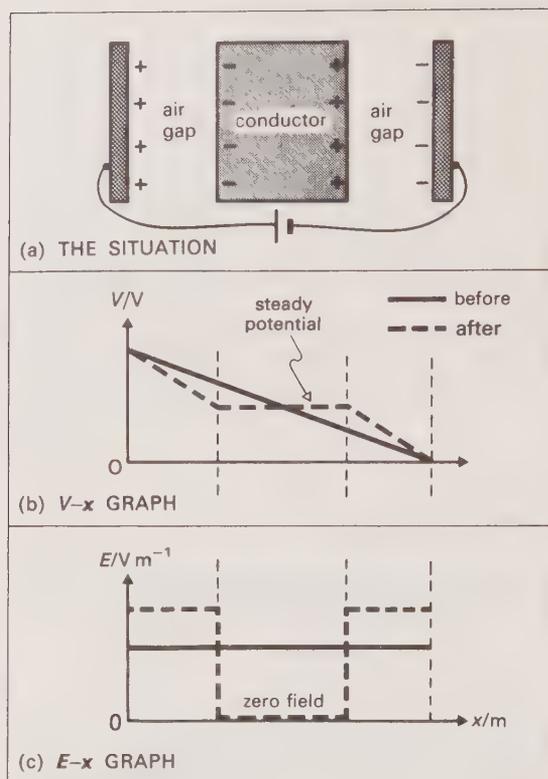


Fig. 45.7. The effect of a conducting plate on values of E and V nearby.

and $V = \left(\frac{1}{4\pi\epsilon_0}\right) \cdot \frac{Q}{a}$ for every point inside

this being the potential of the sphere.

These ideas are brought out in fig. 45.8.

Conductor not isolated

(a) The *charge* on a conductor can only be altered by electrical contact.

(b) The *potential* of a conductor is strongly influenced by its environment. We use this fact in designing a capacitor (p. 359).

Examine fig. 45.9 carefully.

► 45.7 POTENTIAL DIFFERENCES CLOSE TO A LINE OF CHARGE

Gauss's Law was used in 44.9 to show that the magnitude of the radial electric field vector at a distance r from an infinitely long line of charge is given by $E = \lambda/2\pi\epsilon_0 r$, where λ is the linear charge

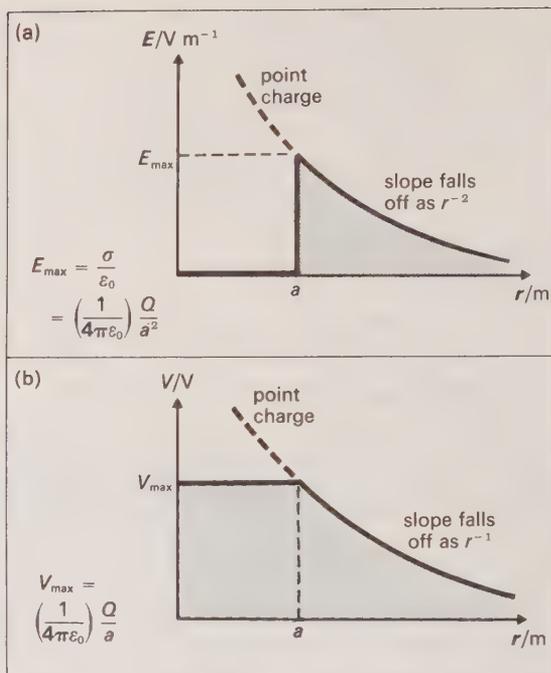


Fig. 45.8. (a) Field and (b) potential established by a charged conducting sphere. (Cf. fig. 20.8.)

density. Using $E = -(dV/dx)$, we find the p.d. between two points distant r_1 and r_2 from the rod ($r_2 > r_1$) to be given by

$$V_2 - V_1 = - \int_{r_1}^{r_2} E \cdot dr$$

If the rod is positively charged then E and r have the same direction.

$$V_2 - V_1 = - \int_{r_1}^{r_2} \frac{\lambda}{2\pi\epsilon_0} \cdot \frac{dr}{r} = - \frac{\lambda}{2\pi\epsilon_0} \ln\left(\frac{r_2}{r_1}\right)$$

$$\text{or } V_1 - V_2 = \left(\frac{\lambda}{2\pi\epsilon_0}\right) \ln\left(\frac{r_2}{r_1}\right)$$

Notes

(a) If V_2 is taken to be zero when $r_2 = \infty$, the above equation requires infinite potential for all finite values of r_1 . This physical impossibility stems from our postulating an *infinite* line of charge, which cannot be realized in practice.

(b) The equation remains useful nevertheless for calculating potential differences between pairs

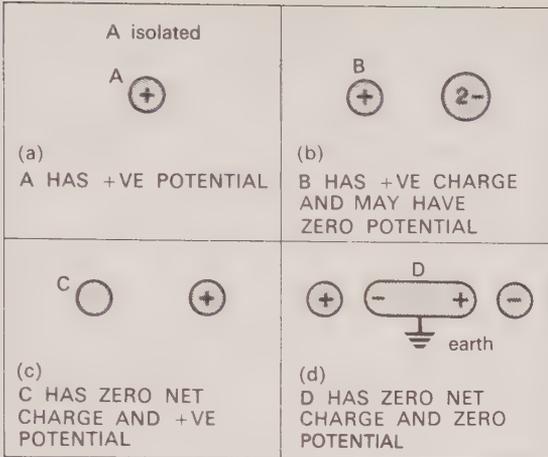


Fig. 45.9. The potential of a body is not controlled by its own charge only.

of points relatively close to a *finite* line of charge. (See the coaxial capacitor on p. 360.)

45.8 CHARGE DISTRIBUTION OVER A CONDUCTOR

It can be shown that the **surface density of charge** on a conductor of varying radii of curvature (fig. 45.10) is inversely proportional to the radius of curvature although the potential is the same at all points whatever the shape.

Thus $\sigma [C\ m^{-2}]$ is large where the curvature is large. But $E = \sigma/\epsilon_0$, so the field is also large. The

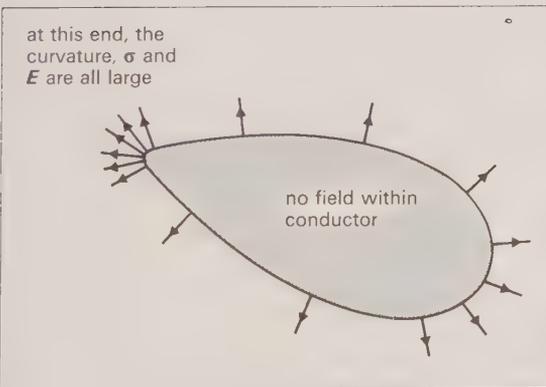


Fig. 45.10. Variations of σ and E over a surface.

charge distribution on a conductor can be investigated experimentally by transferring samples of charge with a **proof-plane** (a small metal disc on an insulating handle) from different parts of the conductor to a d.c. amplifier for measurement. The charge taken up by the proof-plane is directly proportional to the charge density at the point touched because the proof-plane becomes part of the surface during contact.

At a *sharp point* E becomes very high, and this may

- (a) attract or repel ionized particles from the air, or
- (b) in extreme cases *ionize* air molecules.

Movement of charge then neutralizes the charge on the conductor: the excess charge seems to stream off into the air. This is called the **action of points**. It is utilized in the lightning conductor and the field ion microscope. **Corona discharge** is the extreme process.

45.9 THE VAN DE GRAAFF GENERATOR

In fig. 45.11 the comb at A is maintained at a positive potential, and so the moving belt acquires a positive charge. (In a small laboratory version friction between the roller and the belt serves to maintain charge on the belt.) This positive charge induces a negative charge on comb B, which streams onto the belt. The net effect is to transfer positive charge from the belt to the *outside* of the conducting sphere (p. 345).

The limit to the potential that the sphere can acquire is controlled by the air breakdown.

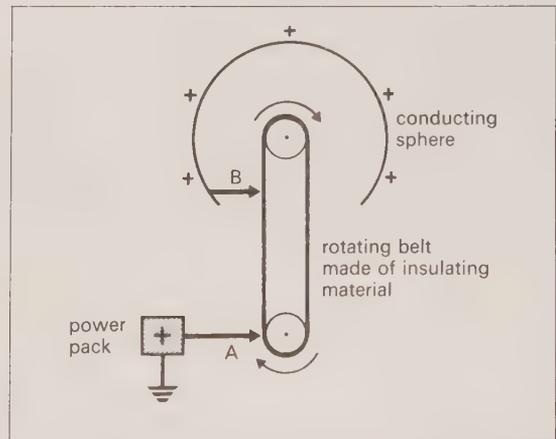


Fig. 45.11. The essentials of the Van de Graaff generator.

45.10 THE GOLD LEAF ELECTROSCOPE

Its construction

There are two basic components:

- (a) the *cap, spine* and *leaves*, which are connected conductors at the same potential, and
- (b) the *case*, also a conductor, but insulated from (a), and usually earthed.

Why the leaves diverge

The leaves diverge when they experience a force because they are in the electric field established between the case (zero potential) and themselves. The presence of this field indicates a potential difference between the case and leaves: the *electroscope is a device which measures this potential difference*. The leaves diverge until they reach an equilibrium position. Suppose

- (a) the case is earthed: then *indirectly* the electroscope measures the *charge* on the cap, since a greater charge leads to a greater p.d. between leaves and case, a greater field, and a greater deflection;
- (b) the case is connected to the cap: a charge placed on the cap increases the potential of *both* basic components *equally*. The leaves are in zero field, and so experience zero force.

Charging the electroscope

- (a) *on contact*, the cap, spine and leaves acquire charge of the same sign as the charging body.
- (b) *By induction*, they ultimately acquire charge of opposite sign. Study fig. 45.12 carefully, in which the case is earthed.

► 45.11 THE ATTRACTED DISC ELECTROMETER

Force on the surface of a charged conductor

In fig. 45.13 we deduce

$$(a) \text{ net field outside} = \frac{\sigma}{2\epsilon_0} + \frac{\sigma}{2\epsilon_0} = \frac{\sigma}{\epsilon_0}$$

$$(b) \text{ net field inside} = \frac{\sigma}{2\epsilon_0} - \frac{\sigma}{2\epsilon_0} = 0$$

These both satisfy the requirements of p. 348 and p. 346.

The outward force exerted on elementary area ΔA is given by

$$F = E'Q$$

where E' is the field set up by the *rest* of the conductor.

$$F = \frac{\sigma}{2\epsilon_0} \cdot (\Delta A\sigma)$$

$$F = \frac{\Delta A\sigma^2}{2\epsilon_0}$$

$$F[\text{N}] = \frac{1}{2} \frac{\Delta A[\text{m}^2]\sigma^2 \left[\frac{\text{C}^2}{\text{m}^4} \right]}{\epsilon_0 \left[\frac{\text{C}^2}{\text{N m}^2} \right]}$$

It will act outwards whatever the sign of σ .

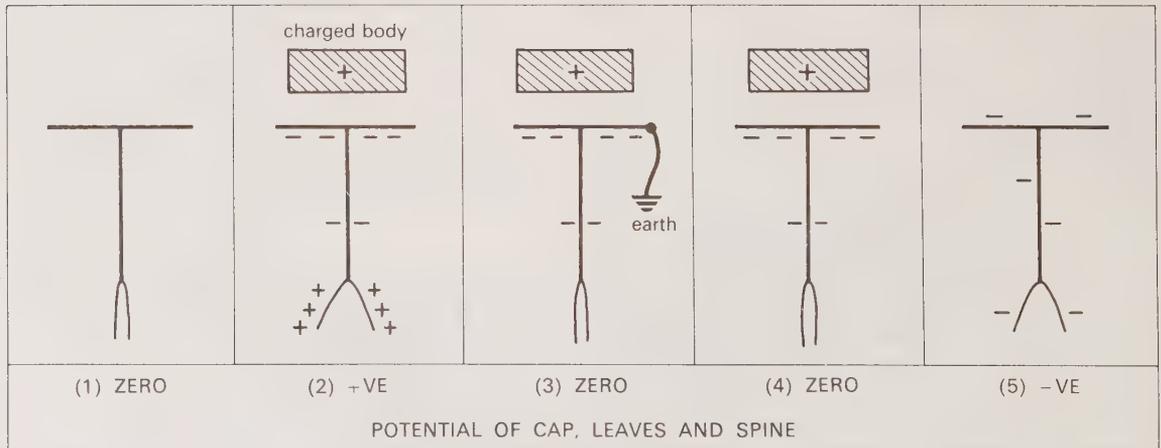


Fig. 45.12. The gold leaf electroscope: successive stages during charging by induction.

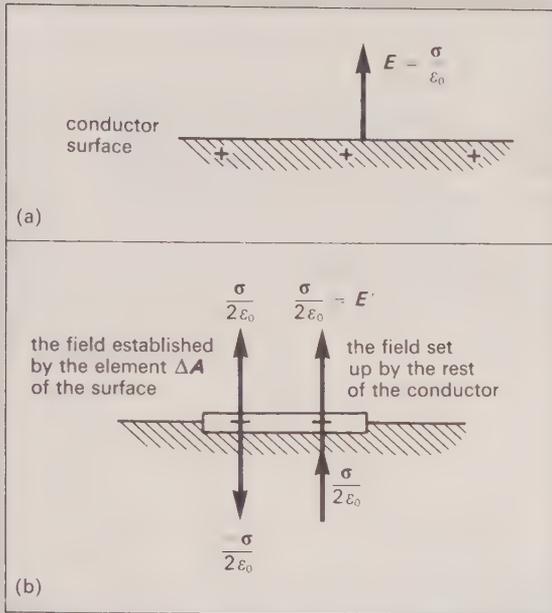


Fig. 45.13. Analysis of the field at the surface of a conductor: (a) the total field, (b) components of the field.

The principle of the electrometer

This is an **absolute** instrument which enables a potential difference to be measured in terms of mechanical quantities. Fig. 45.14 shows only the principle on which it works.

Procedure

- (1) Arrange the upper plate to be in the plane of the guard-ring (p. 360), whose function is to create a uniform field.
- (2) Apply a potential difference V between the plates.
- (3) Remove a mass m to return the central plate to its original height. This compensates for the downward electrical force. Then

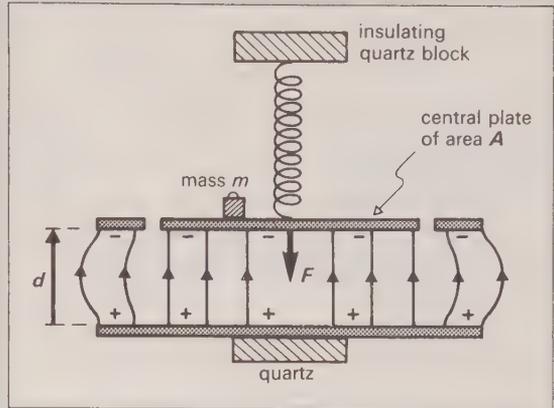


Fig. 45.14. The essentials of the attracted disc electrometer.

$$F = mg = \frac{A\sigma^2}{2\epsilon_0}$$

in which $\frac{\sigma}{\epsilon_0} = E$, and $V = Ed$

$$\text{so } mg = \frac{\epsilon_0 V^2 A}{2d^2}$$

V can be calculated from this equation, since m , g , d , ϵ_0 and A are known.

$$\epsilon_0 = 9 \times 10^{-12} \text{ C}^2 \text{ N}^{-1} \text{ m}^{-2}$$

so F is very small. Typically the electrometer measures potential differences of a few kilovolts. (Note. ϵ_0 does not have a defined value. This means that the electrometer may be regarded as *absolute* only if the assumed value of ϵ_0 is measured by the method involving the speed of light (p. 361), since this requires only mechanical measurements.)

46 CAPACITORS

46.1 CAPACITANCE FOR CHARGE	358	46.6 COMPARISON OF CAPACITANCES	362
46.2 EARTHING	359	46.7 COMBINATIONS OF CAPACITORS	362
46.3 CAPACITORS	359	46.8 ENERGY OF A CHARGED CONDUCTOR	363
46.4 CALCULATION OF CAPACITANCE	360	46.9 THE D.C. AMPLIFIER	364
46.5 MEASUREMENT OF CAPACITANCE AND ϵ_0	361		

46.1 CAPACITANCE FOR CHARGE

Suppose we put a charge Q on an insulated conductor, and that its potential changes by V .

Then we define the **capacitance** C of the conductor by the equation

$$C = \frac{Q}{V}$$

$$C \left[\frac{C}{V} \right] = \frac{Q[C]}{V[V]}$$

Notes

(a) The unit of capacitance, a *coulomb volt⁻¹*, occurs so often that it is convenient to give it the name-**farad** (F). The farad is an extremely large unit, and we use more often

- (i) the *microfarad* $1 \mu\text{F} = 10^{-6} \text{ F}$, and
- (ii) the *picofarad* $1 \text{ pF} = 10^{-12} \text{ F}$

(b) The change in V brought about by a given change in Q depends upon a conductor's size, the nearness of other conductors, and the material surrounding it. It can be shown theoretically and experimentally that, for a given environment,

$$Q \propto V$$

i.e. that C is a *constant*.

(c) Capacitance is numerically equal to the charge in coulombs that changes the potential of a conductor by one volt.

$$[C] = [M^{-1}L^{-2}T^4I^2]$$

(d) Note the analogy with *heat capacity*:

- (i) For heat $\Delta Q = C \Delta \theta$ (p. 189)
- (ii) For electric charge $Q = CV$

where V represents a *change* in potential resulting from a change Q in the total charge.

Example. For an isolated conducting sphere, of radius a and carrying a charge Q ,

$$V = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{Q}{a} \quad (\text{p. 353})$$

so
$$C = \frac{Q}{V} = 4\pi\epsilon_0 a$$

Capacitance always has dimensions

$$[\epsilon_0] [L^1]$$

A sphere of radius $9 \times 10^9 \text{ m}$ (more than $10^3 \times$ that of the Earth) would have a capacitance of about one farad.

The unit of the permittivity constant

Compare
$$\epsilon_0 = \frac{C}{4\pi a}$$

(from the capacitance of an isolated sphere)

with
$$\epsilon_0 = \frac{Q^2}{4\pi F r^2}$$

(from *Coulomb's Law*).

The first equation shows that a possible unit for ϵ_0 is the *farad metre⁻¹*: this unit, which is equivalent to the $C^2 N^{-1} m^{-2}$, is the one usually quoted.

$$\begin{aligned}
 1 \text{ C}^2 \text{ N}^{-1} \text{ m}^{-2} &= 1 \text{ C} (\text{C N}^{-1} \text{ m}^{-1}) \text{ m}^{-1} \\
 &= 1 \text{ C V}^{-1} \text{ m}^{-1} \\
 &= 1 \text{ F m}^{-1}
 \end{aligned}$$

46.2 EARTHING

When a small conducting body shares its excess charge by contact with a large conducting neutral body, it loses most of it. The two bodies acquire the same potential, which is less than that of the small body's original potential.

The Earth has a capacitance given by

$$\begin{aligned}
 C &= 4\pi\epsilon_0 a \\
 &= (\frac{1}{9} \times 10^{-9} \text{ F m}^{-1}) \times (6 \times 10^6 \text{ m}) \\
 &= 7 \times 10^{-4} \text{ F}
 \end{aligned}$$

since it may be regarded as a conducting body.

This capacitance is so large that we say isolated bodies that are **earthed** (i.e. share their charge with the Earth) effectively lose that charge. The body and the Earth acquire a common potential which we call the *practical zero of potential*.

46.3 CAPACITORS

The charge that can be put onto a conductor is limited by the large surface fields that develop, which result in electrical breakdown, sparking and consequent discharge.

A **capacitor** is a system of high capacitance which has been designed for the storage of separated positive and negative charges.

In fig. 46.1 the potential of each conductor is affected by the presence of the other—the *p.d.* between them is reduced.

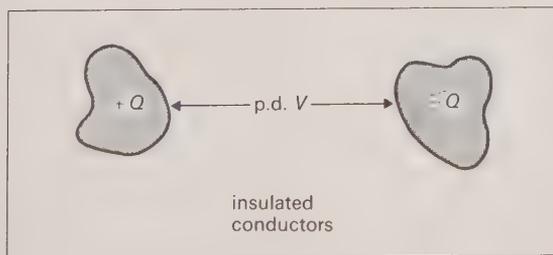


Fig. 46.1. The principle of the capacitor.

The capacitance of the system is

$$\begin{aligned}
 C &= \frac{\text{magnitude of either charge}}{\text{p.d. between conductors}} \\
 &= Q/V
 \end{aligned}$$

Thus for a fixed Q , C is made large by lowering V : this is achieved by bringing a second conductor, which is frequently earthed, very close to the first conductor.

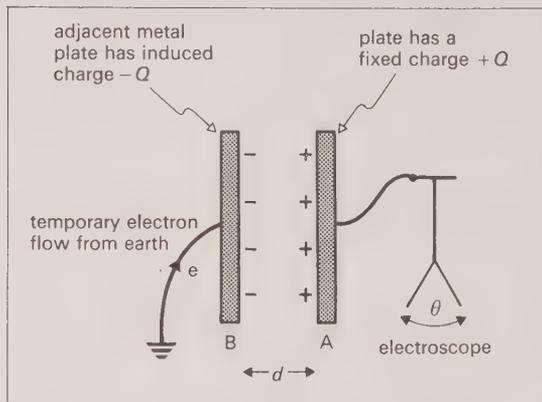


Fig. 46.2. Investigation of the parallel-plate capacitor.

Demonstration of the parallel-plate capacitor

The negative charge induced on plate B (fig. 46.2) attracts and tends to anchor the positive charge on plate A. Since both B and the electrometer case are earthed,

$$\theta \propto V, \text{ the p.d. between the plates}$$

But $C = Q/V$, and Q is fixed, so

$$C \propto 1/\theta$$

If we

- (a) decrease d , then θ decreases, so C increases,
- (b) increase the common plate area A , θ decreases and so C increases.

We have shown experimentally, in a general way

$$C \propto A/d$$

Because the plates attract (as in the electrometer, p. 357) they are kept apart by an insulator (p. 366).

The importance of capacitors

(a) They can be used to establish electric fields of required pattern, e.g. for

- (i) Millikan's experiment, (p. 472)
- (ii) the study of dielectric materials. (p. 365)

- (b) They store not only charge, but *energy* in the electric field between the plates.
- (c) They form an integral part of most a.c. circuits, e.g. in
 - (i) smoothing circuits, (p. 456)
 - (ii) oscillatory circuits, and (p. 486)
 - (iii) time bases. (p. 474)

Capacitors in practice

- (a) *Tuning capacitors* (in radio receivers) have a variable *A* and so a variable *C*.
- (b) *Electrolytic capacitors* have a small *d* and so a large *C*.
- (c) The *guard-ring capacitor* prevents distortion of the electric field lines.

In fig. 46.3 the ring and plate are maintained at the same potential, but are not joined so their charges are kept separate.

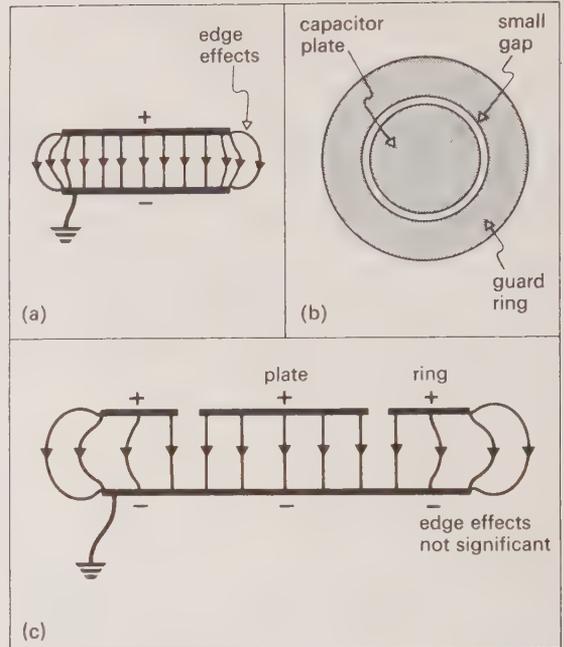


Fig. 46.3. The guard-ring capacitor: (a) the distorted field, (b) plan view, (c) side view (different scale).

46.4 CALCULATION OF CAPACITANCE

Procedure

- (1) Imagine a charge (+*Q*) to be put on one conductor. Then in an *ideal capacitor* a charge (-*Q*) is induced on the other.
- (2) Calculate the change of potential *V* that results.
- (3) Apply the defining equation $C = Q/V$.

The result should have dimensions

$$[\epsilon_0][L^1]$$

(2) is the difficult step, and can be done only when the geometry is simple.

(A) The parallel-plate capacitor

- (1) Refer to fig. 46.4.
- (2) Using $V = Ed$ (p. 351)

$$V = \frac{\sigma}{\epsilon_0}d \quad \text{since} \quad E = \frac{\sigma}{\epsilon_0} \quad (\text{p. 348})$$

$$(3) \quad C = \frac{Q}{V} = \frac{A\sigma}{\frac{\sigma}{\epsilon_0}d}$$

$$C = \frac{\epsilon_0 A}{d}$$

As required, the dimensions of *C* are $[\epsilon_0][L^1]$

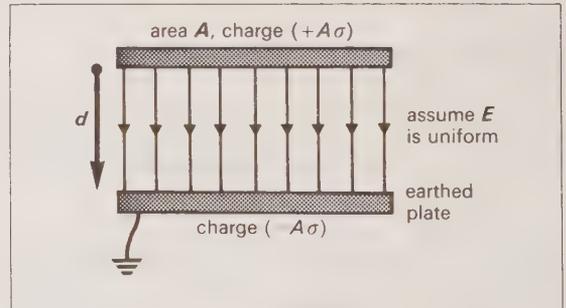


Fig. 46.4. The parallel-plate capacitor.

► (B) The coaxial cylindrical capacitor

- (1) Refer to fig. 46.5.
- (2) In 45.7 the p.d. between two points close to a line of charge was shown to be

$$V_1 - V_2 = \frac{\lambda}{2\pi\epsilon_0} \ln \frac{r_2}{r_1}$$

The p.d. between a point on cylinder A and a point on cylinder B is given by

$$V = \frac{\lambda}{2\pi\epsilon_0} \ln \frac{b}{a} = \frac{Q}{2\pi\epsilon_0 l} \ln \frac{b}{a}$$

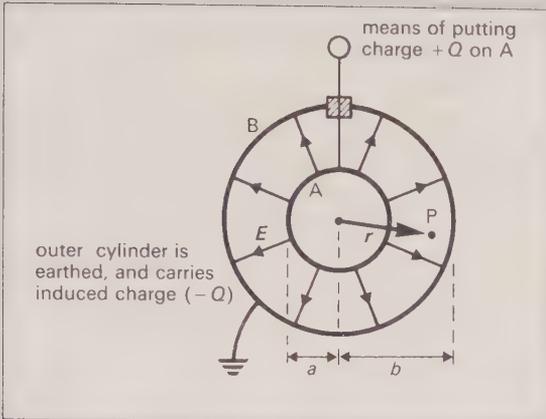


Fig. 46.5. Plane section perpendicular to the axis of a coaxial cylindrical capacitor of length l .

$$(3) \quad C = \frac{Q}{V} = \frac{2\pi\epsilon_0 l}{\ln(b/a)}$$

This has the required dimensions. Except at the ends, where fringing occurs, it enables us to calculate the capacitance per unit length of coaxial cables.

(C) The concentric spheres capacitor

Using the symbols of fig. 46.5 one can show

$$C = 4\pi\epsilon_0 \left(\frac{ab}{b-a} \right)$$

46.5 MEASUREMENT OF CAPACITANCE AND ϵ_0

The measurement of ϵ_0

Although ϵ_0 is defined from the statement of *Coulomb's Law* (p. 337), the situation cannot be used for a direct experimental determination of its size.

(a) The determination of C for a parallel-plate capacitor in the school laboratory (below) enables ϵ_0 to be calculated from $C = \epsilon_0 A/d$, since we need measure only A and d in addition.

(b) *Sir J. J. Thomson* used a cylindrical capacitor with cylindrical guard rings at the ends (1890).

(c) *Rosa and Dorsey* used a concentric spherical capacitor (1907), and obtained an uncertainty of 1 in 10^4 .

(d) Experiments of this kind have all confirmed *Maxwell's* prediction that $c = 1/\sqrt{(\epsilon_0\mu_0)}$. The most accurate way of evaluating ϵ_0 is to assume the truth of this equation, and to calculate the value of

ϵ_0 from the defined value of μ_0 and the measured value of c . $\epsilon_0 = 8.854 \text{ pF m}^{-1}$, with an estimated uncertainty of 7 in 10^9 (1973).

Measurement of capacitance

Refer to fig. 46.6. The capacitor is charged at a frequency f to the p.d. V across the supply, and each time discharged through the sensitive ammeter. At sufficiently high values of f the meter reading is steady.

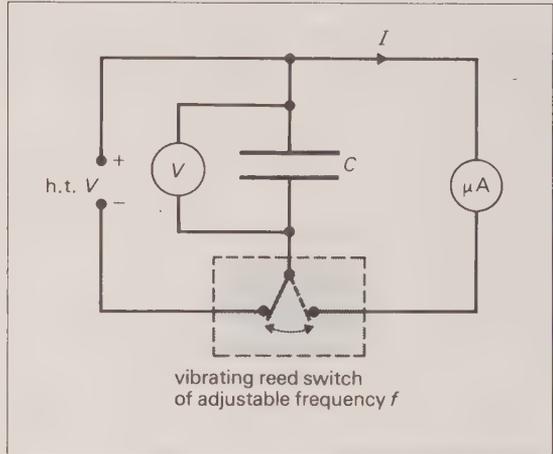


Fig. 46.6. The basic principle of the vibrating reed capacitance meter.

During each time interval $1/f$, a charge $Q = CV$ is passed through the meter.

$$\therefore I = \frac{Q}{(1/f)} = fCV$$

The graph of fig. 46.7 shows the experimental results, from which C is found.

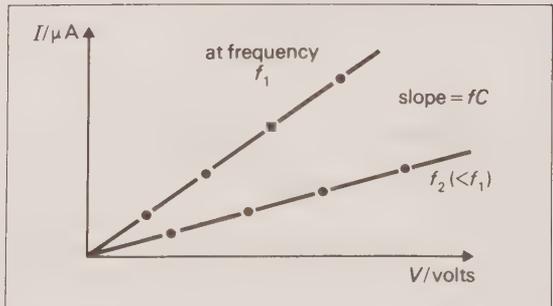


Fig. 46.7. Experimental results.

46.6 COMPARISON OF CAPACITANCES

(A) *By d.c. amplifier* (p. 364)

Since $Q = CV$

$$\frac{dQ}{dt} = I = C \frac{dV}{dt}$$

$$= C \left(\frac{V}{t} \right)$$

if I is steady. When the switch is closed in fig. 46.8 the value of R is adjusted to ensure a steady current.

Suppose that after time t the switch is opened, and for capacitor C_1 the voltmeter reading is V_1 .

The same current passed for the same time into C_2 will give a reading V_2 . Then

$$I = C_1 \left(\frac{V_1}{t} \right) = C_2 \left(\frac{V_2}{t} \right)$$

from which C_1/C_2 can be found.

(B) *By ballistic galvanometer* (p. 421)

(C) *By vibrating reed capacitance meter*
(fig. 46.6)

For fixed values of V and f the value of C is proportional to I .

(D) *By the De Sauty bridge* (fig. 46.9)

When the sound in the headphones is adjusted to a minimum value, the bridge is balanced.

$$V_{AB} = V_{AD} \quad \text{and} \quad V_{BC} = V_{DC}$$

so $I_1 \left(\frac{1}{\omega C_1} \right) = I_2 R_1$ and $I_1 \left(\frac{1}{\omega C_2} \right) = I_2 R_2$

giving $\frac{C_1}{C_2} = \frac{R_2}{R_1}$

Note that ω cancels from the final equation.

These methods enable the results of 46.7 to be tested experimentally.

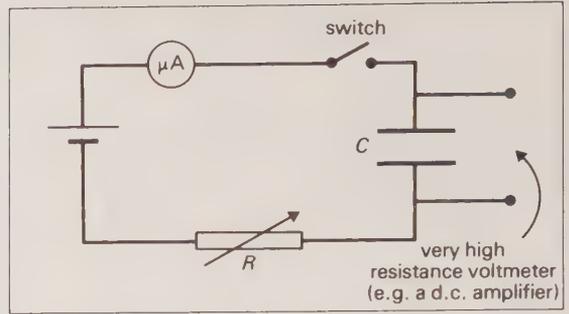


Fig. 46.8. Measurement of capacitance using a steady charging current.

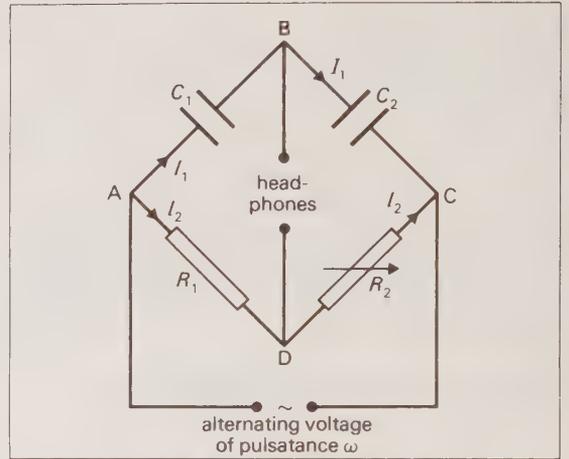


Fig. 46.9. The De Sauty bridge for comparing capacitances.

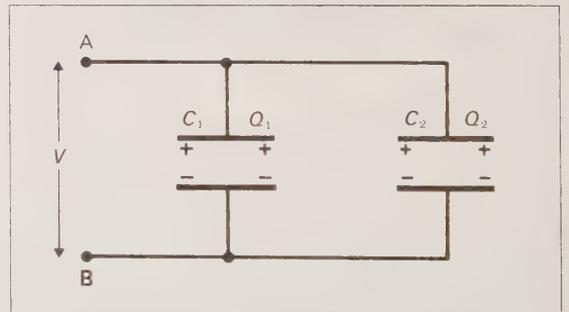


Fig. 46.10. Capacitors in parallel.

46.7 COMBINATIONS OF CAPACITORS

(A) *Capacitors in parallel*

In fig. 46.10 we want the capacitance C of the single capacitor which, connected across AB , would have the same effect as C_1 and C_2 connected as shown.

(a) We have the same p.d. V across each.

(b) $Q = Q_1 + Q_2$

so $CV = C_1V + C_2V \quad (Q = CV)$

$C = C_1 + C_2$

Capacitors in parallel

The result can be extended to any number of capacitors. (Compare with resistors in series, p. 383.)

(B) Capacitors in series

In fig. 46.11 we suppose the charging process has not involved any connection other than those shown, so that (a) below applies.

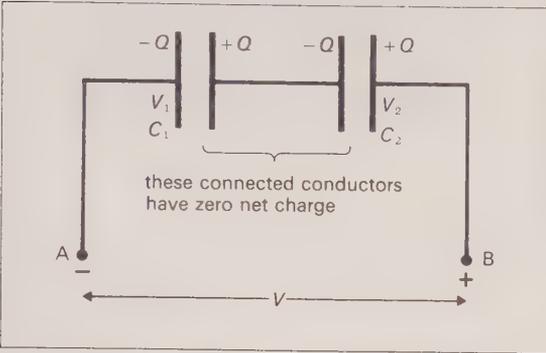


Fig. 46.11 Capacitors in series.

(a) The magnitude Q of the charge induced on each capacitor plate is the same.

(b) $V = V_1 + V_2$

$$\frac{Q}{C} = \frac{Q}{C_1} + \frac{Q}{C_2} \quad \left(V = \frac{Q}{C} \right)$$

$$\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}$$

Capacitors in series

The result can be extended to any number of capacitors. The capacitance of the combination is less than C_1 or C_2 . (Compare with resistors in parallel, p. 384.)

46.8 ENERGY OF A CHARGED CONDUCTOR

Electric potential energy is the work done in assembling a particular configuration of charges (e.g. in putting charge onto a conductor). As a conductor is charged, its potential changes, and it becomes more difficult to add a further charge increment ΔQ .

► Suppose we bring charge δQ from infinity to a conductor already at potential V . Work done $\delta W = V\delta Q$. Total work W done in raising the charge from 0 to Q_0 is

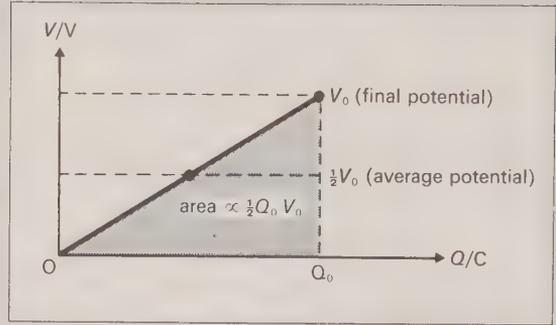


Fig. 46.12. The charging of a conductor.

$$\begin{aligned} W &= \int \delta W = \int_0^{Q_0} V dQ \\ &= \int_0^{Q_0} \frac{Q}{C} \cdot dQ \quad \left(V = \frac{Q}{C} \right) \\ &= \frac{1}{2} \frac{Q_0^2}{C}, \quad \text{since } C \text{ is constant} \end{aligned}$$

$$W = \frac{1}{2} Q_0 V_0 = \frac{1}{2} \frac{Q_0^2}{C} = \frac{1}{2} V_0^2 C$$

$$W[J] = \frac{1}{2} Q_0[C] \times V_0 \left[\frac{J}{C} \right]$$

Fig. 46.12 illustrates the change of potential with charge during the charging process. The energy stored, W , is represented by the area under the graph.

Loss of electrical potential energy on joining two capacitors

On depressing S in fig. 46.13, there will be sparking or flow of charge (current) until the same p.d. exists across each capacitor.

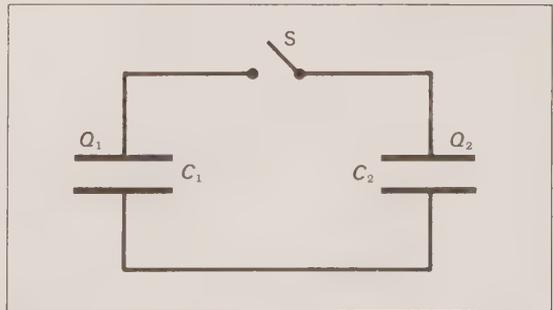


Fig. 46.13. Joining two charged capacitors.

$$\text{Initial energy} = \frac{\frac{1}{2}Q_1^2}{C_1} + \frac{\frac{1}{2}Q_2^2}{C_2}$$

$$\text{Final energy} = \frac{\frac{1}{2}Q^2}{C} = \frac{\frac{1}{2}(Q_1 + Q_2)^2}{C_1 + C_2}$$

Subtract and rearrange terms: loss of electrical potential energy (gain of (e.g.) internal energy)

$$= \frac{\frac{1}{2}(C_2Q_1 - Q_2C_1)^2}{C_1C_2(C_1 + C_2)}$$

Notes

(a) There can never be a gain of electrical p.e. (this is illustrated by the squared term).

(b) The electrical p.e. loss is zero when

$$C_2Q_1 = Q_2C_1$$

or
$$\frac{Q_1}{C_1} = \frac{Q_2}{C_2}$$

i.e. if
$$V_1 = V_2$$

Then there is no movement of charge.

46.9 THE D.C. AMPLIFIER

An **electrometer** is a measuring instrument that operates on a current of 1 pA or less. The type shown in fig. 46.14 uses a very high-gain d.c. amplifier. This instrument amplifies a very small input p.d. to an output p.d. which, being perhaps 10^{10} times as big, is large enough to operate a $100 \mu\text{A}$ moving-coil meter.

Uses

(a) *To measure p.d.* The microammeter connected across the output can be calibrated to measure the *input* p.d. An input p.d. of about 1 volt would normally give full scale deflection (f.s.d.).

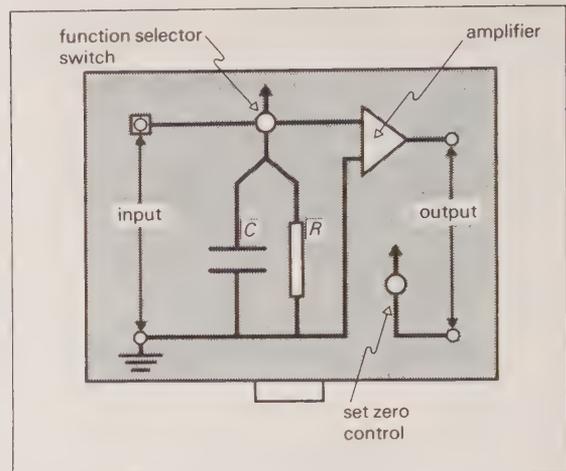


Fig. 46.14. A d.c. amplifier.

(b) *To measure current.* Although the amplifier is basically a voltmeter of impedance about $10 \text{ T}\Omega$, it can be adapted to measure small currents of about 1 pA if these are made to flow through known resistances. (For an example see p. 521.)

(c) *To measure charge.* The method was described on p. 338.

Notes

(a) Types that are not mains-operated are provided with a *gain control* to compensate for the ageing of the amplifier's batteries.

(b) A range of capacitances and resistances can be built into the electrometer.

(c) A *short* button discharges the internal components between readings.

47 DIELECTRIC MATERIALS

47.1	PARALLEL-PLATE CAPACITOR WITH DIELECTRIC	365
47.2	MOLECULAR EXPLANATION OF DIELECTRIC BEHAVIOUR	366
▶ 47.3	DIELECTRICS AND COULOMB'S LAW	367

▶ 47.4	DIELECTRICS AND GAUSS'S LAW	367
▶ 47.5	ELECTRIC DISPLACEMENT D	368
47.6	ENERGY CHANGES IN A CAPACITOR	368

47.1 PARALLEL-PLATE CAPACITOR WITH DIELECTRIC

(The following results can also be shown with the apparatus of fig. 46.2, in which the electroscope is used as an electrostatic voltmeter.)

(A) Capacitors with same $p.d.$

We can use a ballistic galvanometer (fig. 47.1) to show $Q_d > Q_0$, and deduce that the capacitor with the dielectric has the greater capacitance; $C_d > C_0$.

(B) Capacitors with same charge

Experimentally (fig. 47.2) we can show that $V_0 > V_d$, and deduce that $C_d > C_0$, as before.

Suppose the dielectric material fills the space between the plates.

Then we define the **relative permittivity** ϵ_r by the equation

$$\epsilon_r = \frac{C_d}{C_0}$$

(ϵ_r is sometimes called the *dielectric constant*.)

Typical values of ϵ_r are given in the table on p. 366. ϵ_r is dimensionless, and has no unit.

(a) For a capacitor with V kept constant

$$\epsilon_r = \frac{C_d}{C_0} = \frac{Q_d}{Q_0}$$

(b) For Q kept constant

$$\epsilon_r = \frac{C_d}{C_0} = \frac{V_0}{V_d} = \frac{E_0}{E_d}$$

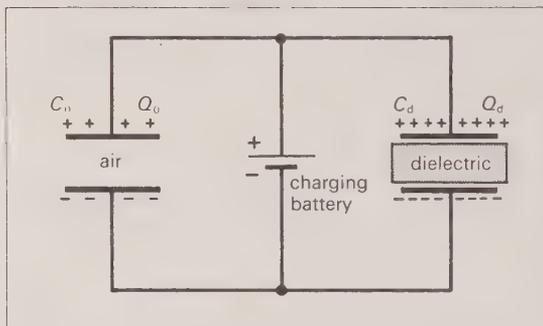


Fig. 47.1. Two capacitors connected to the same battery.

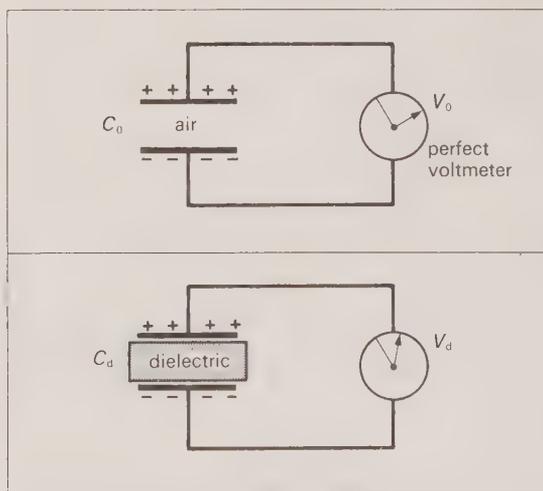


Fig. 47.2. Two capacitors with the same charge.

This idea is discussed in the next section, where E_0 and E_d are defined.

The **electric permittivity** ϵ is defined by the equation

$$\epsilon = \epsilon_r \epsilon_0$$

ϵ is sometimes called the **absolute permittivity**. It has the same dimensions and unit as ϵ_0 .

Capacitors and dielectrics

Dielectric materials are used in capacitors

- (a) to keep the plates apart,
- (b) to raise the capacitance, relative to air, by a factor of about ϵ_r , and
- (c) to reduce the chance of electric breakdown, and thus to enable a large V to be used.

To achieve this they require a large **dielectric strength**, which is the largest potential gradient that can exist in the material.

Material	Relative permittivity ϵ_r , measured at zero frequency	Dielectric strength 10^6 V m^{-1}
vacuum	1 exactly	—
air (at s.t.p.)	1.0006	3
water	80	(conductor)
mica	5	150
polyethylene	2	50

Our capacitance equations become

$$(1) \text{ Parallel-plate} \quad C = \frac{\epsilon A}{d}$$

$$(2) \text{ Coaxial cylinder} \quad C = \frac{2\pi\epsilon l}{\ln(b/a)}$$

ϵ has replaced ϵ_0 .

Measurement of ϵ_r

(a) If the dielectric material fills the space between the plates (as would a fluid), use

$$\epsilon_r = \frac{C_d}{C_0}$$

(b) If the dielectric is a solid slab which does *not* fill the space, measure C_0 and C_d , and use the equation derived in the next section.

47.2 MOLECULAR EXPLANATION OF DIELECTRIC BEHAVIOUR

(a) In a **polar substance** the centre of negative charge in a molecule does not coincide with the centre of positive charge. A molecule has a *permanent electric dipole moment* (p. 342).

In a uniform electric field, the molecules experience

- (i) a torque that tends to align them (see fig. 43.6), but also
- (ii) a tendency to random orientation because of their thermal motion.

Their partial alignment results in a field E' (fig. 47.4) in *opposition* to an external field E_0 .

(b) **Non-polar substances** acquire a temporary *induced* dipole moment in an external field (fig. 47.3).

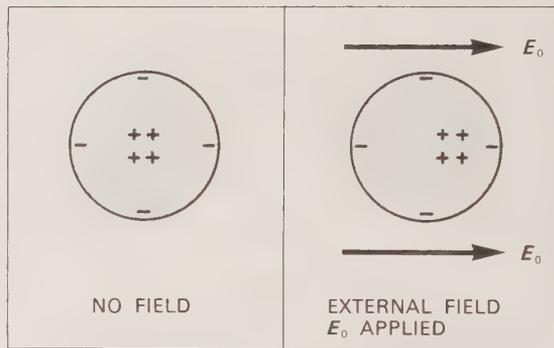


Fig. 47.3. Behaviour of a non-polar molecule.

The temporary dipoles are parallel to E_0 , and establish an opposing field E' .

In both (a) and (b) the material is said to be **polarized** by the *microscopic* charge movement. Any small sample of material (other than at the surface) remains electrically neutral.

(c) **Conductors** (which are not dielectrics) have free charge-carriers. Their behaviour was discussed on p. 354. They are *completely polarized* in that they make *zero* the field within themselves.

In fig. 47.4 we note three fields

- (1) E_0 the external (applied) field,
- (2) E' the internal field that results from the induced surface charges, and
- (3) $E_d = (E_0 - E')$, the resultant internal field.

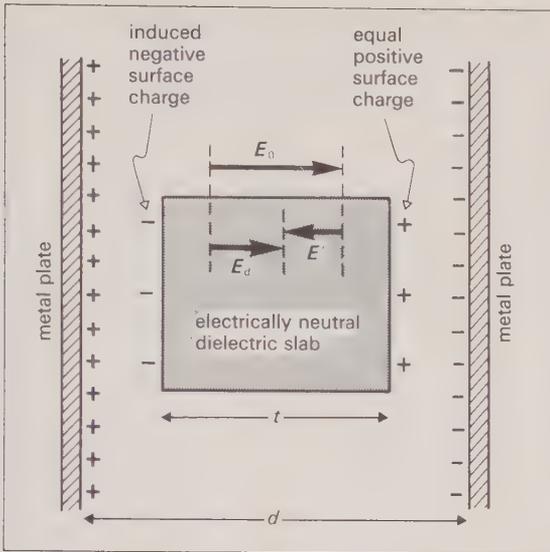


Fig. 47.4. The field inside a dielectric material.

It is E_d that is related to the p.d. V across the dielectric slab (not E_0 or E').

$$\epsilon_r = \frac{E_0}{E_d}$$

Capacitance of a partly filled capacitor

Refer again to fig. 47.4.

If no dielectric is present, the p.d. V_0 is given by

$$V_0 = E_0 d$$

When the dielectric slab is present, the p.d. V_d between the plates

$$\begin{aligned} V_d &= E_0(d - t) + E_d t \\ &= E_0(d - t) + \frac{E_0 t}{\epsilon_r} \\ &= E_0(d - t + t/\epsilon_r) \end{aligned}$$

Hence
$$\frac{C_d}{C_0} = \frac{V_0}{V_d} = \frac{d}{d - t + t/\epsilon_r}$$

from which ϵ_r can be found once C_0 , C_d and d and t are measured.

Electrostatic attraction

Fig. 47.5 shows how a charged body exerts an attractive force on a nearby neutral body which becomes polarized.

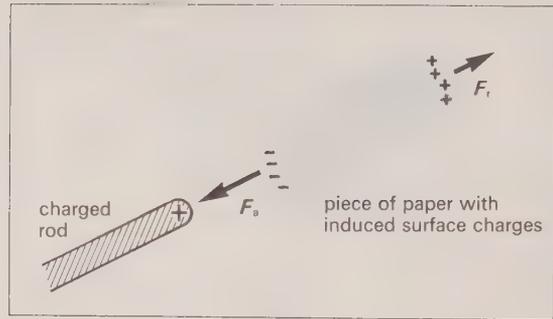


Fig. 47.5. Electrostatic attraction. $F_a > F_r$ since the field is non-uniform.

► 47.3 DIELECTRICS AND COULOMB'S LAW

For a vacuum we write

$$F = \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{Q_1 Q_2}{r^2}$$

When the charges are surrounded by dielectric material, they exert the same *direct* forces on each other as in a vacuum, but the *observed* force is smaller because the dielectric material also exerts a force on each charge, but in the opposite direction.

In a dielectric we write

$$\begin{aligned} F &= \frac{1}{\epsilon_r} \cdot \left(\frac{1}{4\pi\epsilon_0} \right) \cdot \frac{Q_1 Q_2}{r^2} \\ &= \left(\frac{1}{4\pi\epsilon} \right) \cdot \frac{Q_1 Q_2}{r^2} \end{aligned}$$

We modify the results on pp. 340 and 353.

At a point distant r from a point charge Q

$$E = \left(\frac{1}{4\pi\epsilon} \right) \cdot \frac{Q}{r^2} \quad \text{and} \quad V = \left(\frac{1}{4\pi\epsilon} \right) \cdot \frac{Q}{r}$$

These were derived from *Coulomb's Law*, and only apply when field lines are *radial*.

► 47.4 DIELECTRICS AND GAUSS'S LAW

(A) With no dielectric

In fig. 47.6 (overleaf) *Gauss's Law* can be written

$$\epsilon_0 E_0 A = Q$$

so

$$E_0 = \frac{Q}{\epsilon_0 A} \tag{1}$$

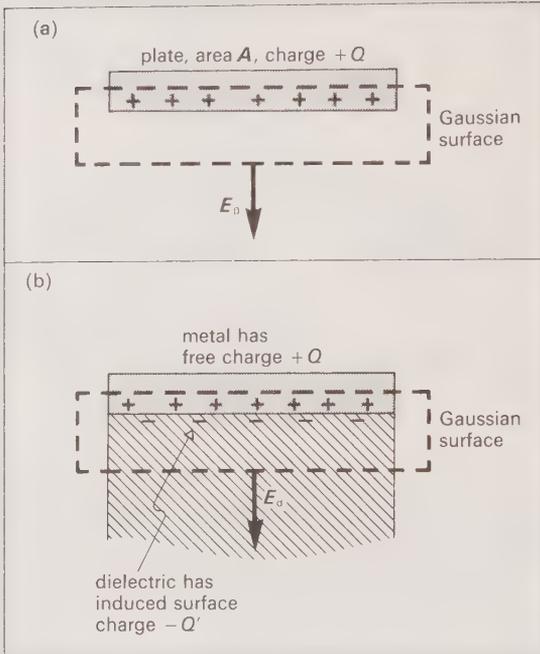


Fig. 47.6. Gaussian surfaces drawn (a) with air between the plates, and (b) with a dielectric material. (The top plate only of a parallel-plate capacitor is shown.)

(B) With dielectric

Gauss's Law can be written

$$\begin{aligned} \epsilon_0 E_d A &= Q - Q' \\ E_d &= \frac{1}{\epsilon_0 A} (Q - Q') \end{aligned} \quad (2)$$

But $E_d = \frac{E_0}{\epsilon_r}$ (p. 365)

so $E_d = \frac{Q}{\epsilon_r \epsilon_0 A}$, using (1)

Combining this with (2), we have

$$(Q - Q') = \frac{Q}{\epsilon_r} \quad (3)$$

This equation demonstrates that when

- (i) $Q' = 0, \epsilon_r = 1$: this corresponds to a vacuum,
- (ii) $Q' < Q, \epsilon_r > 1$: this corresponds to a dielectric material.

(Note that if we have a conducting slab, $Q' = Q$.)

Substitution for $(Q - Q')$ from (3) into (2) gives

$$\begin{aligned} \epsilon_0 E_d A &= \frac{Q}{\epsilon_r} \\ \epsilon_r \epsilon_0 E_d A &= Q \end{aligned}$$

We can now rewrite Gauss's Law as

$$\epsilon_r \epsilon_0 \psi_E = \Sigma \left(\begin{array}{l} \text{free charges inside} \\ \text{Gaussian surface} \end{array} \right)$$

When we insert ϵ_r on the left-hand side, we need only add up the *free or conduction charges* inside the *Gaussian surface*. We then omit the induced *surface or polarization charges*.

The law can be applied in this form to any situation met at this level.

► **47.5 ELECTRIC DISPLACEMENT D**

The **electric displacement D** at a point in an electric field is defined by

$$D = \epsilon_r \epsilon_0 E = \epsilon E$$

$$D \left[\frac{C}{m^2} \right] = \epsilon_r \epsilon_0 \left[\frac{C^2}{N m^2} \right] E \left[\frac{N}{C} \right]$$

D is a *vector* quantity which has the same unit, $C m^{-2}$, as surface density of charge, σ . It is a convenient concept for reformulating Gauss's Law:

$$\left(\begin{array}{l} \text{total electric displacement} \\ \text{flux emerging from a} \\ \text{closed surface} \end{array} \right) = \left(\begin{array}{l} \text{sum of free} \\ \text{charges enclosed} \\ \text{by the surface} \end{array} \right)$$

Displacement flux ψ_D is defined in an analogous way to *field flux* (p. 344), and has the unit of DA , i.e. $(C m^{-2})(m^2) = C$.

47.6 ENERGY CHANGES IN A CAPACITOR

When evaluating energy changes for a capacitor, use

$$W = \frac{1}{2} QV = \frac{1}{2} Q^2/C = \frac{1}{2} V^2 C$$

But if

(a) The *battery is connected*, then V is kept constant. A change in C results in a change in Q .

(b) The *capacitor is insulated*, then Q is kept constant. A change in C results in a change in V .

Examples

(a) If the plates of a parallel-plate capacitor are pulled apart, with the battery connected, then

$$V = \text{constant}$$

so $E = \frac{\sigma}{\epsilon_0}$ decreases,

so Q decreases

and we conclude W decreases. (Charge has been returned to the battery.)

(b) If the plates are pulled apart with the charged capacitor insulated, then

$$Q = \text{constant}$$

so $E = \frac{\sigma}{\epsilon_0}$ is constant

so $V = Ed$ increases

and we conclude W increases. An external agent does work against the attractive forces between the plates: the work done increases the electric potential energy stored in the field.

(c) Suppose a dielectric material is *slowly* inserted between the plates, as in fig. 47.7. The electric field causes work to be done *on* the dielectric slab, and so there is a decrease of electric potential energy. (Think of the gravitational analogy of a stone, held in someone's hand, being allowed to move downwards at constant speed.)

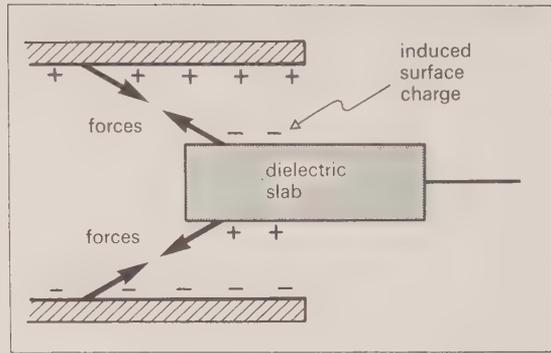


Fig. 47.7. A dielectric material attracted into a parallel-plate capacitor.

Results (a), (b) and (c) can alternatively be obtained by calculating the resulting change in C , and noting what stays constant.

SUMMARY OF ELECTROSTATICS

Defining equations		Laws	Important derivations
Q ϵ_0	$Q = \int I dt$ $F = \left(\frac{1}{4\pi\epsilon_0}\right) \cdot \frac{Q_1 Q_2}{r^2}$	<i>Coulomb</i> $F \propto \frac{Q_1 Q_2}{r^2}$	
E p	$E = F/Q_0$ $p = 2aQ$		$E = \left(\frac{1}{4\pi\epsilon_0}\right) \cdot \frac{Q}{r^2}$
ψ_E	$\psi_E = EA \cos \theta$	<i>Gauss</i> $\epsilon_0 \psi_E = \Sigma Q$	$E = \sigma/\epsilon_0$
V V_{ab}	$V = W/Q_0$ $V_{ab} = W_{ab}/Q_0$		$E = -\frac{dV}{dx}$ $E = \frac{V}{d}$ $V = \left(\frac{1}{4\pi\epsilon_0}\right) \cdot \frac{Q}{r}$ $F = \frac{\Delta A \sigma^2}{2\epsilon_0}$
C	$C = Q/V$		$C = \frac{\epsilon_0 A}{d}$ $C = C_1 + C_2$ parallel $\frac{1}{C} = \frac{1}{C_1} + \frac{1}{C_2}$ series $W = \frac{1}{2} Q_0 V_0 = \frac{1}{2} \frac{Q_0^2}{C}$ $= \frac{1}{2} V_0^2 C$
ϵ_r ϵ D	$\epsilon_r = C_d/C_0$ $\epsilon = \epsilon_r \epsilon_0$ $D = \epsilon_r \epsilon_0 E$	<i>Coulomb</i> $F = \left(\frac{1}{4\pi\epsilon}\right) \cdot \frac{Q_1 Q_2}{r^2}$ <i>Gauss</i> $\epsilon_r \epsilon_0 \psi_E = \Sigma (\text{free charges})$ <i>Gauss</i> $\left(\begin{array}{c} \text{displace-} \\ \text{ment} \\ \text{flux} \end{array}\right) = \left(\begin{array}{c} \text{sum of} \\ \text{free} \\ \text{charges} \end{array}\right)$	$C = \frac{\epsilon A}{d}$

X CURRENT ELECTRICITY

48	CURRENT, CONDUCTORS AND RESISTANCE	372
49	ENERGY TRANSFER IN THE CIRCUIT	385
50	E.M.F. AND CIRCUITS	388
51	PRINCIPLES OF ELECTRICAL MEASUREMENTS	394
52	THE MAGNETIC FIELD DEFINED FROM ITS EFFECT	400
53	THE MAGNETIC FIELD RELATED TO ITS CAUSE	410
54	ELECTRICAL MEASURING DEVICES	418
55	PRINCIPLES OF ELECTROMAGNETIC INDUCTION	424
56	APPLICATIONS OF ELECTROMAGNETIC INDUCTION	430
57	INDUCTANCE	436
58	ALTERNATING CURRENT	441
59	MAGNETIC PROPERTIES OF MATTER	459

48 CURRENT, CONDUCTORS AND RESISTANCE

48.1	THE NATURE OF CURRENT AND ELECTRICAL CONDUCTORS	372
48.2	THE CHARGE-CARRIERS IN DIFFERENT MATERIALS	373
48.3	DEFINITIONS	373
48.4	ELECTRICAL CHARACTERISTICS AND OHM'S LAW	375
▶ 48.5	BAND THEORY OF CONDUCTION THROUGH SOLIDS	376
48.6	THE MOVEMENT OF ELECTRONS THROUGH METALS	378

48.7	INTRINSIC SEMICONDUCTORS	379
48.8	EXTRINSIC (IMPURITY) SEMICONDUCTORS	380
48.9	VARIATION OF RESISTANCE WITH TEMPERATURE	382
48.10	COMPARISON OF METALS AND SEMICONDUCTORS	382
48.11	THE THERMISTOR	382
48.12	COMBINATIONS OF RESISTORS	383
48.13	THE POTENTIAL DIVIDER	384

48.1 THE NATURE OF CURRENT AND ELECTRICAL CONDUCTORS

The nature of current

An **electric current** consists of the movement of electric charge. A steady current exists when there is a systematic drift of charge-carriers. The *conventional* direction of an electric current is that in which there is a *net displacement of positive charge* over a given time interval (fig. 48.1).

Current is taken to be a *primary (fundamental) quantity* in electricity in the SI. The **ampere (A)**, the unit of current, is defined in terms of force and distance on p. 9.

Suppose a conductor carries a current I .

Then the **rate of flow of charge** Q past a given cross-section is defined by the equation

$$I = \frac{dQ}{dt}$$

$$I[\text{A}] = \frac{\Delta Q[\text{coulombs}]}{\Delta t[\text{s}]}$$

Notes

(a) The unit *ampere second (A s)* is called a **coulomb (C)**, and is the *unit for charge*.

(b) The charge Q that passes a given cross-section in a given time is found from

$$Q = \int I dt$$

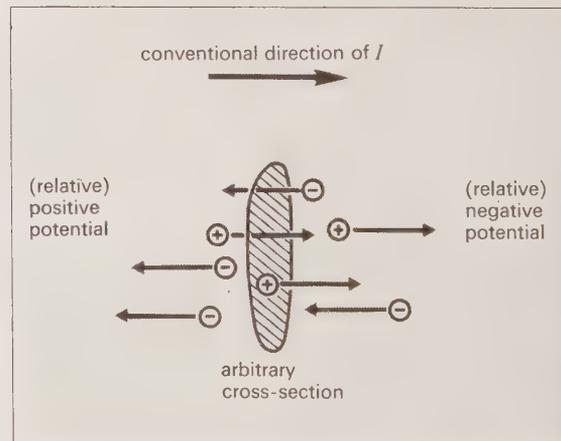


Fig. 48.1. Charge-carriers of both signs contribute to the net current I .

(c) When the current does not vary with time we can use

$$I = \frac{Q}{t} \quad Q = It$$

The nature of an electrical conductor

It was shown in 45.2 that $E = -dV/dx$. When a p.d. exists between two points an electric field is established. Any charged particles placed in that field will experience an electric force, and unless restraining forces are brought into play they will be accelerated.

A material will be a **conductor** if it allows the charged particles inside it to undergo significant *net translational motion* when a p.d. is applied. These particles are then called **mobile charge-carriers**. They may be negative or positive, and in some situations carriers of both signs will be present.

48.2 THE CHARGE-CARRIERS IN DIFFERENT MATERIALS

(a) **Insulators.** The atoms' outer orbital electrons are involved in forming the covalent or ionic bonds which hold the atoms together. They are therefore not mobile, and not available for conduction.

(b) **Metals.** Each atom in a metal lattice supplies one or more *free* electrons to form the bond which holds the positive ions together. The electrons are not attached to any particular parent ion, but are mobile within the boundary of the lattice, and therefore contribute to conduction. (A fuller account is given on p. 376 and p. 377.)

(c) **Semiconductors.** The process is similar to that in metals, but the number density of charge-carriers is much smaller. The carriers can be negative (electrons), positive (holes), or of both signs. (More details are to be found on p. 377 and p. 379.)

(d) **Electrolytes.** An electrolyte is a liquid which undergoes a chemical change when it conducts charge through itself. Solid NaCl consists of rigidly-held positive sodium ions and negative chloride ions (p. 131)—the ions are not free to wander. When NaCl is dissolved in water, water's high relative permittivity enables it to weaken the electrostatic attraction, and this causes the **cations** (positive) to become dissociated from the **anions** (negative). The ions originally present in the crystal have now become mobile.

The electric current in an electrolyte consists of a steady simultaneous drift, in opposite direc-

tions, of the mobile anions and cations. Chemical changes occur at the surfaces of the electrodes.

(e) **Gases.** (i) High-energy photons (such as X-rays) and high-energy particles can cause a neutral gas molecule to be split into its charged constituents. These ions can conduct a steady current so long as the cause of ionization persists. See, for example, the ionization chamber (p. 519).

(ii) If the ions produced in (i) are accelerated through a sufficiently large p.d. they may gain enough energy between collisions to cause further ionization. The process is called a **Townsend discharge**.

(iii) If the p.d. is increased still further the rate of production of ions can equal the rate at which the electrodes collect them. The luminous effect which results is called a **glow discharge**. The *gas discharge tube* was an important subject for research at the turn of the century.

(f) **Charged particles in a vacuum.** The movement of particles is not impeded in a vacuum. Charged particles subjected to an electric field in a vacuum undergo a continuing acceleration, and quickly achieve high speeds, as electrons do in the X-ray tube. (Contrast this with their staccato drift motion through a metal, p. 378.) The current involved is usually small.

48.3 DEFINITIONS

(Word endings: as a general rule a word ending
 ...*or* represents a *device*
 ...*ance* represents a *property of the device*
 ...*ivity* represents a *property of a material*.)

Refer to fig. 48.2 (overleaf).

(a) **Potential difference** V is defined by the equation

$$V_{ab} = \frac{W}{Q}$$

This is discussed on p. 350.

(b) The **resistance** R of the device is defined by the equation

$$R = \frac{V}{I}$$

$$R \left[\frac{V}{A} \right] = \frac{V[V]}{I[A]}$$

The unit volt ampere^{-1} is called an **ohm** (Ω). A conductor has a resistance of 1Ω if there is a p.d. of 1 V across its terminals when a current of 1 A flows through it.

$$[R] = [\text{ML}^2\text{T}^{-3}\text{I}^{-2}]$$

(c) The **conductance** G of the device is defined by the equation

$$G = \frac{I}{V} = \frac{1}{R}$$

$$G[\Omega^{-1}] = \frac{I[\text{A}]}{V[\text{V}]}$$

(The unit Ω^{-1} is called a **siemens** (S).)

Suppose the device shown is a *homogeneous* conductor, and that it has a cross-section of uniform area A (but any shape).

(d) The **resistivity** ρ of the material is defined by the equation

$$R = \frac{\rho l}{A}$$

$$R[\Omega] = \frac{\rho[\Omega \text{ m}]l[\text{m}]}{A[\text{m}^2]}$$

Some values (for room temperature) are given in the table below.

Material	Class	$\rho/\Omega \text{ m}$	$\sigma/\text{S m}^{-1}$
copper	good conductor	1.7×10^{-8}	5.8×10^7
manganin	conductor	4.4×10^{-7}	2.3×10^6
graphite	(poor) conductor	8.0×10^{-6}	1.2×10^5
germanium	semiconductor	0.5	2
silicon	semiconductor	2.5×10^3	4×10^4
quartz	insulator	5×10^{16}	2×10^{-17}

(e) The **conductivity** σ of the material is defined by the equation

$$\sigma = \frac{1}{\rho} = \frac{l}{RA}$$

σ is measured in $\Omega^{-1} \text{ m}^{-1} = \text{S m}^{-1}$, and has the dimensions $[\text{M}^{-1}\text{L}^{-3}\text{T}^3\text{I}^2]$.

Some values (for room temperature) are given in the table above.

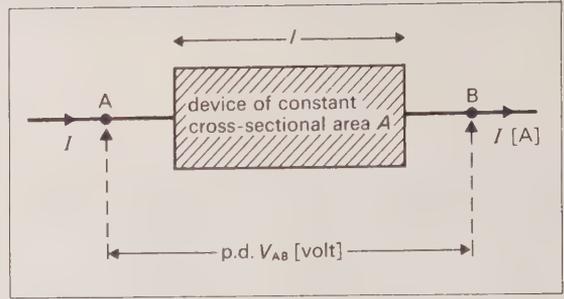


Fig. 48.2. To define V , R , G , ρ and σ .

Note that at room temperature σ and ρ vary from material to material by factors of up to 10^{24} . (See variations of thermal conductivity on p. 238.)

(f) The **electric current density** J at a cross-section of area A is defined by the equation

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

J is measured in A m^{-2} . The maximum safe value of J for copper is about 10^7 A m^{-2} . The concept of current density is used on p. 378 and p. 404.

(g) Suppose that a mobile charge-carrier acquires a mean drift velocity v_d when an electric field E is applied to a conducting material.

The **electric mobility** u of the charge-carriers is defined by the equation

$$u = \frac{v_d}{E}$$

$$u \left[\frac{\text{m}^2}{\text{V s}} \right] = \frac{v_d[\text{m/s}]}{E[\text{V/m}]}$$

The concept of mobility is used on p. 379.

In an electrolyte different types of ion will be present. Each may have a different mobility. This mobility will increase when the temperature rises, since the liquid viscosity is reduced.

48.4 ELECTRICAL CHARACTERISTICS AND OHM'S LAW

Suppose we measure simultaneously the p.d. V across a two-terminal electric device and the current I through it. The graph of V (x -axis, cause) against I (y -axis, effect) is called the **electrical characteristic**. Some examples are given in fig. 48.3.

Non-linear conductors and devices are those whose characteristics are not straight lines: their graphs may, in addition, be asymmetrical relative to the origin. Such devices include, apart from those shown, semiconducting materials (such as carbon, silicon and germanium), electrolytes (such as the liquids in chemical cells), thermionic diodes and metal rectifiers.

For some circuit elements, including the inductor and capacitor, the relation between the current and the applied p.d. is *time*-dependent.

Ohm's Law

Ohm discovered in 1826 that the characteristic shown in fig. 48.3(a) is common to a large class of conducting materials. His law was an experimental result, stated by:

Provided physical conditions (such as temperature) do not alter, then over a wide range of applied p.d.s

$$V \propto I$$

or $V/I = \text{constant}$

for many materials, particularly pure metals and alloys.

$R = \text{constant}$

Ohm's Law

(Note that $R = V/I$ defines R , and that it is *not* a statement of *Ohm's Law*.) The physical reason for a metal having this constant resistivity is discussed in 48.6.

Test of whether a substance obeys Ohm's Law

(A) Ammeter-voltmeter method

Connect the device under test into the circuit as in fig. 48.2. Use an ammeter to measure I , and a voltmeter to measure V . Plot the $V-I$ graph, and see whether it is a straight line.

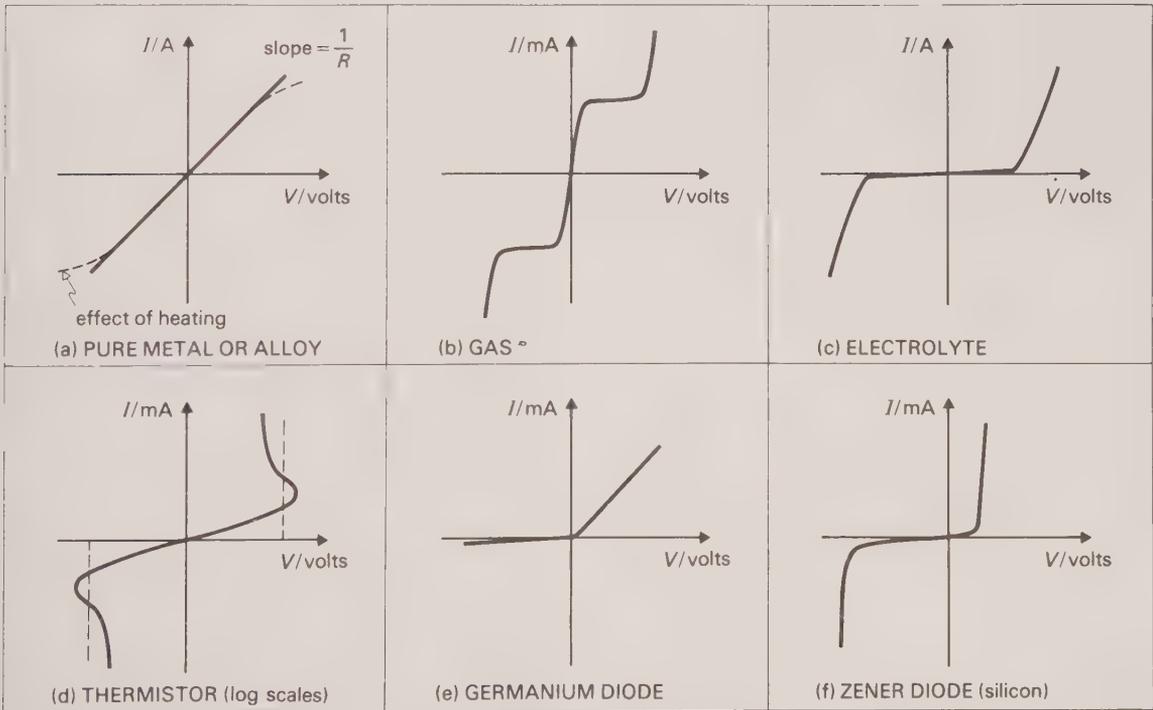


Fig. 48.3. Characteristics of conductors.

(It does not matter that the ammeter and voltmeter have been calibrated using the law, provided that *the material under test* does not form part of the instrument.)

One cannot conclude '*Ohm's Law* is not true', but only '*This material* does not obey the law'.

(B) Energy method

The rate P at which a resistor converts electrical energy to internal energy is given by

$$P = I^2R$$

and can be measured by a wattmeter (p. 423), or by calorimetry. If this is done for a range of currents, the graph of P against I^2 will be a straight line only if the device has a constant resistance.

(C) Wheatstone bridge method (p. 397)

► 48.5 BAND THEORY OF CONDUCTION THROUGH SOLIDS

The formation of energy bands

In *classical* theory the energy of an electron in a solid can have any value. In *quantum* theory certain discrete values only are allowed.

(a) *Isolated atoms*. The energies available to orbital electrons are confined to certain strict values. Some of those for the hydrogen atom are shown in fig. 48.4. (The theory behind these results is given in 63.11.)

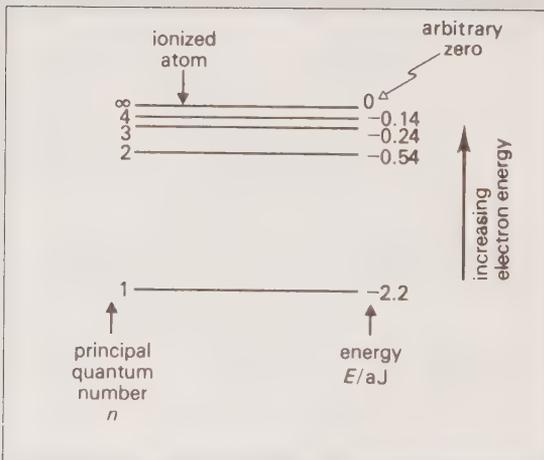


Fig. 48.4. The allowed energy levels of an isolated hydrogen atom.

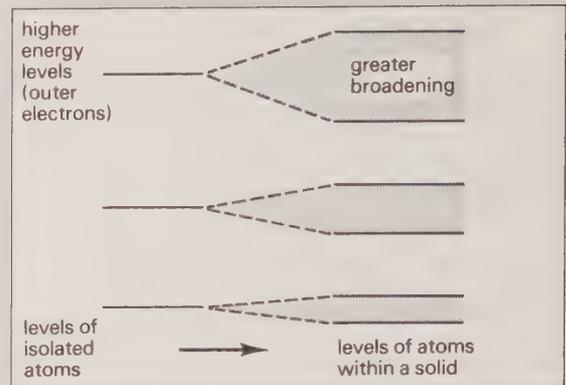


Fig. 48.5. Energy level broadening caused by the atoms of a solid affecting each other's discrete levels.

(b) *Interacting atoms*. When atoms are in a position to affect each other through their electric fields, a greater number of energy levels becomes available to the electrons. The individual lines of the isolated atomic levels are split or *broadened* into a very large number of levels which are very close together (fig. 48.5). Each group of energy levels is called a **band**.

Notes

(a) The outermost orbital electrons are most affected, and so their energy levels are broadened into the widest bands.

(b) Wider bands are produced for tighter packing, since the atomic interaction is more pronounced.

(c) It is possible for a higher energy band to be so broadened as to overlap a lower energy band.

How the bands can explain conduction

At absolute zero the electrons in a given crystal lattice occupy the lowest available energy states. (According to the *Pauli exclusion principle* no two electrons in a given atom can be in precisely the same state at the same time.)

An increase in temperature causes the electrons to move into higher *available* energy states. They move continually between different energy states as they exchange energy.

(a) *Insulators*. Refer to fig. 48.6. If the electrons occupy all the energy states available to them in a given band, they may be unable to change state. This would prevent them from acquiring the energy associated with the systematic drift velocity needed for the net transfer of charge. A band of this kind which is completely filled is a **valence band**.

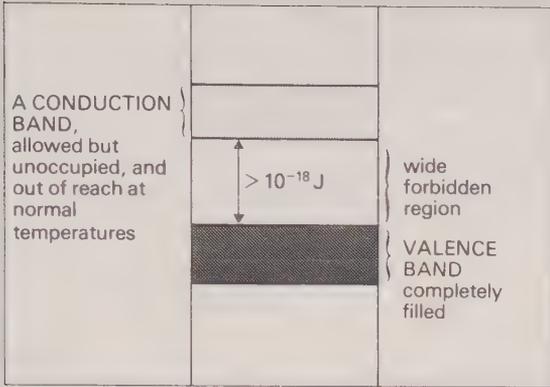


Fig. 48.6. Electron energy level band scheme for an insulator.

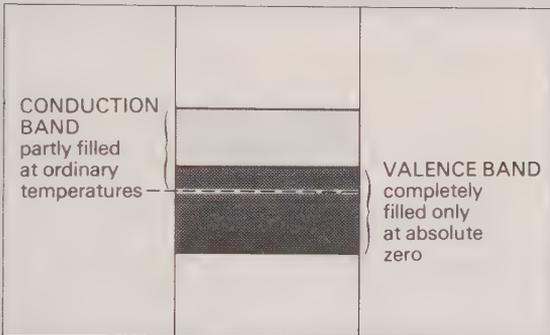


Fig. 48.7 Electron energy band scheme for a conductor in which the valence and conduction bands are overlapping.

(b) **Conductors.** An applied electric field will cause the systematic drift of electrons if those electrons in the highest occupied band do not fill it. Such a partly filled band is called a **conduction band** (fig. 48.7). The valence and conduction bands either have no gap between them, or they overlap.

(c) **Insulator breakdown.** It may be possible for electrons to be lifted across the gap from the valence band to the conduction band by the application of a sufficiently strong electric field, or by an increase in temperature. Those electrons then become mobile, since they can acquire drift k.e. by moving into different energy states in the conduction band.

(d) **Intrinsic semiconductors.** Refer to fig. 48.8. Suppose that a given substance has its highest occupied energy band completely filled at absolute

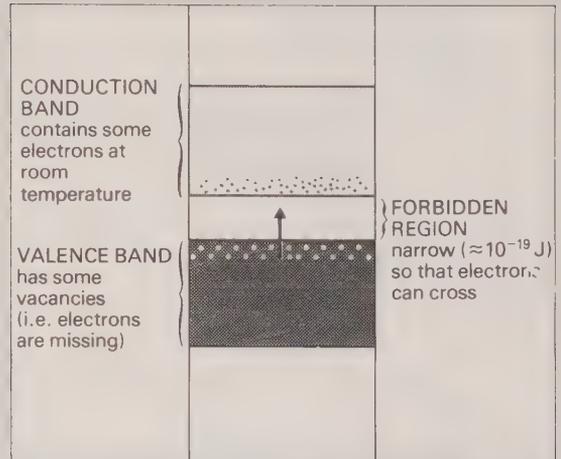


Fig. 48.8. Energy band scheme for an intrinsic semiconductor.

zero, and that the next allowed band does not overlap it. Such a substance would be an insulator.

At ordinary temperatures it is possible for the electrons to be given sufficient energy by thermal agitation to move continually from the valence band across the **forbidden zone** into the conduction band.

Notes

(i) As the temperature increases more electrons can cross the gap, and so the conductivity increases.

(ii) At ordinary temperatures the thermal excitation energy of electrons is about 10^{-20} J . The gap in diamond is close to $85 \times 10^{-20} \text{ J}$, making it an insulator, whereas that for germanium is only $11 \times 10^{-20} \text{ J}$, and so it is a semiconductor. If E_{ex} is the **molar excitation energy** involved in lifting electrons from the valence to the conduction band, the current flowing in the semiconductor is given by

$$I \propto e^{-E_{ex}/RT}$$

for a given applied p.d. The expression $e^{-E_{ex}/RT}$ is a measure of the proportion of free charge-carriers which have sufficient energy for conduction at the given applied p.d. The graph of $\ln I$ against $1/T$ would give a straight line with a negative gradient equal to $-E_{ex}/R$. A high value of E_{ex} would lead to a steep gradient and a small current at a given temperature.

(iii) The conductivity of semiconductors is influenced considerably by the presence of impurities, as discussed in 48.8. A **pure** semiconductor is called an **intrinsic** semiconductor. One **doped** with impurities is called an **extrinsic** semiconductor.

48.6 THE MOVEMENT OF ELECTRONS THROUGH METALS

The current in a metallic conductor

The charge carried by an electron is -1.6×10^{-19} C.

$$1 \text{ C} = 1 \text{ A s}$$

so if a current of 1 A flows for 1 s past a given cross-section in a metal, 6.25×10^{18} electrons pass in the opposite direction.

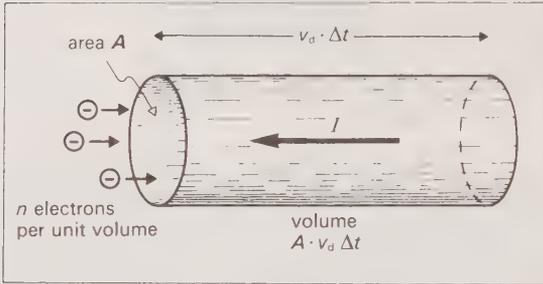


Fig. 48.9. Calculation of the current in a metallic conductor.

Suppose that the electrons in fig. 48.9 carry a charge of magnitude e , move with an average drift speed v_d and that their number density is n . After a time interval Δt all the electrons within the volume of length $v_d \Delta t$ will have passed through the right end face.

The charge ΔQ contained is given by

$$\Delta Q = n(Av_d \Delta t)e$$

so
$$\frac{\Delta Q}{\Delta t} = I = nAev_d$$

$$I = nAev_d$$

$$J = nev_d$$

Since $e = -1.6 \times 10^{-19}$ C, it follows that the direction of current density is opposite to that of the electron average drift velocity.

Compare

(a) the speed of an electrical signal in a conductor ($\approx 3 \times 10^8 \text{ m s}^{-1}$),

(b) the speeds of the electrons in their random motion ($\approx 10^6 \text{ m s}^{-1}$), and

(c) the average drift speed v_d of the electrons ($\approx 10^{-4} \text{ m s}^{-1}$ in a typical conductor carrying a current of 5 A through a cross-section of 1 mm^2).

► The mechanism of current flow

The mobile or conduction electrons in a metal are accelerated by electric forces until they collide inelastically with lattice ions: the motion is repeated very rapidly at short time intervals, as suggested in fig. 48.10(a), being superimposed on the random thermal motion of the electrons.

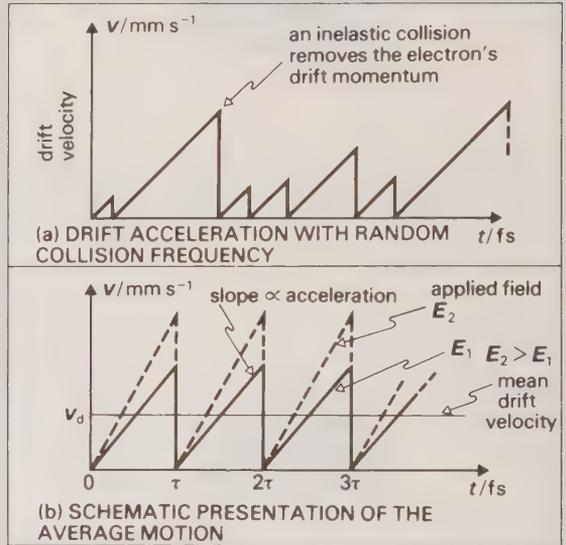


Fig. 48.10. The mechanism of electron movement through a metallic conductor carrying a current.

It can be calculated from the relationships that follow that electrons have a surprisingly large mean free path of about 100 ionic diameters.

The macroscopic effect is a steady current flow, and the gain of internal energy by the lattice causes an increase of temperature. Microscopically electric field energy is converted initially to the mechanical k.e. of the drifting electrons, and then to the k.e. and p.e. of the vibrating lattice ions. The electrons are in thermal equilibrium with these ions.

In a steady field E the electrons have a constant acceleration superimposed on their random motion. Using $a = F/m = Ee/m_e$, we can calculate their average drift velocity v_d from

$$v_d = \frac{1}{2}(\text{maximum drift velocity})$$

$$= \frac{1}{2}at = \frac{1}{2}\left(\frac{Ee}{m_e}\right)\tau$$

where τ is the characteristic time (about 10^{-14} s) for which an electron gains speed between collisions.

But $v_d = J/ne$ (boxed above)

Equating $\frac{J}{ne} = \frac{E\tau}{2m_e}$

or $J = \frac{nEe^2\tau}{2m_e}$ (1)

Now by definition, electrical conductivity

$$\sigma = \frac{l}{RA} = \frac{II}{VA} \quad (\text{def. of } R)$$

$$= \frac{J}{E} \quad (\text{def. of } J)$$

$$J = \sigma E$$

Combining this result with equation (1) above gives

$$\sigma = \frac{ne^2\tau}{2m_e}$$

► The physical basis of Ohm's Law

The last equation indicates that if the values of e and m_e are fixed, the electrical conductivity of a material, and hence the resistance of a given sample, will be independent of current so long as

(a) the number density n of charge-carriers does not depend on I , and

(b) τ is independent of current. Now the frequency of electron collision with the lattice is not changed by the current flow. This is because typically $v_d \approx 10^{-4} \text{ m s}^{-1}$, whereas the random speed $\approx 10^6 \text{ m s}^{-1}$.

These two conditions are obeyed by metals and alloys unless the temperature changes, or unless the material undergoes a change of stress.

Electron mobility

Since $\sigma = J/E$, and $J = nev_d$,

$$\sigma = nev_d/E = neu_e$$

where u_e is the mobility of the electrons.

Thus one aspect of Ohm's law may be expressed in microscopic terms by the statement

At a fixed temperature the mobility of electrons in pure metals and some alloys is independent of current.

In addition, of course, Ohm's law requires n to be constant.

48.7 INTRINSIC SEMICONDUCTORS

The *band theory* of intrinsic (pure) semiconductors is given on p. 377. According to this theory the net transfer of charge necessary for an electric current is caused by

(a) the movement of *electrons*, which in effect have been shaken loose by their parent atoms by thermal agitation, together with

(b) the oppositely directed movement of a position from which an electron is missing. A position such as this is called a **positive hole**: it is a vacancy for an electron, and it is located in the place of a positive ion in the crystal lattice.

The electrons and positive holes in an intrinsic semiconductor *both* contribute to the conductivity, and are called **minority charge-carriers**. (Only about 1 atom in 10^{10} contributes charge-carriers in this way.)

Positive holes

While the holes indicate *locations* in the crystal lattice, *they behave like positive particles*. Under the influence of an electric field they drift through the lattice in the same manner as electrons. The mechanism is the movement of a bound electron (in the valence band) from an atom into a neighbouring positive hole. Holes have a smaller electric mobility than conduction electrons (see table), so the latter contribute more to the conductivity.

Substance	Electron mobility $u_e/\frac{\text{m}^2}{\text{Vs}}$	Hole mobility $u_h/\frac{\text{m}^2}{\text{Vs}}$	At 300 K $\frac{n_e + n_h}{\text{m}^3}$
copper	0.0028	—	1×10^{29}
germanium	0.39	0.19	5×10^{19}
silicon	0.14	0.048	3×10^{16}

Variations of conductivity

The energy necessary to create an electron–hole pair can be provided

- by thermal agitation,
- by photons of electromagnetic radiation, or
- during avalanche breakdown (p. 478).

At equilibrium there is a balance between the regeneration and recombination of electrons and holes: this means that at a particular temperature the product of the concentrations of electrons and holes is a constant. Their sum is a minimum when the two concentrations are equal.

Their product increases rapidly with rise in temperature for both silicon and germanium.

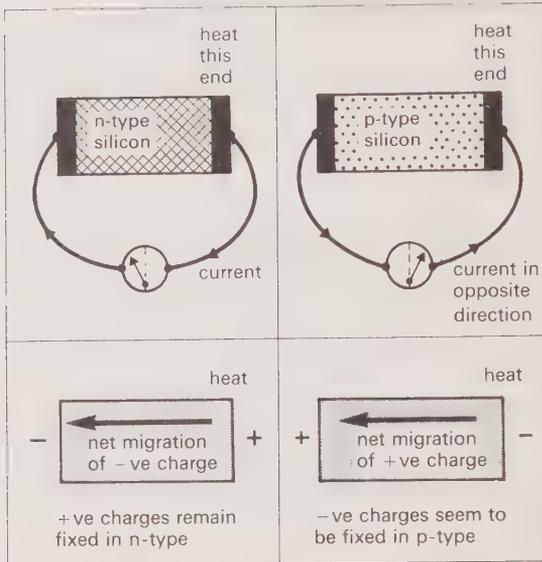


Fig. 48.11. Demonstration of existence of charge-carriers of opposite signs.

48.8 EXTRINSIC (IMPURITY) SEMICONDUCTORS

Seebeck effect

Refer to fig. 48.11. The conclusion to be drawn is that the current-flow mechanism appears to be different in the two different types of silicon. This is confirmed by the *Hall* effect (p. 403).

Their formation

Semiconductors such as silicon can be made extremely pure, but their conductivity is considerably modified by the presence of impurities. (An impurity level of 1 in 10^7 may change the conductivity by a factor of 10^3 .) These impurities can be introduced deliberately: the process is called **doping**.

(a) *The structure of pure silicon.* Refer to fig. 48.12. The silicon atom can be considered to consist of a positive core (the nucleus, and those orbital electrons in closed shells) surrounded by four valence electrons. Silicon is *tetravalent*, and its crystal structure is tetrahedral, like that of diamond.

Any one positive core shares its four valence electrons, one with each of the four near-neighbouring cores. No electrons are mobile at low temperatures since all the valence electrons are involved in forming these **covalent bonds**.

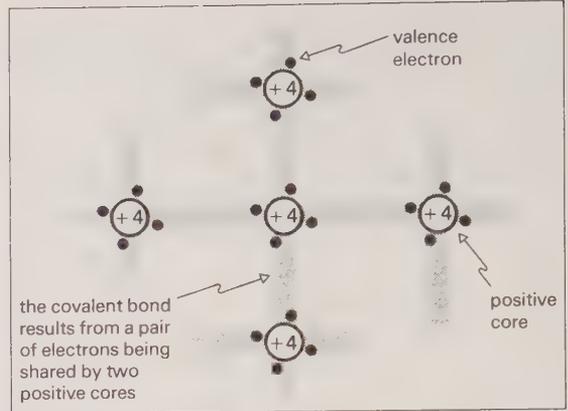


Fig. 48.12. Two-dimensional schematic representation of the three-dimensional arrangement of ions in a silicon crystal lattice.

(b) *The effect of donor impurities.* Refer to fig. 48.13. Pentavalent atoms, such as those of antimony, arsenic and phosphorus, have *five* valence electrons, of which only four are needed to complete the covalent bonding. The fifth electron is bound very weakly to the parent atom, and is easily liberated to become mobile. This liberation requires less energy than would be needed by an electron to escape from the Si-Si bond.

Notes

(i) The conducting material is called **n-type**, since the main contribution to its conductivity comes from the mobile *electrons*. This is revealed experimentally by the *Hall* effect (p. 403).

(ii) The pentavalent material is called a **donor** impurity, since it *gives* electrons to the lattice.

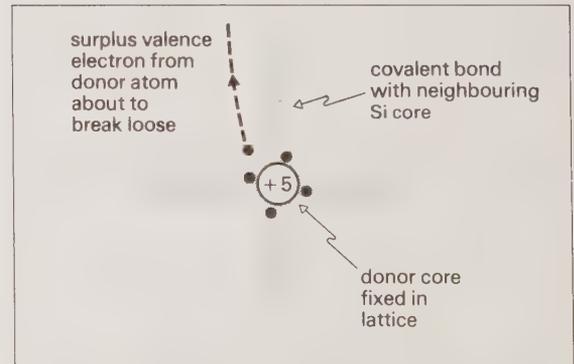


Fig. 48.13. The formation of n-type silicon.

(iii) The electrons are referred to as **majority carriers**. There will always be some hole conduction, and in an n-type material the holes are called **minority carriers**.

(iv) The number density of electrons considerably exceeds that of positive holes.

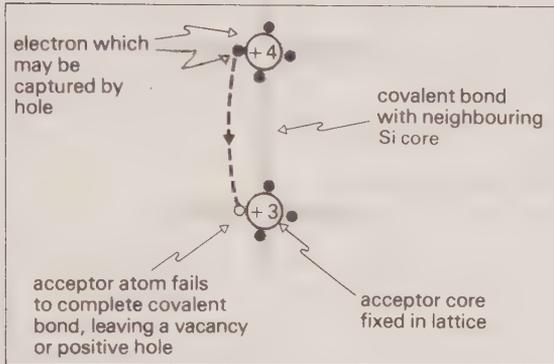


Fig. 48.14. The formation of p-type silicon.

(c) **The effect of acceptor impurities.** Refer to fig. 48.14. Trivalent atoms, such as those of aluminium, boron or indium, have *three* valence electrons, whereas four are needed to complete the covalent bonding. Consequently a positive hole or *vacancy* appears in the lattice, and less energy is needed for this hole to capture an electron from an adjacent silicon atom than is required for an electron to break out of the Si-Si bond. Such a capture causes the net translation of a location at which there is a net positive charge.

Notes

(i) The conducting material is called **p-type**, since the main contribution to its conductivity comes from the mobile *positive holes*. This is revealed experimentally by the *Hall effect* (p. 403).

(ii) The trivalent material is called an **acceptor impurity**, since it *accepts* electrons from the lattice atoms to fill the vacancies that it creates.

(iii) The holes are referred to here as **majority carriers**. There will always be some electron conduction, and in a p-type material the electrons are called **minority carriers**.

(iv) The number density of holes considerably exceeds that of electrons.

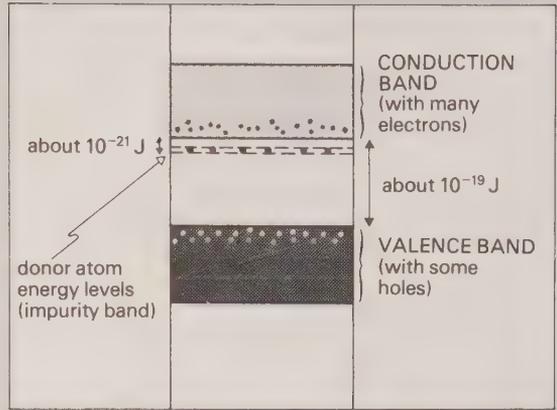


Fig. 48.15. Energy band scheme for an n-type impurity semiconductor. (Not to any scale.)

► **The band theory of extrinsic semiconductors**

(a) **n-type.** Refer to fig. 48.15. The donor atoms contribute electrons whose available closely-spaced energy levels lie just below the conduction band. These *electrons* are moved easily by thermal excitation to available states in the conduction band, where they become *mobile*.

(b) **p-type.** Refer to fig. 48.16. The acceptor atoms contribute electrons whose available closely-spaced energy levels lie just above the valence band. Electrons can move easily into these levels, and so create *mobile positive holes* in the valence band.

As the acceptor energy levels are usually full, electrons in these levels cannot normally take part in conduction.

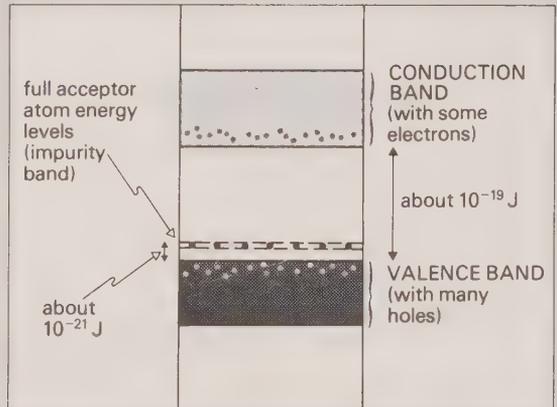


Fig. 48.16. Energy band scheme for a p-type impurity semiconductor. (Not to any scale.)

48.9 VARIATION OF RESISTANCE WITH TEMPERATURE

It was shown in 48.6 that a material's conductivity is given by

$$\sigma = ne^2\tau/2m_e$$

where τ is the characteristic time for which an electron gains speed between collisions, and n is the number density of charge-carriers.

(A) Metallic conductors

(a) n is practically independent of temperature for most metals.

(b) An increase of temperature increases the amplitude of vibration of lattice ions, which reduces τ .

Thus the conductivity falls, and the resistivity rises with increase of temperature, as shown in fig. 48.17.

We define the **temperature coefficient of resistivity** α by the equation

$$\alpha = \frac{1}{\rho} \frac{d\rho}{dt}$$

Near the ice-point the curve is described approximately by

$$\rho = \rho_0[1 + \alpha\theta]$$

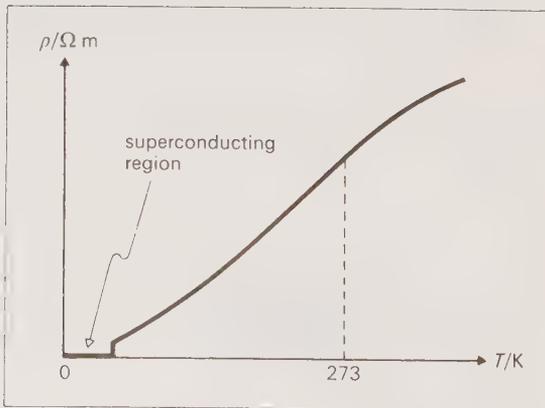


Fig. 48.17. Variation with temperature T of metal resistivity ρ . (The zero resistivity is not shown by all metals.)

where ρ_0 is the resistivity at the ice-point, ρ the resistivity at temperature T , and $\theta = T - 273$ K. For pure metallic elements α is positive, and has a value close to $4 \times 10^{-3} \text{ K}^{-1}$. The variation of resistivity with temperature is utilized in the resistance thermometer.

As the temperature approaches absolute zero, some substances (pure metals, alloys and some compounds) become **superconductors**. Their crystal lattices are unable to take energy from the drifting electrons, and so $\rho = 0$. This effect can be explained only by the quantum theory.

Some alloys, such as constantan, have an extremely small value of α : this makes them useful for constructing standard resistors.

(B) Semiconductors

An increase in temperature

- (i) greatly increases n (p. 377), but
- (ii) as in metallic conductors, reduces τ .

For many semiconductors effect (i) is far more important than (ii) in certain temperature ranges.

Typically $\alpha \approx -6 \times 10^{-2} \text{ K}^{-1}$, being negative because the resistivity is reduced as the temperature increases.

Semiconductor-metal junctions

(a) Where a contact occurs the current is maintained by the recombination of electrons and holes, or by their regeneration, at the semiconductor surface.

(b) The condition of the interface at the contact will determine whether the resistance is to be ohmic or to have the ability to rectify.

48.10 COMPARISON OF METALS AND SEMICONDUCTORS

Refer to the table opposite.

48.11 THE THERMISTOR

This is a non-linear device made of semiconducting material (fig. 48.18, p. 384). The name is a contraction of *thermally sensitive resistor*. A typical characteristic is given in fig. 48.3: its shape can be explained by the large, negative and variable values taken by α at different temperatures. A given value of V might correspond to one or more possible values for I .

Comparison of metals and semiconductors

	Metal	Semiconductor
charge-carriers	free electrons	can be mainly electrons, or positive holes, or both equally
number density at 300 K n/m^{-3}	1.1×10^{29} (Cu)	3×10^{16} (Si)
resistivity at 300 K $\rho/\Omega \text{ m}$	10^{-5} to 10^{-8} 1.7×10^{-8} (Cu)	variable 2.5×10^3 (Si)
Hall effect	usually negative (free electrons), but can be positive (e.g. Zn)	positive (p-type) or negative (n-type), but see below*
typical Hall coefficient $R_H/m^3 \text{ C}^{-1}$	-1.1×10^{-10} (Au)	depends upon doping, but for p-type Si we might have +10
typical temperature coefficient of resistance α/K^{-1}	$+4 \times 10^{-3}$	about -60×10^{-3}
effect of impurities	relatively little	impurity concentration of 1 in 10^7 can alter ρ by a factor of about 10^3 at 300 K
effect of light	very little, unless photon energies can produce an external photoelectric emission	incident photons readily liberate free charge-carriers internally

* For Ge and Si the fact that $u_e > u_h$ makes the Hall coefficient negative even for intrinsic materials.

Applications

- (a) Resistance thermometer
- (i) Small size, sensitive, and with rapid response.
- (ii) Used in the bolometer (p. 244) for the measurement of energy flow in a beam of electromagnetic waves.
- (iii) Useful range is from about 200 K to 600 K.
- (iv) It is unstable, and needs periodic recalibration.
- (b) It can be used to operate relays (in e.g. a fire alarm), since R falls markedly as T increases.
- (c) It is used to limit surge currents through equipment when the current is switched on or off (fig. 48.18(c)). For certain applications (e.g. to limit the initial current through a cold filament) it is connected in series.

48.12 COMBINATIONS OF RESISTORS

See fig. 48.19, overleaf.

(A) In series

Each resistor carries the same current I . But

$$V_{AC} = V_{AB} + V_{BC}$$

(since the electric field is conservative, and energy is conserved), so

$$IR = IR_1 + IR_2$$

$$\therefore \boxed{R = R_1 + R_2} \quad \text{Series}$$

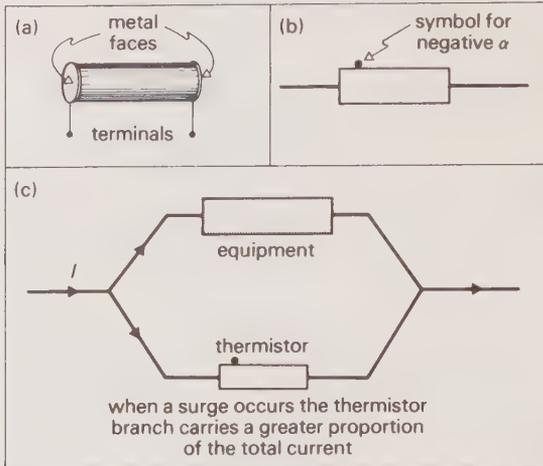


Fig. 48.18. The thermistor: (a) a typical disc-type design, (b) the circuit symbol and (c) its use as a surge suppressor.

(B) In parallel

There exists the same p.d. V across each resistor.
But

$$I = I_1 + I_2$$

(law of conservation of charge), so

$$\frac{V}{R} = \frac{V}{R_1} + \frac{V}{R_2}$$

$$\therefore \frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} \quad \text{Parallel}$$

The expressions can be extended to apply to any number of resistors.

48.13 THE POTENTIAL DIVIDER

A variable resistor is frequently used to control the p.d. across a device, as shown in fig. 48.20.

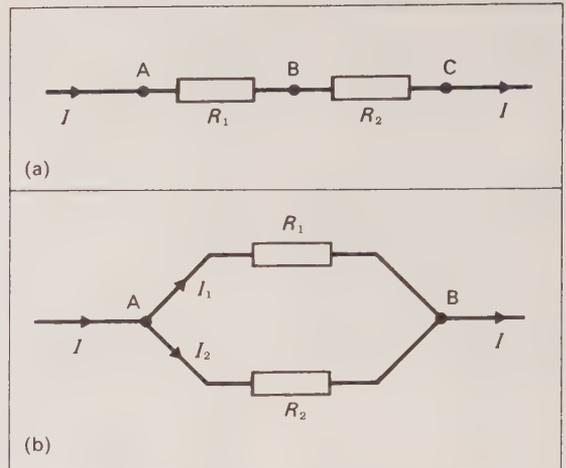


Fig. 48.19. Calculation of equivalent resistance: (a) in series, (b) in parallel.

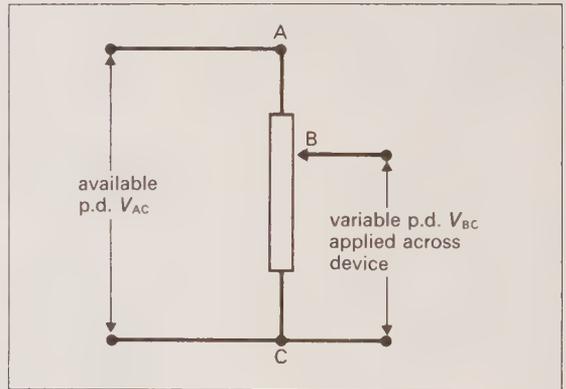


Fig. 48.20. A potential divider.

Since
$$\frac{V_{BC}}{V_{AC}} = \frac{R_{BC}}{R_{AC}}$$

the value of V_{BC} can be varied by moving the sliding contact B on the resistor AC.

(Note that this equation applies only if the device concerned does not draw any current.)

49 ENERGY TRANSFER IN THE CIRCUIT

49.1 ENERGY CONVERSION IN A LOAD	385
49.2 JOULE HEATING	385
49.3 TRANSMISSION OF ELECTRICAL POWER	386

▶ 49.4 THE MAXIMUM POWER THEOREM	386
▶ 49.5 HOW THE CURRENT DIVIDES AT A JUNCTION	386

49.1 ENERGY CONVERSION IN A LOAD

In fig. 49.1 the *source* of electrical energy could be a battery. The *load* might be a resistor, a motor, an accumulator or a combination of devices.

By definition, whatever the nature of the load, a charge Q which passes through the load loses energy W given by

$$V_{AB} = \frac{W}{Q}$$

$$W = VQ$$

The rate at which energy is transferred from the source to the load is given by power

$$P = \frac{dW}{dt} \quad (\text{p. 47})$$

$$= V \frac{dQ}{dt}$$

$P = VI$

$$P \left[\frac{\text{J}}{\text{s}} \right] = V \left[\frac{\text{J}}{\text{C}} \right] I \left[\frac{\text{C}}{\text{s}} \right]$$

$$P[\text{W}] = V[\text{V}]I[\text{A}]$$

The equation $P = VI$ enables us to calculate the rate at which the load converts electrical energy to other forms, e.g.

- (a) if a resistor, to internal energy (below),
- (b) if a motor, to mechanical work done by the motor,
- (c) if an accumulator, to stored chemical energy.

$P = VI$ applies to any circuit element.

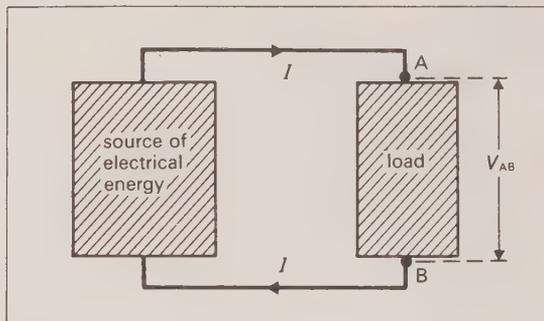


Fig. 49.1. Transfer of energy.

49.2 JOULE HEATING

For a resistor carrying a steady current I , we have

$$\frac{W}{t} = VI \quad (\text{from the def. of } V)$$

and $R = \frac{V}{I}$ (the def. of R)

Combining these expressions, the electrical energy W converted to internal energy of the lattice in time t is given by

$$W = \frac{V^2}{R} t = I^2 R t$$

Joule's Laws

Joule's Laws represent energy conservation when a resistor converts electrical energy into internal energy. The mechanism is given on p. 378. The laws apply *only* to a resistor.

Notes

(a) This energy conversion is *irreversible* (p. 204).

(b) The Department of Trade and Industry measures electrical energy in kilowatt hours (kWh), where 1 kWh is the electrical energy converted to other forms when a power of 1 kW is used for 1 hour.

$$\begin{aligned} \text{Thus } 1 \text{ kWh} &= (1000 \text{ W}) \times (3600 \text{ s}) \\ &= 3.6 \times 10^6 \text{ W s} \\ &= 3.6 \times 10^6 \text{ J} \end{aligned}$$

The kilowatt hour is no longer recognized by physicists as a *unit*.

49.3 TRANSMISSION OF ELECTRICAL POWER

In fig. 49.2

$$\begin{aligned} \text{power generated} &= IV_g \\ \text{power dissipated in cables} &= 2I^2r \\ \text{power available at load} &= IV_g - 2I^2r \end{aligned}$$

The power loss is reduced by making V_g large so that I , and $2I^2r$, are small. An economic balance must be struck with insulation requirements.

Potential differences are stepped up and down using transformers, which function *only* on alternating current (p. 431). (Be sure not to confuse V_g in problems of this type with the potential drop $2Ir$ along the cables.)

► 49.4 THE MAXIMUM POWER THEOREM

We want to calculate the value of R (fig. 49.3) for which the battery supplies maximum power to the external circuit.

External power supplied

$$\begin{aligned} P &= I^2R \\ &= \frac{\mathcal{E}^2R}{(R+r)^2} \quad \text{since } \mathcal{E} = I(R+r) \\ \frac{dP}{dR} &= \mathcal{E}^2 \left[\frac{(R+r) - 2R}{(R+r)^3} \right] \end{aligned}$$

For a stationary value

$$\frac{dP}{dR} = 0$$

so $R + r = 2R$

Either (a) $R = 0$ or $\infty \Omega$, in which case the power supply is zero (minimum), or

(b) $R = r$, which corresponds to maximum rate of energy supply.

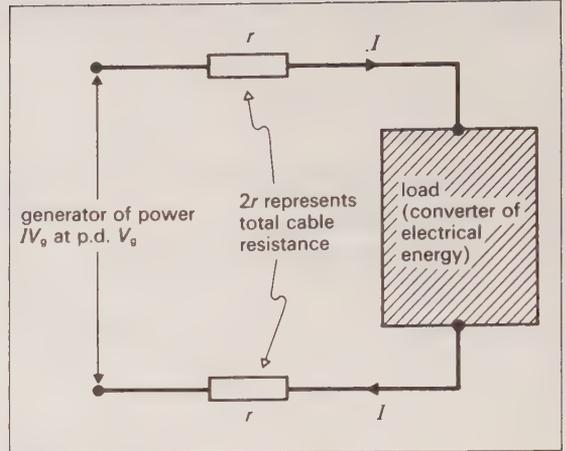


Fig. 49.2. Calculation of power losses.

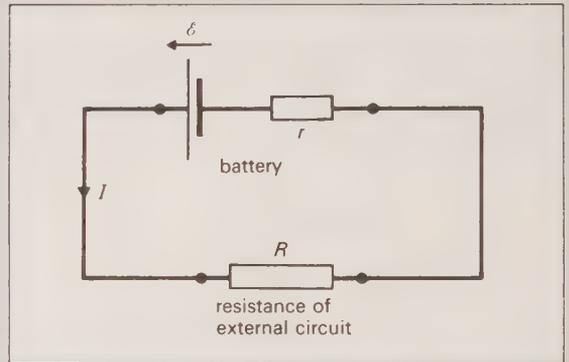


Fig. 49.3. Circuit for maximum power theorem.

Solution (b) shows that any generator, such as a transistor amplifier, delivers maximum *power* to an external load, such as a loudspeaker, when the output impedance of the generator equals the input impedance of the load. Direct connection of a speaker ($Z \sim 10 \Omega$) to an amplifier ($Z \sim 10^3 \Omega$) would give poor results and so a linking device such as a *matching* transformer is required.

Note that when $R = r$ half the electrical energy available is converted to internal energy in the source. This does *not* represent maximum *efficiency*.

► 49.5 HOW THE CURRENT DIVIDES AT A JUNCTION

In fig. 48.19(b) we want to find how I_1 and I_2 are related if least power is to be dissipated as internal energy in the resistors.

The power dissipated

$$\begin{aligned} P &= I_1^2 R_1 + I_2^2 R_2 \\ &= I_1^2 R_1 + (I - I_1)^2 R_2 \\ \frac{dP}{dI_1} &= 2I_1 R_1 + 2(I - I_1)(-1)R_2 \end{aligned}$$

For a stationary value

$$\frac{dP}{dI_1} = 0$$

$$\begin{aligned} \therefore I_1 R_1 &= (I - I_1) R_2 \\ &= I_2 R_2 \end{aligned}$$

The current splits so that

- (a) $I \propto 1/R$, and
- (b) $IR = \text{constant}$. We have the same p.d. across each resistor. This result was assumed on p. 384.

(We have assumed R_1 and R_2 to be independent of current. The p.d. across each resistor remains equal even if this is not true, but the power dissipated is then not a minimum.)

50 E.M.F. AND CIRCUITS

50.1 ELECTROMOTIVE FORCE	388
50.2 THE CIRCUIT EQUATION	389
50.3 KIRCHHOFF'S RULES	389

50.4 THERMOELECTRICITY	391
▶ 50.5 THE RC CIRCUIT	392

50.1 ELECTROMOTIVE FORCE (E.M.F.)

A seat of e.m.f. is a device which can supply energy to an electric current. It does this by producing and maintaining a potential difference (and of course an electric field) between two points to which it is attached.

For example a chemical cell causes forces to be exerted on positive charge-carriers which take them from a point of low potential (negative terminal) to a point of higher potential (positive terminal), and vice versa for negative charges. It exerts a non-conservative force on the charge-carriers to drive them against the electrostatic force in the seat of e.m.f., and in so doing achieves a reversible transformation (p. 204) between non-electrical and electrical energy.

Notes

(a) The e.m.f. exists whether the current flows or not (see the theory of the transformer with the secondary on open circuit (p. 431)).

(b) An e.m.f. can exist in a vacuum—a conductor does not have to be present. (See the discussion of induced e.m.f. on p. 426.)

Examples. Each of these examples has its counterpart (indicated in brackets) in which electrical energy is converted to non-electrical energy:

(a) the electrochemical cell or battery (the voltmeter),

(b) the thermocouple, using the *Seebeck effect*, p. 391 (the *Peltier effect* at a metal junction),

(c) the photoelectric or photovoltaic cell (the fluorescence of a phosphor),

(d) the electric generator or dynamo (the electric motor).

Joule heating is a non-reversible effect, so we do not call the p.d. across a resistor an e.m.f.

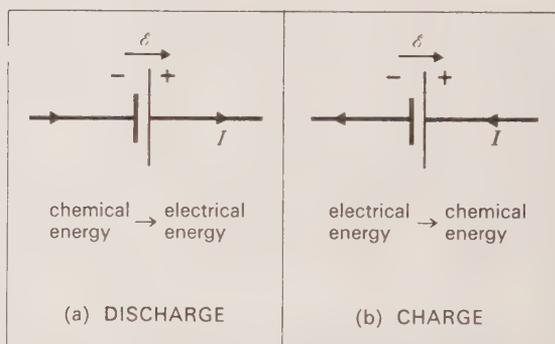


Fig. 50.1. Energy conversions within a chemical cell.

Definition of e.m.f.

Suppose that non-electrical energy W is converted to electrical energy when a charge Q is taken through the source (seat of e.m.f.).

Then we define the e.m.f. \mathcal{E} of the source by the equation

$$\mathcal{E} = \frac{W}{Q}$$

$$\mathcal{E} [\text{volts}] = \frac{W[\text{J}]}{Q[\text{C}]}$$

Notes

(a) The unit of e.m.f. is the same as that of p.d.—the *volt*.

(b) \mathcal{E} equals the p.d. between the terminals of the source when no current flows through the source—when it is on open circuit.

(c) If a charge Q passes through such a seat of e.m.f. in the direction shown in fig. 50.1(a), the energy W given to it is $\mathcal{E}Q$.

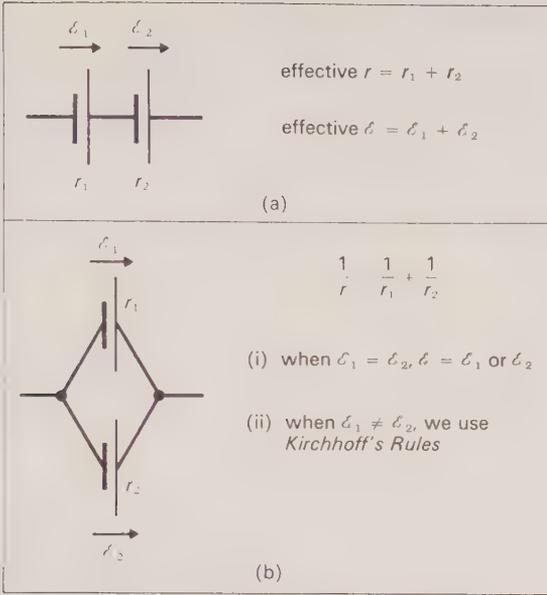


Fig. 50.2. Cells connected (a) in series, and (b) in parallel.

Combination of cells

Fig. 50.2 shows the effective values of e.m.f. and resistance when several cells are used together.

50.2 THE CIRCUIT EQUATION

Rate of energy conversion in a cell

From the previous section

$$W = \mathcal{E}Q$$

so differentiating

$$\frac{dW}{dt} = \mathcal{E} \frac{dQ}{dt}$$

$$P = \mathcal{E}I$$

The power of a cell is the rate at which it converts chemical energy into electrical energy, and is given by

$$P = \mathcal{E}I$$

Application to the circuit (fig. 50.3)

Rate of production of electrical energy by the cell

$$= \mathcal{E}I$$

Rate of dissipation of this energy (as Joule heating)

(a) in the external resistance R , $= I^2R$,

(b) in the resistance r of the cell (its internal resistance), $= I^2r$.

By the law of the conservation of energy for the circuit, these are equal:

$$\mathcal{E}I = I^2R + I^2r$$

$$\mathcal{E} = I(R + r)$$

The Circuit Equation

The p.d. across the cell terminals

$$= V_{AB} = IR \quad (\text{def.})$$

This equals the p.d. across the cell

$$= (\mathcal{E} - Ir)$$

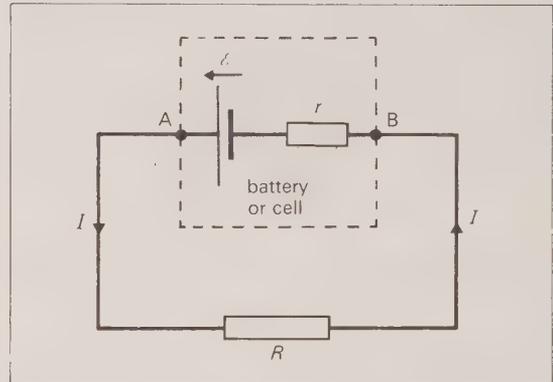


Fig. 50.3. The simple circuit.

The terminal p.d. of the cell is less than the e.m.f. \mathcal{E} by Ir (the lost volts), because the internal resistance of the cell dissipates electrical energy.

As mentioned in the previous section,

$$V_{AB} = \mathcal{E} \quad \text{when} \quad I = 0$$

The current can be calculated from

$$I = \frac{\mathcal{E}}{R + r} = \frac{V_{AB}}{R} = \frac{\mathcal{E} - V_{AB}}{r}$$

Fig. 50.4 (overleaf) shows how the potential varies from point to point in the circuit.

50.3 KIRCHHOFF'S RULES

What the rules do

They (a) summarize the conditions of current and p.d. in circuits with steady currents, and

(b) state the conservation laws of

(i) electric charge, and

(ii) energy

using the quantities associated with circuits.

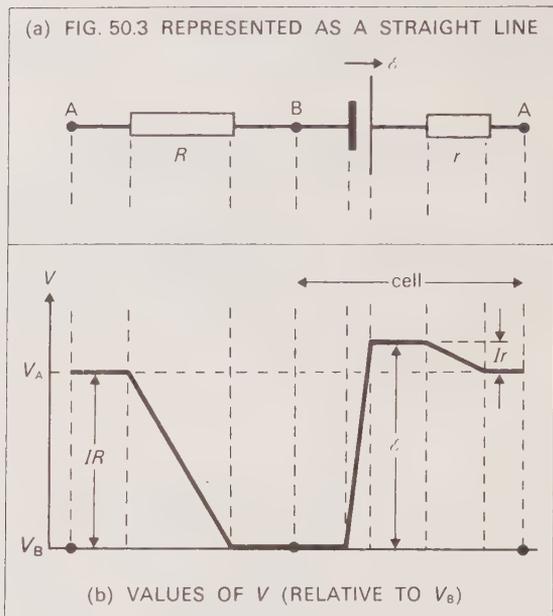


Fig. 50.4. Variation of potential from point to point in a simple circuit.

Thus they do not introduce any new ideas, and can be compared with the equation of continuity and the Bernoulli equation for fluid flow, which follow from the laws of conservation of mass and energy respectively.

Their statement

Rule I The algebraic sum of the currents at a junction is zero.

$$\Sigma I = 0$$

The rule expresses the fact that charge may not accumulate when we are dealing with steady currents.

Rule II Round any closed circuit the algebraic sum of the e.m.f.s is equal to the algebraic sum of the products of current and resistance.

$$\Sigma \mathcal{E} = \Sigma IR$$

The Loop Equation

The rule expresses the law of conservation of energy.

Solution of circuit problems

- (1) Always draw a circuit diagram.
- (2) Make sure that all the quantities given are expressed in appropriate units (e.g. A not mA).
- (3) Quote equations in symbols before substituting numbers and units.
- (4) The equations

$$V = IR, \quad P = VI$$

and

$$W = VIt = VQ$$

apply to particular devices, since in each case V is the p.d. between two particular points. State which points you are considering.

- (5) Use of Kirchhoff's Rules:

(a) **Rule I:** Mark on your circuit diagram unknown currents in terms of known currents using as few unknowns as possible. If the wrong sense is selected for a current, this will be shown by a negative sign when the equations are solved. (Use of $x, y, z,$ etc. for current symbols leads to fewer errors than $I_1, I_2, I_3,$ etc.)

(b) **Rule II:** Traverse each electrical device in a particular loop in the same sense (either clockwise or anticlockwise). The term algebraic means paying regard to sign (fig. 50.5).

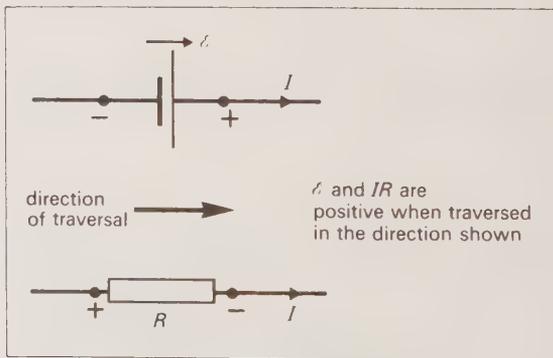


Fig. 50.5. The sign convention for Kirchhoff's Rule II.

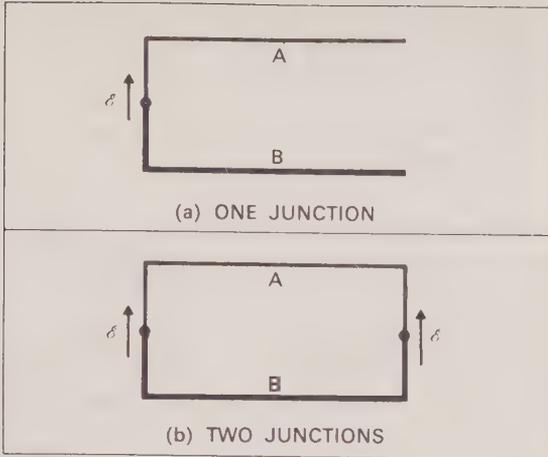


Fig. 50.6 Contact potential differences.

50.4 THERMOELECTRICITY

Contact potential difference

Different metals do not necessarily contain the same number density n of conduction electrons. If two metals are placed in contact, the conduction electrons diffuse across the boundary, showing a net flow in a direction determined by the work functions of the metals concerned. The resulting p.d. across the boundary, which is equal to the difference of the work function potentials, prevents further diffusion and so a *dynamic equilibrium* is reached. (See fig. 50.6(a).)

If the two metals are joined a second time (fig. 50.6(b)) there is no current because the second contact p.d. is equal in magnitude, but opposite in direction, to the first.

The Seebeck effect (1821)

If the temperatures of the two junctions are different, the two p.d.s will be different, and a current will flow. *The temperature difference causes an e.m.f.* (See fig. 50.7.) The device is called a **thermocouple**. The effect is the **Seebeck effect**.

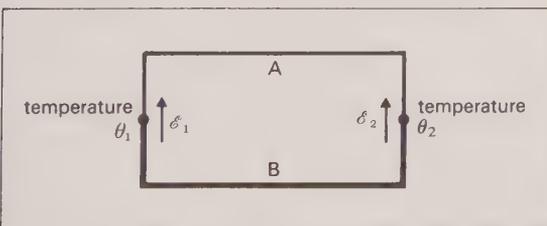


Fig. 50.7. The Seebeck effect.

► The Peltier effect (1834)

When a current is passed through a junction between dissimilar metals, the junction is heated or cooled. *The current causes a temperature difference.* This is the **Peltier effect**.

This effect is *reversible*, being proportional to I . (Cf. *Joule heating* which is proportional to I^2 .) Thus a heating effect becomes a cooling effect if the direction of current flow is reversed, and vice versa. (This is why we consider the *Seebeck effect* to produce an *electromotive force* as opposed to a *potential difference*.)

When the junctions are at the same temperature, the *Peltier effect* that accompanies a current in the circuit causes a temperature difference between junctions. The *Seebeck effect* that results takes the form of an e.m.f. which tries to drive a current in opposition to the original current (an example of *Le Chatelier's Principle*).

Laws of thermoelectricity

(A) Law of intermediate metals

If a wire C of a third metal is connected as in fig. 50.8, the net e.m.f. of the circuit (and so of the thermocouple) is not changed, provided the two A–C junctions are at the same temperature. This means that a meter *could* be used in such a circuit to measure the current. To measure the e.m.f. we would choose a potentiometer (p. 397).

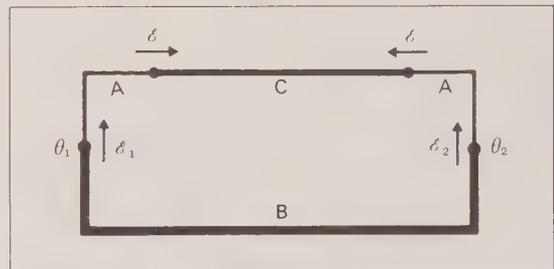


Fig. 50.8. The law of intermediate metals.

► (B) Law of intermediate temperatures

In fig. 50.9 (overleaf)

(a) the net e.m.f.s are $(\mathcal{E}_2 - \mathcal{E}_1)$ and $(\mathcal{E}_3 - \mathcal{E}_2)$ respectively,

(b) the net e.m.f. is $(\mathcal{E}_3 - \mathcal{E}_1)$.

Thus the e.m.f. when the junctions are at θ_3 and θ_1 equals the sum of the e.m.f.s when the junctions are kept first at θ_2 and θ_1 , and then at θ_3 and θ_2 .

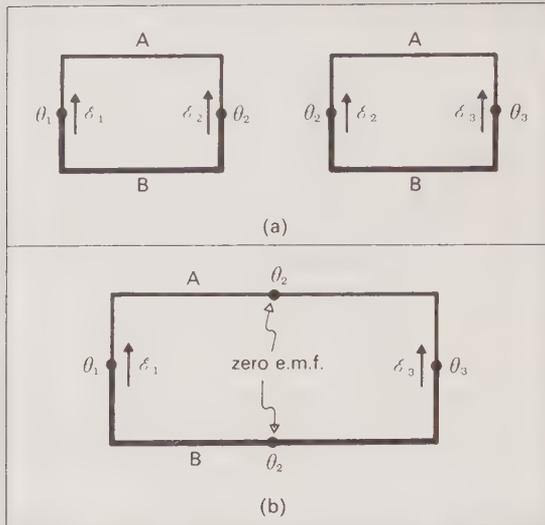


Fig. 50.9 The law of intermediate temperatures: (a) thermocouples separate, and (b) joined in series.

The e.m.f. of a thermocouple

The size of the e.m.f. (\approx mV) is too small to be useful for the large-scale production of electrical energy from thermal energy.

The e.m.f. varies with the temperature difference between the junctions according to fig. 50.10.

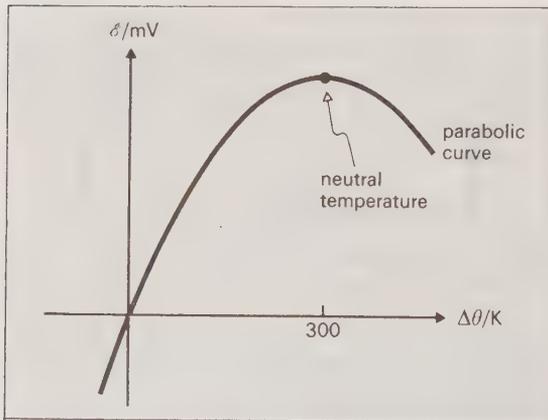


Fig. 50.10. Variation with temperature difference $\Delta\theta$ of a typical thermoelectric e.m.f.

The thermoelectric e.m.f. can be used as a *thermometric property*, i.e. to measure temperature. The e.m.f.–temperature curve is always approximately a *parabolic curve*. To measure a temperature, one

(a) places one junction where its temperature is kept steady (usually at 273 K), and

(b) chooses a pair of pure metals, alloys or semiconductors which

(i) give an approximately linear graph in the required temperature range, and

(ii) produce a big e.m.f.

(Note that fig. 50.10 would have shown a *straight line* if we had plotted \mathcal{E} against temperatures measured on the *thermocouple temperature scale*.)

► 50.5 THE RC CIRCUIT

(A) Growth of charge on capacitor plates

Connect switch to terminal A in fig. 50.11. The *loop equation* (p. 390) can be quoted as

$$\Sigma \mathcal{E} = \Sigma V$$

where V represents the *decrease of potential* in traversing a device.

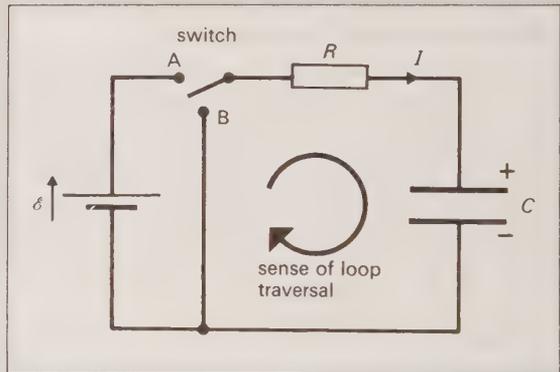


Fig. 50.11. Circuit for charge and discharge of a capacitor through a resistor.

Using $C = Q/V$, and traversing the loop clockwise

$$\mathcal{E} = IR + \frac{Q}{C}$$

$$\therefore \mathcal{E} = R \frac{dQ}{dt} + \frac{Q}{C} \quad \text{since } I = \frac{dQ}{dt}$$

This differential equation in Q has the solution

$$Q = \mathcal{E}C(1 - e^{-t/RC})$$

and this can be checked by substitution.

It follows that the charging current is given by

$$I = \frac{dQ}{dt} = \frac{\mathcal{E}}{R}(e^{-t/RC})$$

These results are shown in fig. 50.12.

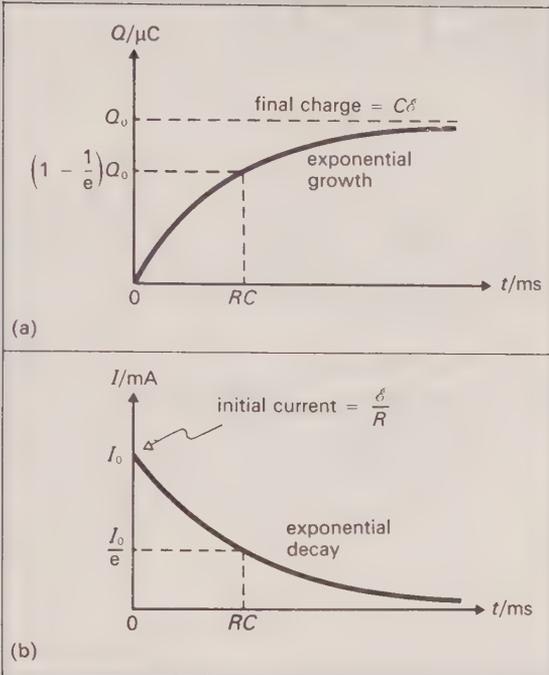


Fig. 50.12. Graphs showing variation of (a) Q and (b) I against t for a capacitor being charged.

The graphs show

(a) when $t = 0$,

$$Q = 0 \quad I_0 = \frac{\mathcal{E}}{R}$$

(b) when $t \rightarrow \infty$

$$Q \rightarrow Q_0 = C\mathcal{E} \text{ (the final charge)}$$

$$I \rightarrow 0$$

(c) when $t = RC$, then

$$Q = (1 - e^{-1})Q_0$$

i.e. Q is 0.63 of its final value Q_0 .

Similarly
$$I = \frac{I_0}{e} = 0.37 I_0$$

RC is called the **capacitive time constant** (τ_c)

$$\tau_c = RC$$

$$\tau_c [\text{s}] = R \left[\frac{\text{V}}{\text{C s}^{-1}} \right] C \left[\frac{\text{C}}{\text{V}} \right]$$

E.g. when $R = 10^3 \Omega$, and $C = 10^{-6} \text{ F}$, then

$$\tau_c = RC = 10^{-3} \text{ s.}$$

(B) Discharge through a resistor

After a time $t \gg RC$ throw the switch from A to B.

The loop equation becomes

$$0 = IR + \frac{Q}{C}$$

$$\therefore 0 = R \frac{dQ}{dt} + \frac{Q}{C}$$

which has the solution

$$Q = Q_0 e^{-t/RC}$$

where $Q = Q_0$ when $t = 0$.

When $t = RC$,

$$Q = \frac{Q_0}{e} = 0.37 Q_0$$

This is shown in fig. 50.13.

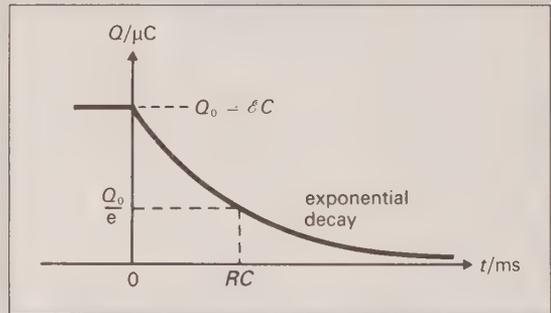


Fig. 50.13. Decay of charge on the plates of a discharging capacitor.

During discharge

$$\begin{aligned} I &= \frac{dQ}{dt} \\ &= -\frac{Q_0}{RC} e^{-t/RC} \\ &= -\frac{\mathcal{E}}{R} e^{-t/RC} \end{aligned}$$

The minus sign indicates that, in the absence of the battery, the current I flows in a direction opposite to that shown in fig. 50.11. (Refer to the note on Kirchhoff's Rule on p. 389.)

51 PRINCIPLES OF ELECTRICAL MEASUREMENTS

51.1	LABORATORY STANDARDS	394
51.2	AMMETERS AND VOLTMETERS	394
51.3	THE PRINCIPLE OF THE POTENTIOMETER	395
51.4	USES OF THE POTENTIOMETER	396

51.5	BRIDGE CIRCUITS	397
51.6	THE MEASUREMENT OF RESISTANCE	398
51.7	MULTIMETERS	399
51.8	APPROXIMATE CALIBRATION OF A VOLTMETER	399

51.1 LABORATORY STANDARDS

The **basic standards** are used in standardizing laboratories, where operations are carried out with an accuracy of 1 in 10^5 . These standards are based on fundamental definitions. They include

(a) the *Lorenz rotating disc apparatus* (p. 432) which is used to measure the resistance of standard coils,

(b) a *standard mutual inductance* (p. 440) which is used for the measurement of inductance and capacitance,

(c) a *current balance* (p. 421) which is used to measure

- (i) current, and then
- (ii) e.m.f. (by using a standard resistance coil).

In practice a laboratory will usually have a standard cell which has been calibrated by reference to the basic standards. For example the e.m.f. of one design of *Weston cadmium cell* at 293 K is 1.0186 V. Since it is a primary cell, ideally it should never be required to deliver a current of more than a few μA .

51.2 AMMETERS AND VOLTMETERS

Ammeters and voltmeters are usually adaptations of the moving-coil galvanometer (p. 418). Suppose that the galvanometer has a resistance R_G , and that full scale deflection (f.s.d.) is produced by a

current of I_G through the galvanometer. The maximum p.d. that should be applied across the terminals is given by

$$V_G = I_G R_G$$

(A) Use as an ammeter

To measure currents up to I_m , add a **shunt** (a resistor of small resistance R_S) placed in parallel, such that

$$I_G R_G = I_S R_S \quad \text{where} \quad I_S = (I_m - I_G)$$

The ammeter as a whole

- (i) has a very low resistance, and
- (ii) is connected in series with the circuit (so one has to break the circuit to insert the ammeter).

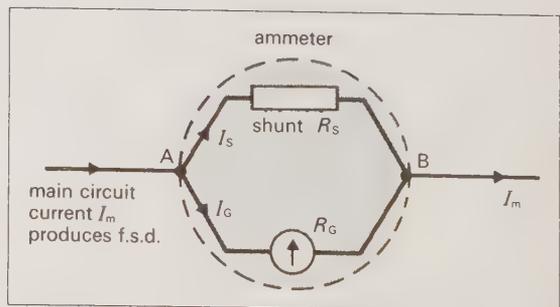


Fig. 51.1. Use of a galvanometer as an ammeter.

(B) Use as a voltmeter

To measure V_{AB} in fig. 51.2, add a **bobbin** (a resistor of large resistance R_B) in series. If V_m is the largest p.d. to be measured, then

$$V_m = I_G R_G + I_G R_B$$

The voltmeter as a whole

- (i) has a very high resistance, and
- (ii) is connected across the points A and B whose p.d. is to be measured (i.e. in parallel with the device concerned).

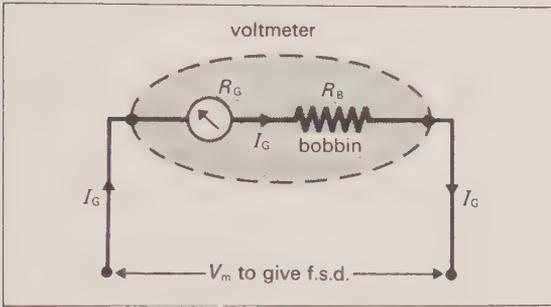


Fig. 51.2. Use of a galvanometer as a voltmeter.

A typical teaching laboratory galvanometer, for which 15 mA gives f.s.d., may have a resistance of 5.0 Ω . If it is converted into

- (a) an ammeter of f.s.d. 1.5 A, the ammeter would have a resistance of 0.05 Ω , or
- (b) a voltmeter of f.s.d. 1.5 V, the voltmeter would have a resistance of 100 Ω .

Ideal instruments

(a) The ideal *ammeter* has zero resistance so that when inserted into a circuit, it does not reduce the current that was previously flowing.

(b) The ideal *voltmeter* has infinite resistance so that it takes no current. A finite resistance causes it to take current from the circuit, and to lower the p.d. between the points to which it is connected. It registers this disturbed value. If its resistance is relatively high, the error may not be significant.

A cathode ray oscilloscope or potentiometer can be used as a voltmeter when the resistance of an ordinary voltmeter is not large relative to the device with which it is connected in parallel. These instruments will not alter significantly the p.d. they are measuring.

51.3 THE PRINCIPLE OF THE POTENTIOMETER

Fig. 51.3 shows a potentiometer circuit in which the wire is being calibrated by using a standard cell.

Potential differences are balanced against the p.d. between A and P when the galvanoscope gives a zero reading.

Causes for lack of balance

No balance point can be found if

- (a) the high-potential terminal of the p.d. under test is connected to the low-potential terminal of the potentiometer wire, and/or
- (b) the p.d. under test exceeds the p.d. between the points A and B. (In fig. 51.3, if $V_{AB} < \mathcal{E}$.)

Notes

(a) When the jockey P is correctly positioned no current flows between the potentiometer circuit and that to which it is connected. The *potentiometer does not disturb the other circuit*, whereas a voltmeter does.

(b) Using a cell of known e.m.f. \mathcal{E} , one can calibrate the wire AB as for use with a particular accumulator. (The e.m.f. of the accumulator need not be known, but it must be steady.) The wire

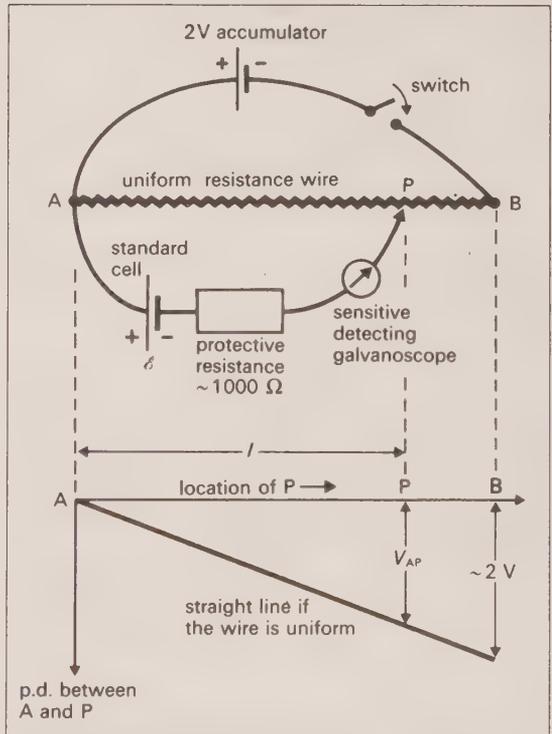


Fig. 51.3. The potentiometer principle.

then provides a continuously variable p.d. for comparison purposes.

(c) Best results are obtained when AP is as large as possible. (Cf. the metre bridge where the jockey is near the mid-point of the uniform wire.)

51.4 USES OF THE POTENTIOMETER

The potentiometer is a device which compares (or measures) potential differences.

(A) Measurement of p.d.

The potentiometer is applied exactly as would be a voltmeter. See fig. 51.4.

The potentiometer can be used to calibrate a voltmeter.

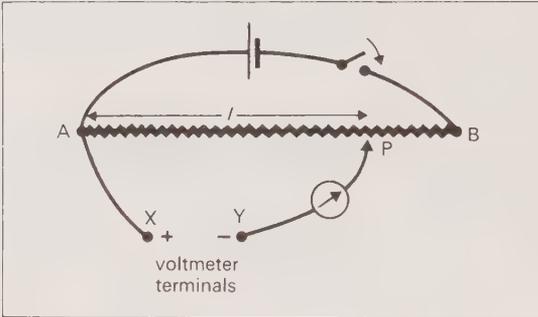


Fig. 51.4. The potentiometer as a voltmeter.

(B) Comparison of e.m.f.s

In fig. 51.3

$$\frac{\mathcal{E}_1}{\mathcal{E}_2} = \frac{l_1}{l_2}$$

One does compare genuine *e.m.f.s* (not terminal p.d.s) because at balance the cells are on open circuit.

(C) Measurement of current

The potentiometer measures the p.d. V across a standard resistance R carrying the current.

Then

$$I = \frac{V}{R}$$

This method is often used to calibrate an ammeter.

(D) Comparison of resistances

In fig. 51.5 the two resistors are carrying the same current.

$$V_1 = IR_1, \quad V_2 = IR_2$$

So
$$\frac{V_1}{V_2} = \frac{R_1}{R_2}$$

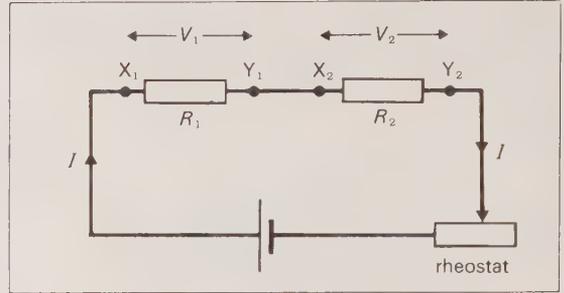


Fig. 51.5. Comparison of two resistances.

V_1 and V_2 are measured successively by connecting the terminals X and Y of fig. 51.4

- (a) to X_1 and Y_1 , and
- (b) to X_2 and Y_2 ,

Then
$$\frac{R_1}{R_2} = \frac{l_1}{l_2}$$

(E) Measurement of internal resistance

(1) Record the balance length l_0 when the switch is open: this corresponds to the e.m.f. \mathcal{E} .

(2) Record the length l when the switch is closed: this corresponds to the terminal p.d. ($\mathcal{E} - Ir$).

$$\therefore \frac{\mathcal{E} - Ir}{\mathcal{E}} = \frac{l}{l_0}$$

Using $\mathcal{E} = I(R + r)$, we have

$$\frac{R}{R + r} = \frac{l}{l_0}$$

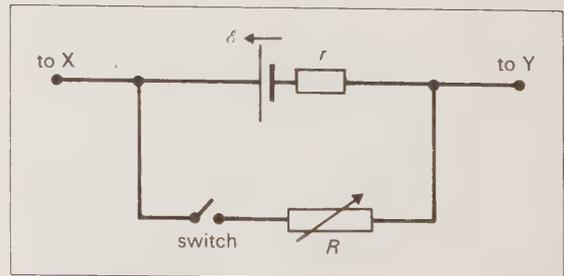


Fig. 51.6. Circuit for measuring internal resistance.

which on rearranging gives

$$\frac{1}{l} = \left(\frac{r}{l_0}\right) \frac{1}{R} + \frac{1}{l_0}$$

Because R is a known variable resistance, we can plot a straight-line graph of $1/l$ against $1/R$. The slope is r/l_0 , from which r can be found. (Note that if the graph is not a straight line, it may be that r is non-ohmic.)

(F) Measurement of a very small p.d. (\approx mV)

In principle one wants to extend the wire (to a length of perhaps 200 m) so that the balance length can be measured without a large experimental error.

In practice this is achieved by adding a large resistance in series with a metre wire: its value can be adjusted to suit the p.d. being measured.

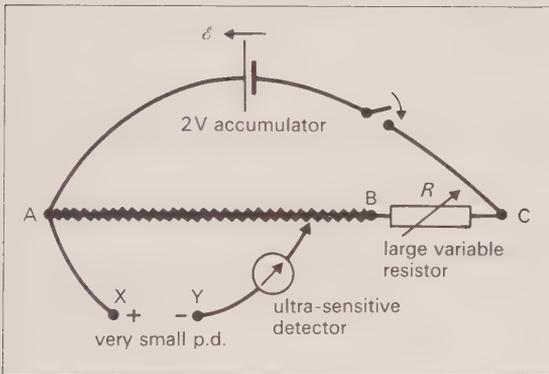


Fig. 51.7. Adapting the potentiometer to compare very small p.d.s.

In fig. 51.7,

$$V_{AB} = \left(\frac{r}{R+r}\right) V_{AC}$$

For example to measure a p.d. of up to 0.1 V we might have $V_{AC} \sim 2 \text{ V}$; $r = 4 \Omega$ and we would arrange $R = 76 \Omega$.

If the wire were to be calibrated by a standard cell we would need to put about half of this resistance between A and the accumulator. We could then measure (rather than compare) the small potential differences.

Applications

(a) To measure the e.m.f. of a thermocouple, which may be \approx millivolts.

(b) To compare small resistances. To prevent contact resistance, connections are made as shown in fig. 51.8.

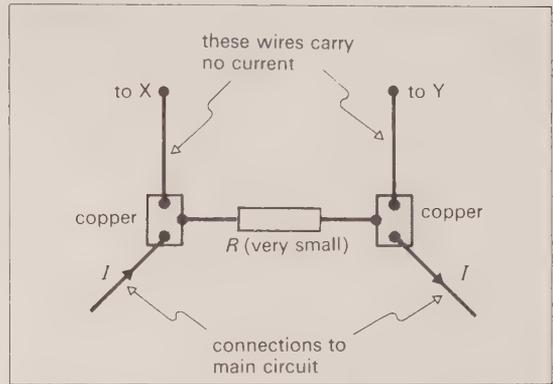


Fig. 51.8. Method of connecting a potentiometer to a very small resistance.

Because R is so small, its connecting wires must not include any resistance wire.

51.5 BRIDGE CIRCUITS

(A) The Wheatstone bridge

The resistors of fig. 51.9 are arranged in a Wheatstone bridge network.

When G shows zero deflection, B and D are at the same potential.

$$\therefore V_{AB} = V_{AD} \quad \text{and} \quad V_{BC} = V_{DC}$$

so

$$I_1 P = I_2 R \quad \text{and} \quad I_1 Q = I_2 S$$

Dividing

$$\frac{P}{Q} = \frac{R}{S}$$

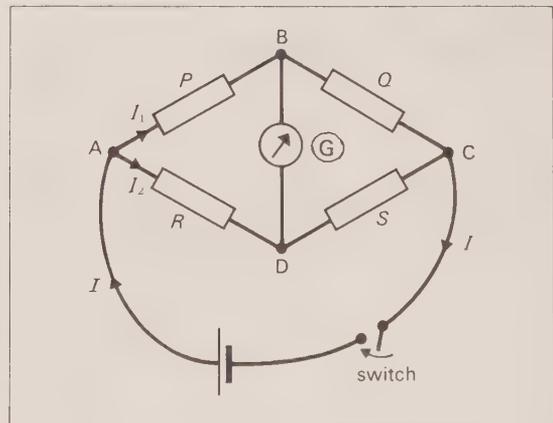


Fig. 51.9. The Wheatstone bridge network.

If any three of P , Q , R and S are known, the fourth can be found. P and Q should be interchanged to obtain a second value for their ratio.

Applications

- (a) The resistance thermometer (p. 185).
- (b) The bolometer (p. 244).
- (c) **To test whether a material obeys Ohm's Law.**

Make R and S of identical resistance coils (using any material). Make P a long thick wire, and Q a short thin wire, of the material under test.

(1) Using a small current, change the length of resistance Q so that $P = Q$, and the bridge is balanced.

(2) Change the main circuit current by a significant amount. If the bridge remains balanced (at least until heating occurs), then the material of P and Q is ohmic.

This material could now be used in conjunction with a potentiometer to calibrate a moving-coil instrument, which could then be used to test further materials as indicated on p. 375.

(B) The metre bridge

The metre bridge is the practical application of the *Wheatstone* network principle in which the ratio of two of the resistances is deduced from the ratio of their balance lengths.

The balance condition in fig. 51.10

$$\frac{P}{Q} = \frac{R}{S}$$

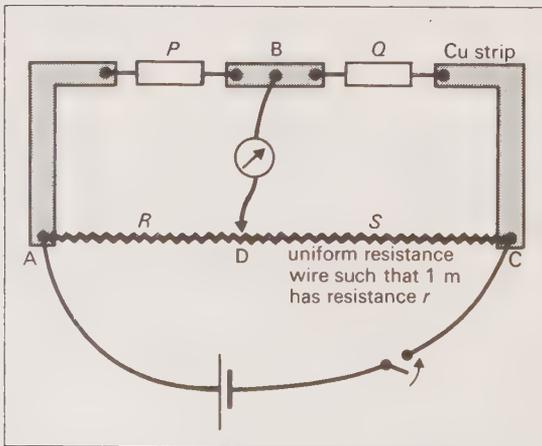


Fig. 51.10. The metre bridge.

becomes

$$\begin{aligned} \frac{P}{Q} &= \frac{rl_{AD}}{rl_{CD}} \\ &= \frac{l_{AD}}{l_{CD}} \end{aligned}$$

Notes

(a) For accuracy D should be near the mid-point of AC . (Cf. the potentiometer, where P should be near one end.)

(b) P and Q should again be interchanged, and an average value calculated for their ratio. (This procedure also makes it possible to eliminate end-corrections.)

(c) The resistance wire is heavily soldered to the copper strips so that the resistance of the join is negligible.

(d) The e.m.f. of the driving cell need not be steady, and need only be large enough to make the galvanometer deflect significantly when the bridge is just not balanced.

51.6 THE MEASUREMENT OF RESISTANCE

These methods are collected here for reference.

(A) Ammeter-voltmeter method

Using the defining equation $R = V/I$, measure I using an ammeter in series, and V with a voltmeter in parallel.

(B) Substitution method

In the circuit of fig. 51.11

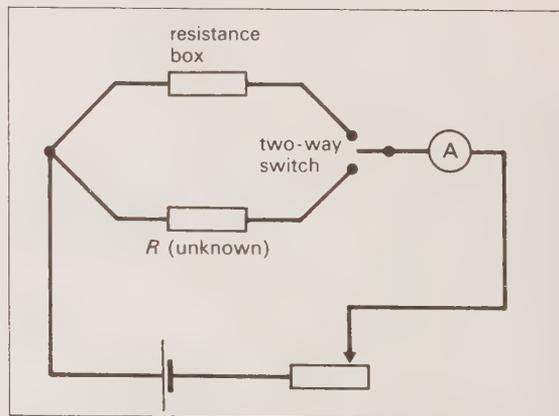


Fig. 51.11. Substitution method for R .

(1) using the unknown resistance R , adjust the rheostat until the ammeter gives a suitable reading,

(2) using the resistance box, find the (known) value that gives that same reading.

(C) Absolute measurement

Details of the *Lorenz* method are given under electromagnetic induction (p. 432).

(D) Comparison method

Compare the unknown with a known resistance, using

- (i) the *Wheatstone* or metre bridge, or
- (ii) the potentiometer (modified if the resistances are small).

51.7 MULTIMETERS

A multimeter consists of a moving-coil galvanometer, adapted so that several different scales can be used. For measuring current ranges, some internal resistors in series form a shunt across the meter. For measuring p.d. ranges, more internal resistors in series form a multiplier in series with the meter. A rotary switch can be moved to the range required. Resistance can also be measured

by using the switch position that connects an internal battery and rheostat in series with the meter and the unknown resistance. Before connecting the resistor, the terminals of the meter should be short-circuited and zeroed. The zero resistance reading will correspond to the maximum current value.

51.8 APPROXIMATE CALIBRATION OF A VOLTMETER

A suitable metal block is heated through about 10 K using an electrical coil with a steady current. The voltmeter to be calibrated is connected across the coil and the position of the pointer is marked. The electrical energy supplied can be calculated in terms of the unknown p.d. The block is returned to room temperature and is now heated mechanically using a friction band. The temperature rise should be roughly the same as before and should be achieved in about the same time so that the heat losses in each experiment are similar enough to be neglected. The energy supplied mechanically can be calculated from the friction force, the radius of the metal block and the number of revolutions. The values of the energy supplied per kelvin in the two experiments are now equated, leaving the p.d. as the only unknown so that the meter can be calibrated (in joules per coulomb).

52 THE MAGNETIC FIELD DEFINED FROM ITS EFFECT

52.1 THE MAGNETIC FORCE ON A MOVING CHARGE	400	52.8 THE HALL PROBE	405
52.2 DEFINITION OF B	401	52.9 THE FORCE ON A CURRENT-CARRYING CONDUCTOR	405
52.3 MAGNETIC FLUX Φ	401	52.10 THE TORQUE ON A CURRENT-CARRYING COIL	406
52.4 THE MOTION OF CHARGED PARTICLES IN A MAGNETIC FIELD	402	52.11 ELECTROMAGNETIC MOMENT m	407
▶ 52.5 THE CYCLOTRON	403	▶ 52.12 ENERGY AND DIPOLE ORIENTATION	408
52.6 THE HALL EFFECT	403	52.13 MEASUREMENT OF ELECTROMAGNETIC MOMENT	409
52.7 THE HALL COEFFICIENT	404		

52.1 THE MAGNETIC FORCE ON A MOVING CHARGE

Quite apart from the *Coulomb* forces, a *moving* electric charge exerts a force on another *moving* charge. The first moving charge sets up a **magnetic field**, and the second experiences a **magnetic force** because it *moves* in the magnetic field. We represent the magnitude, direction and sense of the magnetic field by the **magnetic flux density B** , which is also called the magnetic *induction*, or simply the magnetic field.

A magnetic field can be represented by **magnetic field lines**, in the same way that we represent an electric field by electric field lines. The magnetic field line is not real: it is a concept developed as an aid to visualizing the effect of a field. Field lines are drawn as in fig. 52.1 such that

(a) the *tangent* to a field line at a point gives the *direction* of B at that point, and

(b) the *number* of field lines drawn per unit cross-sectional area is proportional to the *magnitude* B . If the field is uniform, the field lines are evenly spaced.

Consider a beam of charged particles moving in a region where there exists a magnetic field, as in fig. 52.2.

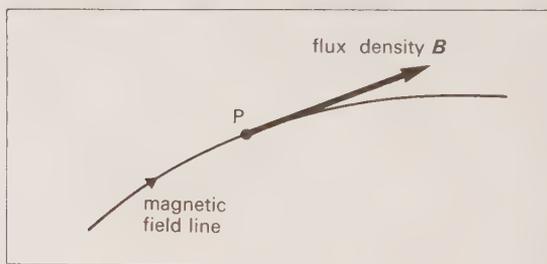


Fig. 52.1. The magnetic field line.

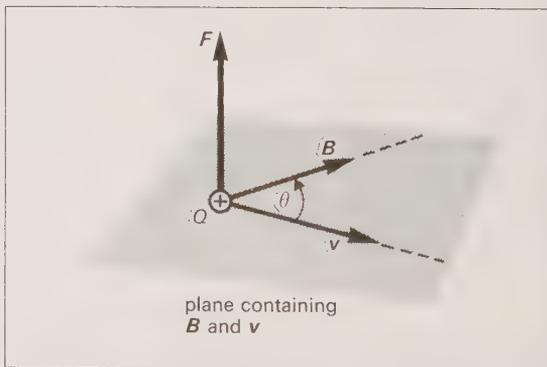


Fig. 52.2. The magnetic force on a moving charge.

By experiment, one finds

- (a) the magnetic force F is perpendicular to the particle velocity v ,
- (b) $F \propto Q$, the magnitude of the charge, but that the direction of F is reversed for a charge of opposite sign,
- (c) $F \propto$ magnitude of v ,
- (d) $F = 0$ for a particular direction of v (the direction of the magnetic field),

$$(e) \left(\begin{array}{c} \text{the magnitude} \\ F \end{array} \right) \propto \left(\begin{array}{c} \text{resolved part of } v \\ \text{normal to the line of} \\ B \end{array} \right)$$

i.e. $F \propto v \sin \theta$.

F is thus a maximum when θ is $\pi/2$ rad.

52.2 DEFINITION OF B

These experimental results are summarized by

$$F \propto Qv \sin \theta$$

We define the magnitude of **magnetic flux density** B by the equation

$$F = BQv \sin \theta$$

$$B = \frac{F}{Qv \sin \theta}$$

$$B \left[\frac{\text{N}}{\text{A m}} \right] = \frac{F[\text{N}]}{Q[\text{C}]v \left[\frac{\text{m}}{\text{s}} \right] \sin \theta}$$

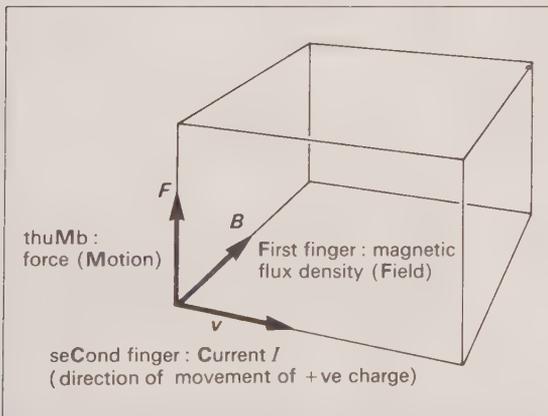


Fig. 52.3. Schematic representation of Fleming's left-hand motor rule.

B is a *pseudovector* quantity (p. 19) whose direction follows from **Fleming's left-hand motor rule** (fig. 52.3).

Notes

(a) The unit for B is the $\text{N A}^{-1} \text{m}^{-1}$, but for convenience we define a new unit called the **weber** (Wb), such that

$$1 \text{ Wb m}^{-2} = 1 \text{ N A}^{-1} \text{m}^{-1}$$

or $1 \text{ Wb} = 1 \text{ N m A}^{-1} = 1 \text{ V s}$

The unit Wb m^{-2} for B is consistent with the idea that B is a flux density.

(b) The SI unit for B is called a **tesla** (T), defined so that

$$1 \text{ T} = 1 \text{ Wb m}^{-2} = 1 \text{ N A}^{-1} \text{m}^{-1}$$

(c) $[B] = [\text{MT}^{-2}\text{I}^{-1}]$.

(d) The equation

$$B = \frac{F_m}{Qv \sin \theta}$$

which defines B , should be compared with

$$E = \frac{F_e}{Q}$$

which defines E . Note particularly the different directional relationships.

(e) Because B is a *pseudovector* quantity, we must add two or more flux densities by the parallelogram law.

Orders of magnitude for B

In Earth's field

$$B \approx 5 \times 10^{-5} \text{ T}$$

At centre of circular coil of 1000 turns, each of radius 10^{-1} m and carrying 1.5 A $B \approx 1 \times 10^{-2} \text{ T}$

Between poles of large electromagnet $B \approx 2 \text{ T}$

52.3 MAGNETIC FLUX Φ

Refer to fig. 52.4 (overleaf).

The **magnetic flux** $\Delta\Phi$ passing through the small area ΔA shown is defined by

$$\Delta\Phi = B \cos \theta \Delta A$$

$$\Delta\Phi[\text{Wb}] = B \cos \theta \left[\frac{\text{Wb}}{\text{m}^2} \right] \Delta A[\text{m}^2]$$

Φ has the dimensions $[\text{ML}^2\text{T}^{-2}\text{I}^{-1}]$

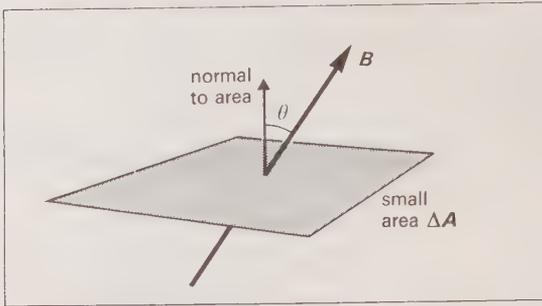


Fig. 52.4. Definition of magnetic flux Φ .

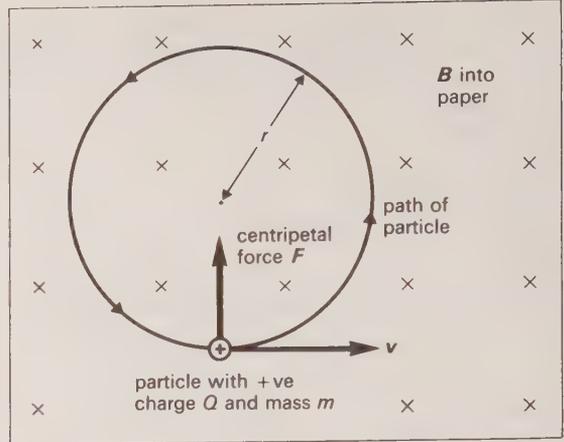


Fig. 52.5. Charged particles describing circular orbits in a uniform magnetic field.

Notes

(a) For a uniform magnetic flux density cutting a plane surface of area A

$$\Phi = BA \cos \theta$$

(b) When B is perpendicular to the surface

$$\Phi = BA$$

(c) Because of the way in which Φ is defined (by scalar product), it is a scalar quantity even though B and A are vectors.

The concept of magnetic flux is fundamental to the ideas of electromagnetic induction. We can visualize Φ as being proportional to the number of magnetic field lines that cut through a hypothetical surface placed in a field.

52.4 THE MOTION OF CHARGED PARTICLES IN A MAGNETIC FIELD

(a) If the direction of v is parallel or antiparallel to B ,

$$F = 0$$

The trajectory is a straight line.

(b) If v is perpendicular to B , then $\theta = \pi/2$ rad so

$$F = BQv$$

and the direction of F is perpendicular to the plane containing B and v at the instantaneous particle position (fig. 52.5).

The particles, moving under a constant centripetal force, describe a circle.

Using

$$F = ma$$

we have

$$BQv = \frac{mv^2}{r}$$

$$BQ = \frac{mv}{r} = m\omega$$

This shows that ω , and hence the period of revolution $T (= 2\pi/\omega)$, are independent of the speed. The radius r of the orbit is determined by the speed.

► (c) If the velocity direction makes an angle θ with B (fig. 52.6), then the particle moves in a helix, the axis of which is parallel to B .

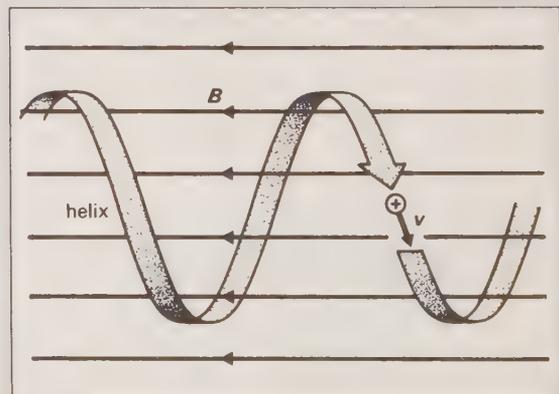


Fig. 52.6. The helical path of a charged particle moving in a magnetic field.

Its motion is the result of superposing

- (i) a uniform circular motion in which it has speed $v \sin \theta$ in a plane perpendicular to the direction of B , on
- (ii) a steady axial speed of magnitude $v \cos \theta$ along the direction of B .

The frequency f of this circular motion is given by

$$f = \frac{\omega}{2\pi} = \frac{Q}{2\pi m} B$$

and is called the **cyclotron frequency**. It does *not* depend on v , θ or r , but (for a given particle) is controlled only by B .

► 52.5 THE CYCLOTRON

Refer to fig. 52.7. The purpose of the machine is to accelerate ions from the source at the centre to high energies, before causing them to be deflected from the dees so as to bombard a target.

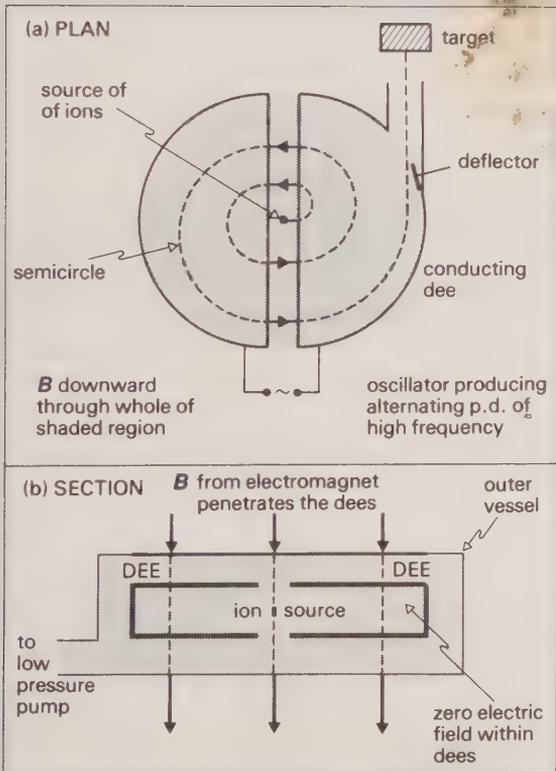


Fig. 52.7. The essential features of the cyclotron.

(a) The frequency f of the alternating p.d. is adjusted to equal the cyclotron frequency $Bq/2\pi$ (above), where q is the ions' specific charge.

(b) As the k.e. increases, f remains unchanged while r increases: *synchronism* is preserved.

(c) The energy W acquired by an ion can be found from

$$W = \left(\begin{array}{c} \text{total number of} \\ \text{transitions between} \\ \text{the dees} \end{array} \right) \times \left(\begin{array}{c} \text{p.d. between the} \\ \text{dees at the} \\ \text{relevant instant} \end{array} \right) \times \left(\begin{array}{c} \text{charge} \\ \text{on} \\ \text{ion} \end{array} \right)$$

(d) The relativistic increase of mass with gain of speed complicates the synchronism between the electric field variation and the rotation of the ions. This fact makes the cyclotron unsuitable for electrons, since they acquire very high speeds at relatively low energies.

Orders of magnitude

This simple design of cyclotron might have a dee diameter of 4.5 m, use a B of 2 T, and produce ions of energy 12 pJ. It has now been superseded, but provided useful information about the nucleus.

52.6 THE HALL EFFECT

Fig. 52.8 (overleaf) shows the consequence (for the majority carriers) of applying a transverse magnetic field to samples of *extrinsic* n-type and p-type material.

A magnetic field B causes the electric fields E_p and E_n , which are the effects. E_p and E_n are oppositely directed, and so a voltmeter connected between the points X and Y would give deflections in opposite directions, corresponding to the different polarities of the p.d.s.

Intrinsic semiconductors do show a small p.d.: this is caused by the mobility of the electrons being greater than that of the holes. The polarity of the p.d. is usually the same as that of an n-type extrinsic material.

Metals usually show the same effect as n-type material, but to a much smaller degree. Some metals (e.g. Zn) give a p.d. of the opposite polarity.

The Hall effect enables us to identify the sign of majority charge-carriers in a conductor.

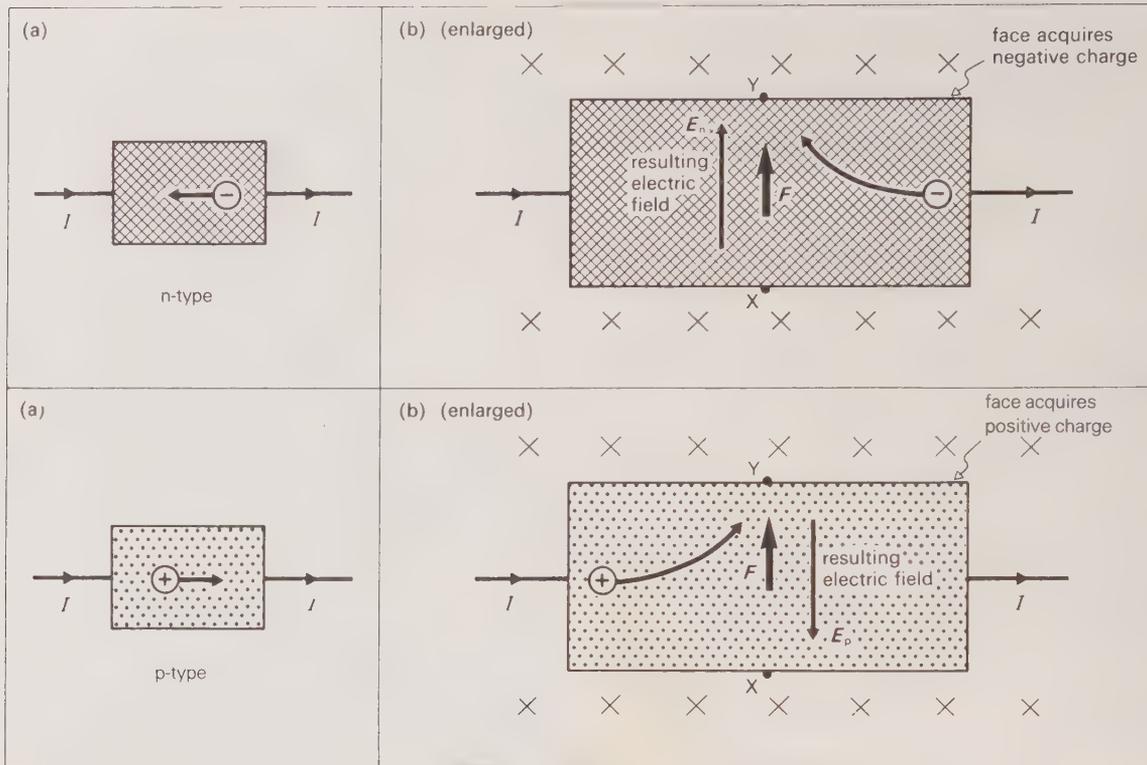


Fig. 52.8. The Hall effect: (a) zero magnetic field, and (b) magnetic field applied into paper.

52.7 THE HALL COEFFICIENT

General theory

Suppose that the material shown in fig. 52.9 (opposite) has a number density n of majority charge-carriers, each carrying a charge e .

(a) Since $I = nAev_d$ (p. 378)

the average drift velocity

$$v_d = \frac{I}{nAe} = \frac{J}{ne}$$

where J is the current density (p. 374).

(b) The transverse magnetic force F_m on each carrier is given by

$$\begin{aligned} F_m &= Bev_d \\ &= \frac{BeJ}{ne} \\ &= BJ/n \end{aligned}$$

(c) When the resulting lateral charge movement

has set up a potential gradient between X and Y, there will be an electric force F_e on the particles. This force is along the same line as F_m , but is oppositely directed.

$$\begin{aligned} F_e &= EQ \\ &= \left(\frac{V_{XY}}{w}\right)e \end{aligned}$$

since the size of E equals the potential gradient.

(d) When equilibrium is established

$$\begin{aligned} F_m &= -F_e \\ \frac{BJ}{n} &= \left(\frac{V_H}{w}\right)e \end{aligned}$$

$V_{XY} \equiv V_H$ is called the **Hall p.d.**

$$\begin{aligned} V_H &= \frac{BJw}{ne} \\ &= \frac{BI}{tne} \end{aligned} \tag{1}$$

since $A = tw$.

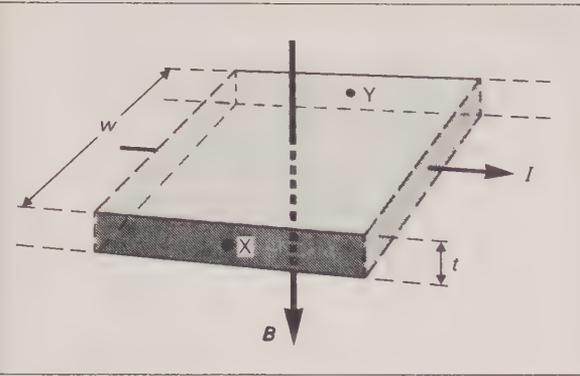


Fig. 52.9. The derivation of the Hall coefficient.

Definition

For the situation of fig. 52.9 the **Hall coefficient** R_H is defined by the equation

$$V_H = R_H(B)w$$

$$V_H \left[\frac{N m}{C} \right] = R_H \left[\frac{m^3}{C} \right] B \left[\frac{N}{A m} \right] J \left[\frac{A}{m^2} \right] w [m]$$

Notes

(a) Comparison with equation (1) shows that

$$R_H = \frac{1}{ne}$$

which is immediately consistent with the unit $m^3 C^{-1}$ for R_H .

(b) A material has a *negative* coefficient if the carriers primarily responsible for the *Hall* effect

are *electrons* (since e is then negative). It is positive if they are positive holes (p. 379).

(c) When the *Hall* effect is being demonstrated in the laboratory the points X and Y are selected by experiment (using the method implied in fig. 52.10) to ensure that $V_{XY} = 0$ when $B = 0$. The current through a specimen is a consequence of the potential gradient in the direction of J .

Orders of magnitude

The value of R_H enables us to calculate n , the number density of charge-carriers, as well as their sign. We can use this to find the number of charge-carriers provided by each parent atom.

	n/m^{-3}	$R_H/m^3 C^{-1}$	Typical $V_H/\mu V$
Cu	10^{29}	-5×10^{-11}	1
p-type Ge	10^{20} (variable)	$+10^{-2}$ (variable)	10^4

V_H is much larger for semiconducting materials as both $R_H (=1/ne)$ and v_d are greater than for metals.

52.8 THE HALL PROBE

It was shown in the previous section that the Hall p.d.

$$V_H \propto B$$

This result is used in the **Hall probe** to compare magnetic fields (fig. 52.10). The device is small and is mounted on the end of a long narrow handle for easy exploration of a field.

Procedure

- (1) The slider is adjusted, in the absence of a B -field, until X and Y are at the same potential.
- (2) The p.d. subsequently recorded between the terminals X and Y by a sensitive voltmeter is a measure of the resolved part of B perpendicular to the wafer.

52.9 THE FORCE ON A CURRENT-CARRYING CONDUCTOR

An electric current is an assembly of moving charges, and so we expect a current-carrying conductor to experience a (magnetic) force from the agent responsible for establishing any magnetic field round the conductor.

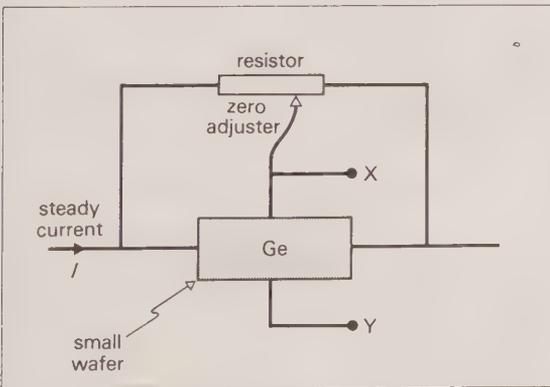


Fig. 52.10. The components of the Hall probe for comparing magnetic fields.

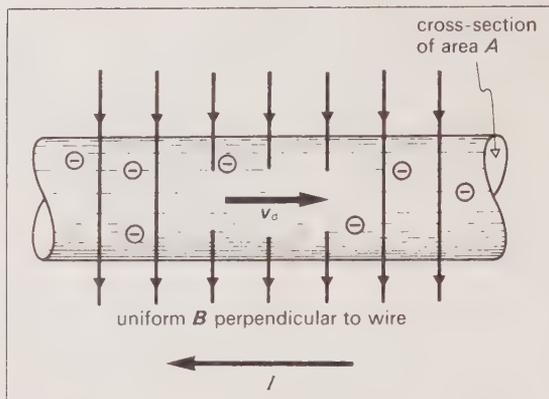


Fig. 52.11. Calculation of the magnetic force on a current-carrying conductor.

Refer to fig. 52.11.

n [m^{-3}] is the number density of conduction electrons. The size of the force F_e experienced by one electron is given by $F_e = Bev_d$, in which the drift speed

$$v_d = \frac{I}{nAe} \quad (\text{p. 378})$$

$$\therefore F_e = Be \left(\frac{I}{nAe} \right)$$

and is directed out of the paper.

In a length l of conductor, there are nAl conduction electrons. The electrons experience a total force F , where

$$F = (nAl)Be \left(\frac{I}{nAe} \right) = BIl$$

This is also the force experienced by the conductor, as the electrons are not free to move outside the lattice, and experience a restraining force from the lattice.

The direction of the force acting on the wire can be found from *Fleming's left-hand motor rule*: it is perpendicular to the plane defined by B and the wire.

If the field makes an angle θ with the wire

$$F = BIl \sin \theta$$

$$F[\text{N}] = B \left[\frac{\text{N}}{\text{A m}} \right] I[\text{A}] l[\text{m}] \sin \theta$$

Note that this equation could have been used to define B , but that it is more fundamental to use the force on a moving charge.

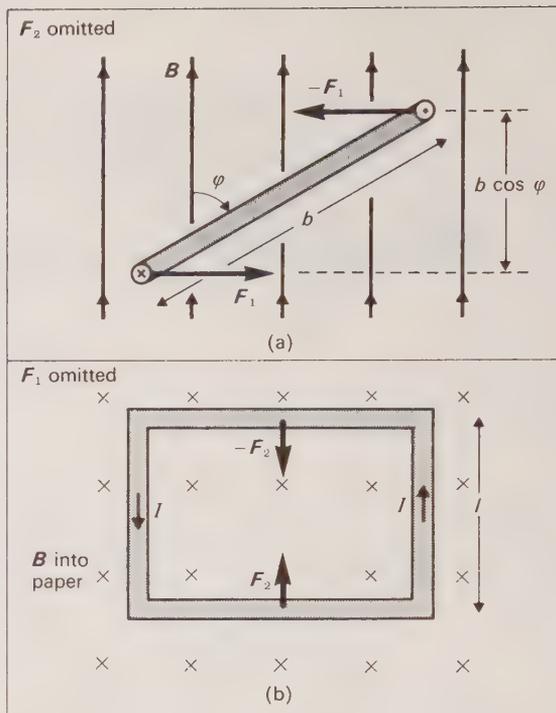


Fig. 52.12. Calculation of the resultant torque acting on a rectangular coil: (a) plan, and (b) section.

52.10 THE TORQUE ON A CURRENT-CARRYING COIL

(A) Uniform field

Consider a rectangular coil of length l and breadth b placed in a uniform magnetic field of flux density B (fig. 52.12).

The resultant of the forces F_2 is zero, since they have the same line of action.

The forces F_1 constitute a couple, since they are equal but antiparallel.

The torque acting on the coil

$$\begin{aligned} T &= F_1 \times (\text{arm of couple}) \quad (\text{p. 67}) \\ &= BIl \times b \cos \phi \\ &= BI \times \text{area} \times \cos \phi \end{aligned}$$

Note that here ϕ is the angle between the plane of the coil and B .

(B) Radial field

If we use circular pole-pieces and a soft-iron cylindrical core, then the flux density in the air

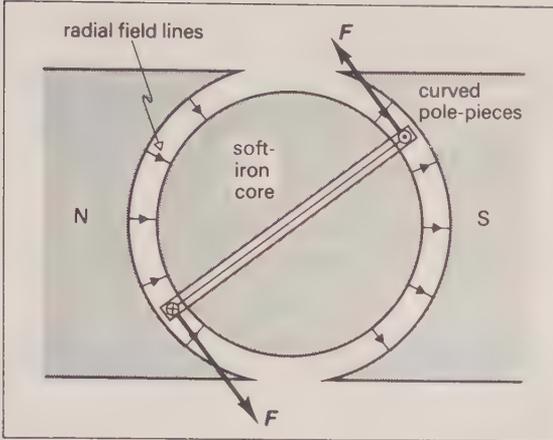


Fig. 52.13. A radial field produces a constant torque.

gap is radial (fig. 52.13). This means that the plane of the coil is always parallel to B , so

$$\phi = 0 \text{ always}$$

and

$$\text{torque } T = IAB$$

(C) Torque acting on any coil

It can be shown that for coils of any shape, carrying N turns of wire, and having area A

$$T = NIAB \cos \phi$$

$$T[\text{N m}] = NI[\text{A}]A[\text{m}^2]B\left[\frac{\text{N}}{\text{A m}}\right] \cos \phi$$

where ϕ is defined in fig. 52.12.

These results are applied in the moving-coil galvanometer (p. 418) and d.c. motor (p. 434).

52.11 ELECTROMAGNETIC MOMENT m

(A) Current-carrying coil

We will now work in terms of the angle θ (rather than ϕ), where $\theta = (90^\circ - \phi)$, and is defined in fig. 52.14.

Since

$$T = NIAB \sin \theta$$

$$T_{\text{max}} = NIAB$$

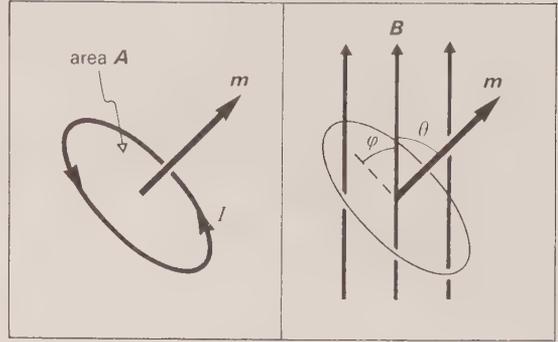


Fig. 52.14. Electromagnetic moment m of a coil.

The **electromagnetic moment** m of a current-carrying coil is defined by the equation

$$m = \frac{T_{\text{max}}}{B}$$

$$m[\text{A m}^2] = \frac{T_{\text{max}}[\text{N m}]}{B\left[\frac{\text{N}}{\text{A m}}\right]}$$

m is thus *numerically* equal to the torque required to hold the coil with its normal at right angles to a uniform flux density of size

$$B = 1 \text{ T} = 1 \text{ Wb m}^{-2} = 1 \text{ N A}^{-1} \text{ m}^{-1}$$

Notes

(a) m is a *vector* whose direction is given in fig. 52.14. It is sometimes called simply **magnetic moment**.

(b) Since $m = NIA$, we see at once that A m^2 is a consistent unit for m , which has the dimensions $[\text{L}^2 \text{ I}]$.

(c) The torque T acting on the coil tends to align m and B , and has a magnitude

$$T = mB \sin \theta$$

where θ is the angle between B and the *normal* to the coil.

(d) By defining m in terms of B rather than H (p. 465), we are following the *Sommerfeld* approach to electromagnetism. (A definition expressed in terms of H would follow the *Kennelly* approach.)

(B) Permanent magnetic dipole

A current-carrying coil and a permanent bar magnet have, to a first approximation, similar *external*

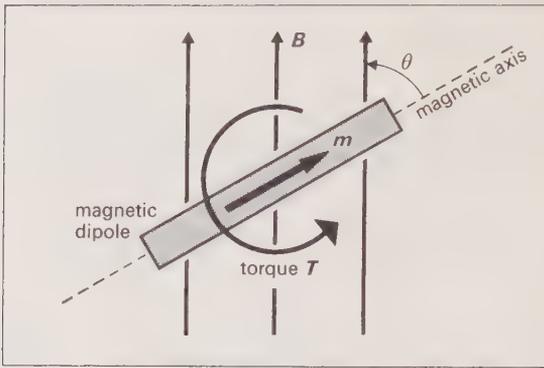


Fig. 52.15. The electromagnetic moment m of a permanent magnetic dipole.

fields. This suggests that we can extend the concept of electromagnetic moment to such a dipole (fig. 52.15).

The electromagnetic moment m of the dipole is again defined by the equation

$$m = \frac{T_{\max}}{B}$$

Notes

(a) m is a vector whose direction is given in fig. 52.15.

(b) The torque T acting on the dipole tends to align m and B , and is given by $T = mB \sin \theta$. Here θ is the angle between B and the dipole's magnetic axis.

(C) Comparison with electric dipole moment

The reader should study fig. 52.16 and its equations very carefully.

To carry the analogy further, we can consider the magnet to have poles, called N and S, near the ends, and with which we can associate a particular magnetic pole strength. The direction of m is from S to N, and its magnitude would be proportional to

$$(\text{pole strength}) \times 2d$$

where $2d$ is the pole separation. This idea is not developed in this book.

► 52.12 ENERGY AND DIPOLE ORIENTATION

Suppose that in fig. 52.14 an external torque T increases the value of θ by $\delta\theta$, the axis of T being parallel to that of $\delta\theta$. The external agent does work

$$\begin{aligned} \delta W &= T \cdot \delta\theta \\ &= mB \sin \theta \delta\theta \end{aligned}$$

If we define the p.e. E_p of the system (the dipole and the currents responsible for the external field) such that $E_p = 0$ when $\theta = \pi/2$ rad, then

$$\begin{aligned} E_p &= \int_{\pi/2 \text{ rad}}^{\theta} mB \sin \theta \, d\theta \\ &= [-mB \cos \theta]_{\pi/2 \text{ rad}}^{\theta} \\ &= -mB \cos \theta \end{aligned}$$

Notes

(a) The fact that E_p is negative when $\theta < \pi/2$ rad results from our arbitrary choice of the zero of p.e.

(b) When $\theta = \pi$ rad, the p.e. is a maximum. This occurs when the directions of m and B are anti-parallel.

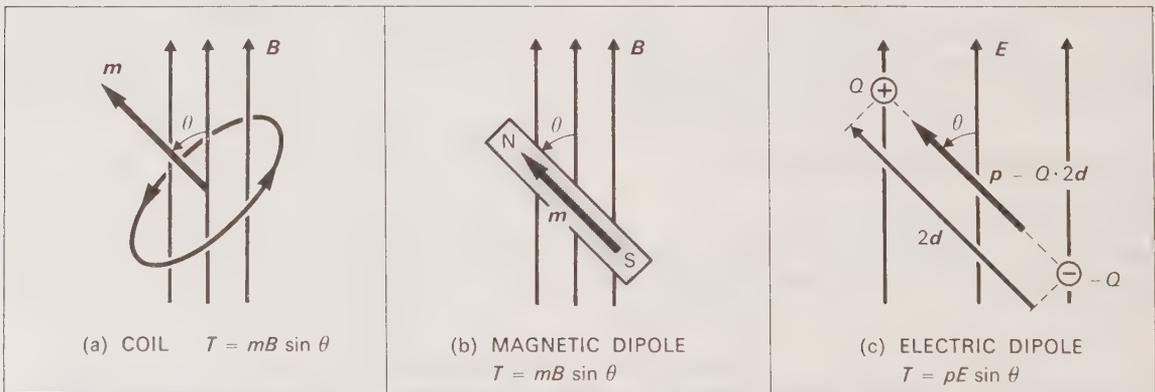


Fig. 52.16. Comparisons of the electromagnetic moments of a coil and magnetic dipole with the moment of an electric dipole.

(c) An amount of energy $2mB$ is needed to change θ from 0 to π rad. This result is important in spectroscopy.

52.13 MEASUREMENT OF ELECTROMAGNETIC MOMENT

(a) Use the *defining equation*

$$T = mB \sin \theta$$

Measure values of T , B and θ , from which m can be found.

(b) *Oscillation method*

Suspend the dipole so that it can rotate about its centre of mass. Allow it to oscillate about its equilibrium position in a uniform magnetic field B (such as that of the Earth).

The torque

$$T = -mB \sin \theta = I\ddot{\theta}$$

where the negative sign shows that the torque acts so as to decrease θ , and I is the moment of inertia of the suspended system.

For small oscillations

$$\ddot{\theta} = -\left(\frac{mB}{I}\right)\theta$$

The motion is simple harmonic, of time period

$$\frac{2\pi}{\omega} = 2\pi\sqrt{\frac{I}{mB}}$$

which is independent of the amplitude.

m can be calculated from a knowledge of the time period, I and B . We assume that the suspension fibre exerts zero torque.

53 THE MAGNETIC FIELD RELATED TO ITS CAUSE

53.1	THE MAGNETIC FIELD CREATED BY MOVING CHARGES	410
53.2	THE BIOT–SAVART LAW	410
53.3	COMPARISONS OF THREE IMPORTANT FIELDS	411
▶ 53.4	THE MAGNETIC FIELD ESTABLISHED BY COMPLETE CIRCUITS	412

53.5	SUMMARY OF SECTION 53.4	414
53.6	EXPERIMENTAL STUDY OF MAGNETIC FIELDS	415
▶ 53.7	AMPÈRE'S LAW	415
53.8	THE MAGNETIC FORCES EXERTED ON EACH OTHER BY CURRENT-CARRYING CONDUCTORS	415
53.9	SUMMARY OF THE MAGNETIC FIELD	417

53.1 THE MAGNETIC FIELD CREATED BY MOVING CHARGES

In chapter 52 we discussed the *effect* of a magnetic field on a moving charge. This effect was used to define the vector B .

In this chapter we discuss the *cause* of the magnetic field, which is the movement of electric charge. One *moving* charge exerts a force on a second *moving* charge (and vice versa) through their *magnetic interaction*.

(a) *Oersted* (1820) demonstrated that a magnetic field exists near a current-carrying wire. We infer that it is the moving charge in the wire that creates the field.

(b) *Rowland* (1878) demonstrated that such a field also exists close to an insulated charged disc which is rotated at high speed. This shows more fundamentally that electric charges in motion create the field. (Its magnitude was $\sim 10^{-7}$ tesla.)

In both experiments the existence of the field was shown by the torque experienced by a small magnetic dipole.

53.2 THE BIOT–SAVART LAW

Statement of the law

Suppose a positive charge Q moves with velocity v as shown in fig. 53.1.

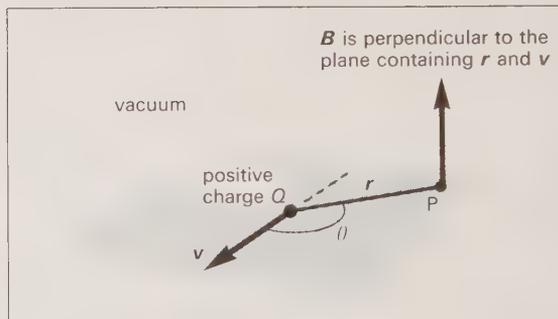


Fig. 53.1. The field created by a moving charge.

The magnitude of the magnetic field B at P is given by

$$B \propto \frac{Qv \sin \theta}{r^2}$$

where r and θ are defined in the diagram, which also shows the direction of B .

The permeability of free space μ_0

We write the constant of proportionality as $\mu_0/(4\pi)$ in the *rationalized* SI, leading to the equation

$$B = \left(\frac{\mu_0}{4\pi} \right) \cdot \frac{Qv \sin \theta}{r^2}$$

μ_0 is called the **permeability of free space**, or sometimes simply the **magnetic constant**.

Notes

(a) According to this defining equation μ_0 can have the unit $\text{Wb A}^{-1} \text{m}^{-1}$, but is more commonly expressed in the equivalent unit *henry m*⁻¹ (p. 437). It has the dimensions $[\text{MLT}^{-2}\text{I}^{-2}]$.

(b) The *ampere* is defined (p. 9) in such a way that we assign a value to μ_0 . (It cannot, therefore, be found by experiment.)

Thus

$$\mu_0 = 4\pi \times 10^{-7} \text{ Wb A}^{-1} \text{m}^{-1}$$

(c) The introduction of the 4π into this equation (*rationalization*) means that it does not appear in *Ampère's Law* (p. 415).

The Biot–Savart law for a current element

A charge Q moving at velocity v sets up a magnetic field identical to that established by a current I

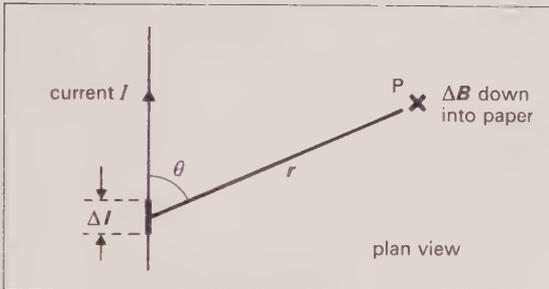


Fig. 53.2. The field set up by a current element $I \Delta l$.

in a conductor of length Δl (fig. 53.2), where

$$Qv = I \Delta l$$

Then the contribution ΔB of the **current element** $I \Delta l$ to the total value of B at P is given by

$$\Delta B = \left(\frac{\mu_0}{4\pi} \right) \cdot \frac{I \Delta l \sin \theta}{r^2}$$

when Δl is very small.

Notes

(a) This situation could have been used to quote the law, but it is more fundamental to use the field set up by a moving charge.

(b) The law can be tested experimentally only with difficulty, but is justified because all its *predictions* are in accord with experimental results.

(c) This equation is used to find B in simple situations (below).

The direction of the magnetic field

We defined the sense and direction of B on p. 401. We work out the sense and direction from either

- Maxwell's right-hand cork-screw rule, or
- the right-hand grip rule.

Figs. 53.1 and 53.2 both illustrate the result.

53.3 COMPARISONS OF THREE IMPORTANT FIELDS

Field type	Cause of field	Definition of field strength	Usual statement of inverse square law	Calculation of field from point cause	Gauss's law (for a closed surface)
gravitational	m	$g = \frac{F}{m}$	Newton $F = G \frac{m_1 m_2}{r^2}$	$g = G \frac{m}{r^2}$	$\left(\frac{1}{4\pi G} \right) \psi_g = -\Sigma m$
electric	Q	$E = \frac{F}{Q}$	Coulomb $F = \left(\frac{1}{4\pi\epsilon_0} \right) \frac{Q_1 Q_2}{r^2}$	$E = \left(\frac{1}{4\pi\epsilon_0} \right) \frac{Q}{r^2}$	$\epsilon_0 \psi_E = \Sigma Q$
magnetic	$Qv \sin \theta$	$B = \frac{F}{Qv \sin \theta}$	Biot–Savart (right)	$B = \left(\frac{\mu_0}{4\pi} \right) \frac{Qv \sin \theta}{r^2}$	$\Phi = 0$

► 53.4 THE MAGNETIC FIELD ESTABLISHED BY COMPLETE CIRCUITS

Procedure

- (1) Write down ΔB in terms of a single variable, and
- (2) sum or integrate between the limits applicable to the problem.

The results of this section are summarized on p. 414 for the student who wants to bypass the mathematics.

(A) B at the centre of a circular coil

Suppose the coil has N turns each of radius r and carries a current I . The field at the centre is given by

$$\begin{aligned}
 B &= \Sigma \Delta B \\
 &= \Sigma \left(\frac{\mu_0}{4\pi} \right) \frac{I \Delta l \sin \theta}{r^2} \\
 &= \left(\frac{\mu_0}{4\pi} \right) \frac{I \sin 90^\circ}{r^2} \Sigma \Delta l \\
 &= \frac{\mu_0}{4\pi} \cdot \frac{I}{r^2} \cdot N 2\pi r \\
 \therefore B &= \frac{\mu_0 I N}{2} \cdot \frac{1}{r} \\
 B \left[\frac{\text{Wb}}{\text{m}^2} \right] &= \frac{\mu_0 \left[\frac{\text{Wb}}{\text{A m}} \right] \cdot I [\text{A}] \cdot N}{2r [\text{m}]}
 \end{aligned}$$

(B) B at any point on the axis of a circular coil

In fig. 53.3 we sum ΔB , taking its direction into account.

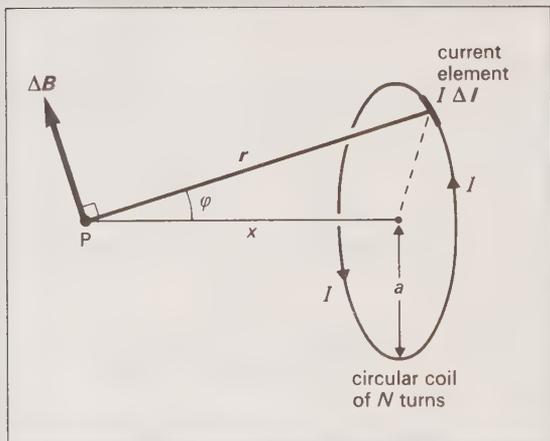


Fig. 53.3. Field at any point along the axis of a coil.

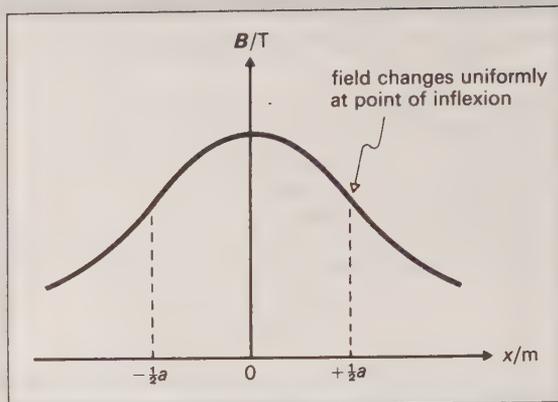


Fig. 53.4. Variation with distance x of the value of B along the axis of a coil.

(a) By symmetry those components of ΔB perpendicular to the axis sum to zero.

(b) Each ΔB has a component parallel to the axis, of magnitude.

$$\Delta B \cdot \cos(90^\circ - \phi) = \left(\frac{\mu_0}{4\pi} \right) \frac{I \Delta l \sin 90^\circ}{r^2} \cdot \sin \phi$$

The magnetic field at P is given by

$$\begin{aligned}
 B &= \Sigma \left(\frac{\mu_0}{4\pi} \right) \frac{I \Delta l}{r^2} \cdot \sin \phi \\
 &= \left(\frac{\mu_0}{4\pi} \right) \cdot \frac{2\pi a N I \sin \phi}{r^2} \\
 &= \frac{\mu_0 I N}{2} \cdot \frac{a^2}{(a^2 + x^2)^{3/2}}
 \end{aligned}$$

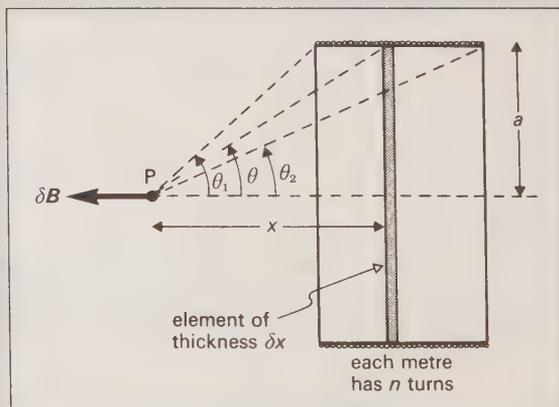


Fig. 53.5. Calculation of B due to a solenoid.

This curve is plotted in fig. 53.4. It is nearly linear at the point of inflexion, given by $x = \pm(a/2)$. (See the *Helmholtz coil system*, below.)

(C) B along the axis of a solenoid

Using result (B), the contribution δB of the infinitesimal element δx of fig. 53.5 is given by

$$\delta B = \frac{\mu_0 a^2 (n \delta x) I}{2(a^2 + x^2)^{3/2}}$$

where $(n \delta x)$ is the number of turns in the element.

Write all the variables in terms of θ :

$$x = a \cot \theta$$

$$\delta x = -a \operatorname{cosec}^2 \theta \delta \theta$$

$$(x^2 + a^2) = a^2 \operatorname{cosec}^2 \theta$$

Then
$$\delta B = \frac{\mu_0 a^2 n (-a \operatorname{cosec}^2 \theta \delta \theta) I}{2a^3 \operatorname{cosec}^3 \theta}$$

$$= -\frac{\mu_0}{2} n I \sin \theta \delta \theta$$

Note that the radius a cancels.

Let $\delta \theta \rightarrow 0$, and integrate

$$\begin{aligned} B &= \int_{\theta_1}^{\theta_2} -\frac{\mu_0}{2} n I \sin \theta \, d\theta \\ &= \frac{\mu_0}{2} n I (\cos \theta_2 - \cos \theta_1) \end{aligned}$$

Special cases

(a) When P is inside an *infinitely long* solenoid

$$\left. \begin{aligned} \theta_2 &= 0 \\ \theta_1 &= \pi \text{ rad} \end{aligned} \right\} \text{so } B = \mu_0 n I$$

(b) For P inside a toroidal solenoid (a **toroid**, such as that in fig. 57.2, p. 437), $B = \mu_0 n I$.

(c) For P at one end of a long solenoid

$$\left. \begin{aligned} \theta_2 &= 0 \\ \theta_1 &= \frac{\pi}{2} \text{ rad} \end{aligned} \right\} B = \frac{\mu_0 n I}{2}$$

Note carefully that n has unit *turns* m^{-1} , whereas N was measured in *turns*.

It can be shown, using *Ampère's Law* (p. 415), that B is constant over the whole cross-section of a long straight solenoid.

(D) The Helmholtz coil arrangement

Two identical coils of radius a are placed a distance a apart. This achieves a magnetic field of excep-

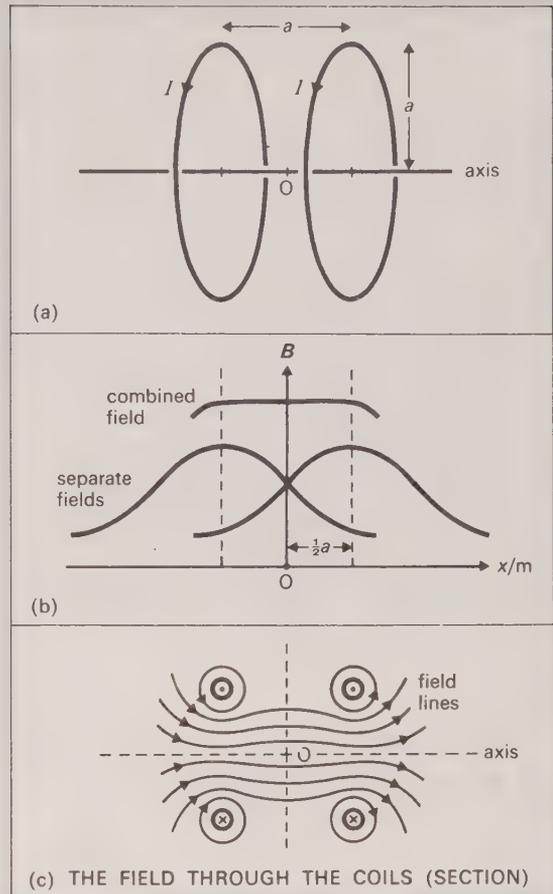


Fig. 53.6. The Helmholtz coils: (a) the arrangement, (b) and (c) the resulting field.

tional uniformity over a large region. (See, for example, p. 471.)

At the mid-point O , $x = a/2$.

$$\text{Using result (B) } B = 2 \frac{\mu_0 I N}{2} \cdot \frac{a^2}{\left[a^2 + \left(\frac{a}{2} \right)^2 \right]^{3/2}}$$

$$= \frac{8\mu_0 N I}{\sqrt{125} a}$$

(E) The field due to a long straight wire

In fig. 53.7 (overleaf) the contribution δB of the infinitesimal element δl is given by

$$\delta B = \left(\frac{\mu_0}{4\pi} \right) \cdot \frac{I \delta l \sin \theta}{r^2}$$

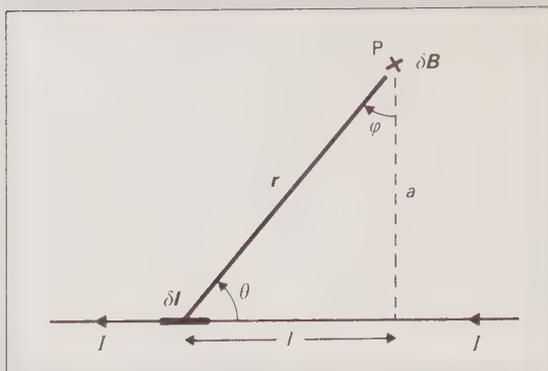


Fig. 53.7. Field due to a long straight wire.

Write all the variables in terms of ϕ :

$$\begin{aligned} l &= a \tan \phi \\ \delta l &= a \sec^2 \phi \delta \phi \\ r &= a \sec \phi \\ \sin \theta &= \cos \phi \end{aligned}$$

Then

$$\begin{aligned} \delta B &= \left(\frac{\mu_0}{4\pi} \right) \cdot \frac{I a \sec^2 \phi \delta \phi \cos \phi}{a^2 \sec^2 \phi} \\ &= \left(\frac{\mu_0}{4\pi} \right) \frac{I}{a} \cos \phi \delta \phi \end{aligned}$$

Let $\delta \phi \rightarrow 0$, and integrate:

$$\begin{aligned} B &= \left(\frac{\mu_0}{4\pi} \right) \cdot \frac{I}{a} \int_{\phi_1}^{\phi_2} \cos \phi \, d\phi \\ &= \left(\frac{\mu_0}{4\pi} \right) \cdot \frac{I}{a} (\sin \phi_2 - \sin \phi_1) \end{aligned}$$

For the special case of an *infinitely long wire*

$$\left. \begin{aligned} \phi_2 &= \frac{\pi}{2} \text{ rad} \\ \phi_1 &= -\frac{\pi}{2} \text{ rad} \end{aligned} \right\} B = \frac{\mu_0 I}{2\pi a}$$

53.5 SUMMARY OF SECTION 53.4

Situation	Symbols	Result
(a) Point at centre of circular coil	r = coil radius N = number of turns	$B = \frac{\mu_0 IN}{2r}$
(b) Any point on axis of circular coil	a = coil radius N = number of turns x = distance of point from coil centre	$B = \frac{\mu_0 IN}{2} \cdot \frac{a^2}{(a^2 + x^2)^{3/2}}$
(c) Point inside a very long solenoid or toroid Point at one end of a long solenoid	n = number of turns per unit length	$B = \mu_0 n I$ $B = \frac{\mu_0 n I}{2}$
(d) Point at centre of Helmholtz coil system	N = number of turns on each coil a = coil radius	$B = \frac{8\mu_0 NI}{\sqrt{125}a}$
(e) Point close to a very long straight wire	a = perpendicular distance from point to wire	$B = \frac{\mu_0 I}{2\pi a}$
In each result B can be written $\frac{(\text{dimensionless constant})\mu_0 I}{\text{length}}$.		Note that $[n] = [L^{-1}]$.

53.6 EXPERIMENTAL STUDY OF MAGNETIC FIELDS

The justification for the *Biot–Savart* law lies in the agreement between the predictions summarized in the table opposite and the values of B that are actually observed. This section indicates how the fields are measured.

(a) *Search coils.* (i) If the circuit whose field is to be investigated carries an *alternating* current then it will establish an alternating field. If this field is made to thread a small search coil an alternating e.m.f. will be induced across the coil (p. 426). The e.m.f. can be measured by a c.r.o., and has a size proportional to the peak value of B , and to the frequency.

(ii) If the circuit carries a *direct* current the field can be found by using a search coil in conjunction with a ballistic galvanometer in the way described on p. 428. Experimentally the method is not accurate, but it is the only one available in certain circumstances (see p. 459).

(b) *Force on a wire.* In principle the method used for the absolute measurement of current (p. 421) can be adapted for finding B if I is known, using $F = BIl$. In practice the method is not sensitive.

(c) *The Hall probe* (p. 405). This is probably the best way of comparing the fields in question. Allowance is necessary for the Earth's field, so the method is not suitable for very small fields. (Cf. (a)(i).)

configuration of current-carrying conductors in a vacuum (p. 465).

Uses of the law

(a) For finding the magnetic field in situations which possess the necessary geometrical symmetry. It can be used for the long straight wire, the toroid and the solenoid.

(b) When applied to a long straight solenoid, it shows that B is constant over the whole cross-section (i.e. has the same value at any distance from the axis).

The law is an alternative to the *Biot–Savart Law*. (Cf. *Coulomb's Law* and *Gauss's Law*.)

53.8 THE MAGNETIC FORCES EXERTED ON EACH OTHER BY CURRENT-CARRYING CONDUCTORS

At all points along the right wire of fig. 53.8 the field B created by the left wire is given by

$$B = \frac{\mu_0 I_1}{2\pi a}$$

Using $F = BIl$, the magnitude of the force F experienced by a length l of the right wire is given by

$$F = \left(\frac{\mu_0}{2\pi}\right) \cdot \frac{I_1 I_2}{a} \cdot l$$

► 53.7 AMPÈRE'S LAW

Demonstration of Ampère's Law

The field produced by an infinitely long straight conductor is given by

$$B = \frac{\mu_0 I}{2\pi a}$$

so $B(2\pi a) = \mu_0 I$

This illustrates **Ampère's Law** which is

$$\oint B \cdot dl = \mu_0 I$$

where

(a) $\oint B \cdot dl$ is the *line integral* round a closed loop, and

(b) I is the current enclosed by the loop.

The law holds for any closed path about any

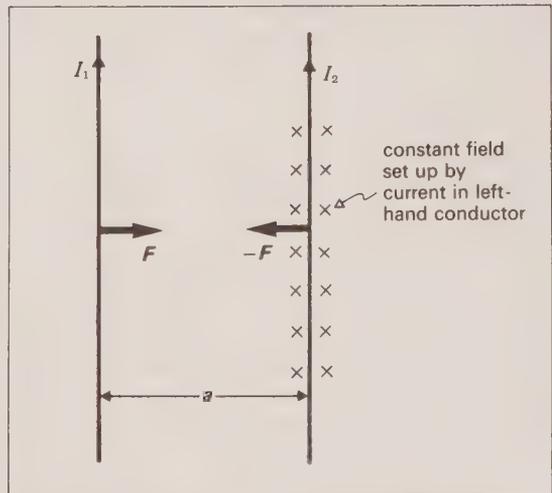


Fig. 53.8. The magnetic forces exerted by current-carrying conductors.

Notes

(a) When the currents are in the same direction, the conductors attract each other. When they are in opposite directions, they repel. This can be shown by *Fleming's left-hand motor rule*.

(b) *Newton's Third Law* tells us that two such circuits exert equal and opposite forces on each other.

(c) The two wires of fig. 53.8 exert magnetic forces on each other even when

(i) they are both *electrically neutral*, and

(ii) they are both made of material which is *not ferromagnetic*.

This *fundamental magnetic effect* was first studied by *Ampère* (1822).

Correlation with the ampere

Suppose we assume the definition of the ampere:

The ampere is that steady current which, flowing in two infinitely long straight parallel conductors of negligible circular cross-sectional area placed 1 m apart in a vacuum, causes each wire to exert a force of 2×10^{-7} N on each metre of the other wire.

Then in

$$F = \left(\frac{\mu_0}{2\pi}\right) \cdot \frac{I_1 I_2}{a} \cdot l$$

we have

$$F = 2 \times 10^{-7} \text{ N}$$

$$I_1 = I_2 = 1 \text{ A}$$

$$a = l = 1 \text{ m}$$

Substituting, we find

$$\mu_0 = 4\pi \times 10^{-7} \text{ Wb A}^{-1} \text{ m}^{-1}$$

(As an alternative procedure we could have taken the assigned value of μ_0 , and shown that it leads to a result consistent with the definition of the ampere.)

The definition of the current standard of the SI

(1) *assigns* a value to μ_0 ,

(2) enables us to define the *coulomb*, which leads us

(3) to an *experimental* value for ϵ_0 (p. 337).

► The relationship between μ_0 and ϵ_0

(a) To define the ampere, we put

$$\mu_0 = 4\pi \times 10^{-7} \text{ J s}^2 \text{ C}^{-2} \text{ m}^{-1}$$

(b) By experiment, we find

$$\epsilon_0 = 8.854 \times 10^{-12} \text{ J}^{-1} \text{ C}^2 \text{ m}^{-1}$$

So
$$\frac{1}{\mu_0 \epsilon_0} \approx 9 \times 10^{16} \text{ m}^2 \text{ s}^{-2}$$

and
$$\frac{1}{\sqrt{(\mu_0 \epsilon_0)}} \approx 3 \times 10^8 \text{ m s}^{-1}$$

More advanced analysis predicts that the speed c of electromagnetic radiation in free space is given by

$$c = \frac{1}{\sqrt{(\mu_0 \epsilon_0)}}$$

The above calculation suggests that this agrees with the experimental evidence, and in fact this relationship is used for the accurate measurement of ϵ_0 (p. 361).

53.9 SUMMARY OF THE MAGNETIC FIELD

Definitions		Laws	Important results
B	$F = BQv \sin \theta$		
m	$m = \frac{T_{\max}}{B}$		$F = BIl$
Φ	$\Delta\Phi = B \cos \theta \Delta A$		$T = NLAB \sin \theta$
			$T = mB \sin \theta$
μ_0	$B = \left(\frac{\mu_0}{4\pi}\right) \frac{Qv \sin \theta}{r^2}$	<i>Biot-Savart</i> $B \propto \frac{Qv \sin \theta}{r^2}$	$\Delta B = \left(\frac{\mu_0}{4\pi}\right) \frac{I \Delta l \sin \theta}{r^2}$
			coil $B = \frac{\mu_0 NI}{2r}$
			long solenoid $B = \mu_0 nI$
			straight wire $B = \frac{\mu_0 I}{2\pi a}$
			$F = \left(\frac{\mu_0}{2\pi}\right) \cdot \frac{I_1 I_2 \cdot l}{a}$
		<i>Ampère</i> $\oint B \, dl = \mu_0 I$	

54 ELECTRICAL MEASURING DEVICES

54.1	SURVEY OF INSTRUMENT TYPES	418
54.2	THE MOVING-COIL GALVANOMETER	418
54.3	WAYS OF ADAPTING THE GALVANOMETER	420
▶ 54.4	THEORY OF THE BALLISTIC GALVANOMETER	420

54.5	USES OF THE BALLISTIC GALVANOMETER	421
54.6	A SIMPLE CURRENT BALANCE	421
54.7	THE ELECTRODYNAMOMETER	422
54.8	THE WATTMETER	423

54.1 SURVEY OF INSTRUMENT TYPES

Those marked † are described in this chapter.

(A) Those that measure steady or alternating quantities

- (a) †Modified *moving-coil galvanometer*
- (b) †*Current balance*
- (c) †*Electrodynamometer*
- (d) *Cathode ray oscillograph* (p. 473)
- (e) *Electrostatic voltmeters*, such as
 - (i) the gold leaf electroscope, (p. 356)
 - (ii) the quadrant electrometer,
 - (iii) the attracted-disc electrometer. (p. 356)

They are not very sensitive instruments, and measure p.d.s greater than about 50 V.

- (f) *Diode voltmeter*
- (g) *Hot-wire ammeter*
- (h) *Thermocouple ammeter* (p. 443)
- (i) *Moving-iron instruments*, which can be made attractive or repulsive. They

- (i) have a non-uniform scale,
- (ii) are hardly affected by the Earth's field,
- (iii) need calibrating,
- (iv) are ballistic in nature, and are usually damped (e.g. by a dashpot mechanism). Many of these instruments give a deflection which is proportional to the *square* of an alternating quantity. Their readings are discussed further on p. 443.

(B) Those only suitable for steady quantities

- (a) †*Moving-coil galvanometer*
- (b) *Pivoted-magnet types*

A bar magnet, suitably pivoted in a restoring magnetic field and a deflecting field, gives a deflection that depends on the relative magnitudes of the fields.

This is the basis of the *tangent*, *sine*, *Helmholtz* and *astatic pair* galvanometers. They

- (i) can be made absolute if the Earth's field is known,
- (ii) are not very accurate,
- (iii) have non-linear scales,
- (iv) are of academic interest only.

The applications of the d.c. amplifier are discussed in 46.9.

54.2 THE MOVING-COIL GALVANOMETER

It is shown on page 407 that a coil of N turns each of area A and carrying a current I experiences a torque T given by

$$T = NIAB$$

when its plane is *parallel* to a field B .

Such a coil is the basis of the moving-coil galvanometer (fig. 54.1, opposite).

The coil system is usually either

- (a) pivoted on jewelled bearings for currents exceeding 5×10^{-5} A, or
- (b) suspended from a torsion fibre for currents less than 5×10^{-5} A.

Galvanometer sensitivity

Suppose the magnitude of the restoring torque T is given by $T = c\theta$ (p. 148) when the coil is deflected through angle θ [rad].

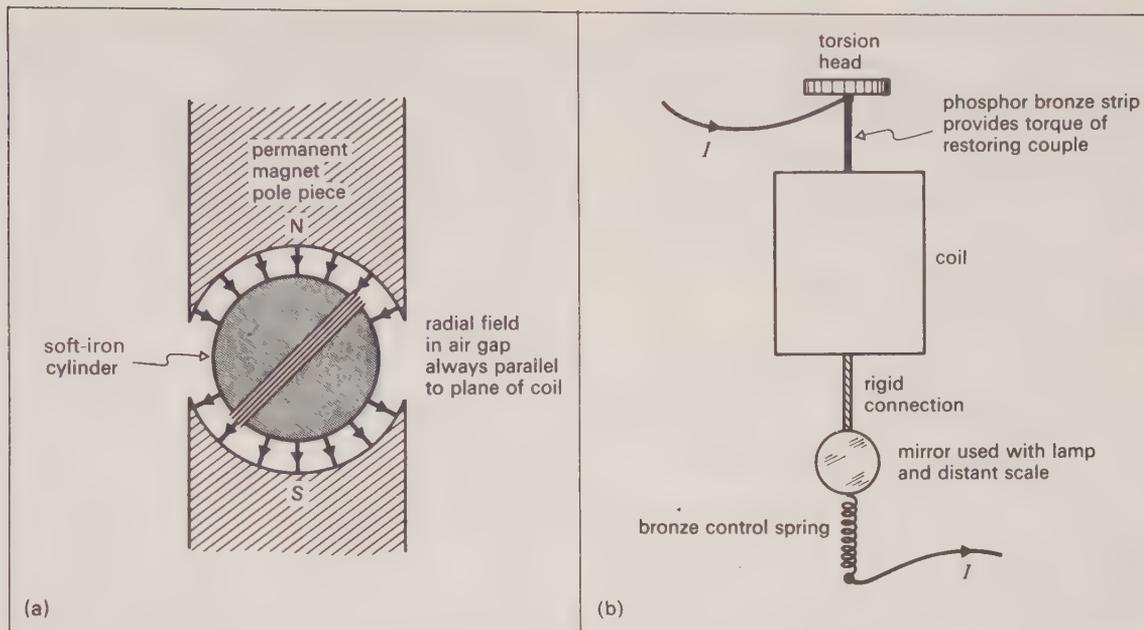


Fig. 54.1. A moving-coil galvanometer showing (a) the coil in the radial magnetic field (plan), (b) a suspension system (elevation).

$$T[\text{N m}] = c \left[\frac{\text{N m}}{\text{rad}} \right] \theta[\text{rad}]$$

When the coil is in equilibrium

$$NIAB = c\theta$$

$$\therefore \theta = \left(\frac{BAN}{c} \right) I$$

The **current sensitivity** S is defined by the equation

$$S_I = \frac{\theta}{I}$$

$$S_I \left[\frac{\text{rad}}{\text{A}} \right] = \frac{\theta[\text{rad}]}{I[\text{A}]}$$

Although the unit of S is rad A^{-1} , in practice currents as large as 1 A are not passed through galvanometers of this type.

From the definition,

$$S_I = \frac{BAN}{c}$$

so we see that *high sensitivity* can be achieved by using

(a) a magnet which gives a *large value of B* (but its size must remain constant).

(b) a coil of *large area A* wound with a *large number of turns N* (but its size must be manageable),

(c) a restoring system of *small c* (consistent with the long period of swing that this would produce).

The sensitivity is sometimes quoted as (I/θ) $[\text{A rad}^{-1}]$. Numerically this is the current required to give unit deflection. Its value might be $\sim 10^{-8}$ A for 1 mm deflection on a scale placed 1 m away.

High current sensitivity is desirable in circuits of high resistance.

Voltage sensitivity

Suppose that a p.d. V across the galvanometer produces a deflection θ . The **voltage sensitivity** is defined by the equation

$$S_V = \frac{\theta}{V} = \frac{\theta}{IR} = \frac{S_I}{R} = \frac{BAN}{cR}$$

High voltage sensitivity is desirable in circuits of relatively low resistance, such as those of the *Wheatstone bridge* and *potentiometer*.

Notes on the galvanometer

(a) $\theta \propto I$, so it cannot measure alternating values directly.

(b) The scale is linear when B is radial, so calibration is simple.

(c) External magnetic fields are relatively unimportant.

(d) It can be made very sensitive without loss of accuracy (but there is a lower limit imposed by *Brownian motion* (p. 135)).

(e) A suspension system is delicate, and significantly affected by excessive vibration.

(f) It is very widely used.

54.3 WAYS OF ADAPTING THE GALVANOMETER

(a) As an *ammeter*—add a **shunt** in parallel with a pivoted coil system (p. 394).

(b) As a *voltmeter*—add a high-resistance **bobbin** in series with the coil (p. 395).

(c) To measure *alternating quantities*

(i) Put a rectifier in series with the coil. (N.B. This destroys the a.c. by converting it to a half-wave pulsating d.c. (p. 457).)

(ii) To obtain full-wave rectification *through the instrument* use the rectifier bridge of fig. 54.2.

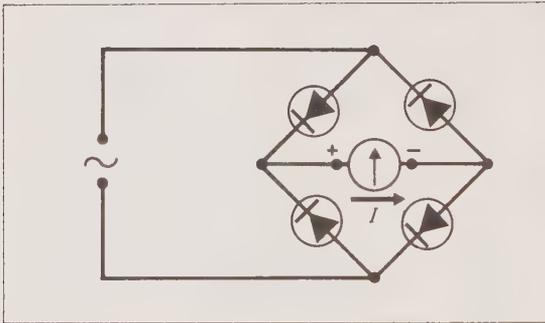


Fig. 54.2. Rectifier bridge for moving-coil galvanometer.

Dead-beat and ballistic use

(a) If the galvanometer is intended to attain a *steady value* quickly, damping is necessary to prevent oscillation. The galvanometer is **critically damped** (*dead-beat* movement) when oscillation is just prevented. Refer to fig. 11.4 (p. 93) for an illustrative diagram.

This is achieved by winding the coil on a *metallic frame*: the large induced currents in the frame give *electromagnetic damping*. (Refer to eddy currents, p. 430, for a discussion of the mechanism.)

(b) In a galvanometer adapted to measure charge, an impulsive torque sets the coil swinging. The damping is made as small as possible by placing a large resistance in series with the coil to reduce the induced current. The coil then swings freely (*ballistic* movement).

► 54.4 THEORY OF THE BALLISTIC GALVANOMETER

Principle

(1) Pass a charge Q through the coil in a time Δt which is $\ll \frac{1}{4}$ of the time period of swing.

(2) The suspended system receives an impulsive torque which is finished before the angular displacement is significant.

(3) The k.e. given to the coil is converted by rotation to p.e. stored in the system that provides the restoring torque.

Calculation

(a) The instantaneous torque

$$T = NIAB$$

The impulsive torque (angular impulse)

$$\begin{aligned} \int T dt &= \int_0^{\Delta t} NIAB dt \\ &= NAB \int_0^{\Delta t} I dt \\ &= NABQ \end{aligned}$$

Put K = moment of inertia of coil system

ω_0 = initial angular velocity

Then

angular impulse = change of angular momentum

$$NABQ = K\omega_0$$

(b) The initial k.e. of the coil = $\frac{1}{2}K\omega_0^2$.

Put θ_m = final deflection of coil

Then the final p.e. stored by the suspension fibre is

$$\int_0^{\theta_m} c\theta d\theta = \frac{1}{2}c\theta_m^2$$

Equating

$$\begin{aligned} \frac{1}{2}K\omega_0^2 &= \frac{1}{2}c\theta_m^2 \\ \frac{1}{2}K\left(\frac{NABQ}{K}\right)^2 &= \frac{1}{2}c\theta_m^2 \end{aligned}$$

So

$$Q = \left(\frac{\sqrt{cK}}{NAB}\right)\theta_m$$

$$Q \propto \theta_m$$

Notes

(a) (θ/Q) [rad C⁻¹] is called the **charge sensitivity**. It can be found by calculation or by calibration.

(b) Because neither air nor electromagnetic damping can be *eliminated*, the observed value of θ may have to be corrected to find the true θ_m by noting successive values of θ during the subsequent oscillation. (Again refer to fig. 11.4 on p. 93.)

(c) θ_m is frequently called the corrected **throw** of the galvanometer.

54.5 USES OF THE BALLISTIC GALVANOMETER

The previous paragraph is summarized by the equation

$$Q = k\theta_m$$

where θ_m is the corrected first throw, and k is a constant.

(Note that where values of θ are to be compared, no correction is necessary.)

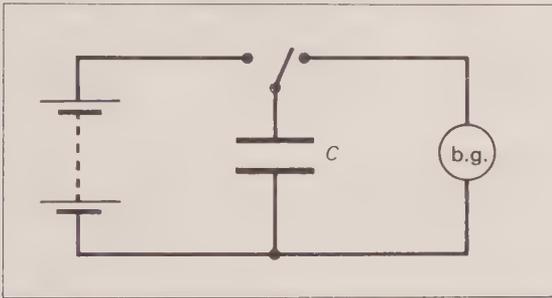


Fig. 54.3. The ballistic galvanometer on open circuit.

(A) Use on open circuit

Fig. 54.3 shows a circuit for comparing the capacitances of two capacitors.

Connect two capacitors in turn, (1) to a battery which charges them to a p.d. V , and (2) to the ballistic galvanometer.

Then $Q_1 = C_1V = k\theta_{1,m}$

$Q_2 = C_2V = k\theta_{2,m}$

Dividing $\frac{C_1}{C_2} = \frac{\theta_{1,m}}{\theta_{2,m}}$

We do not need to know the e.m.f. of the battery.

(B) Calibration on closed circuit

When the b.g. is used on *closed* circuit, the damping is altered significantly. It must be calibrated in the circuit in which it is to be used (fig. 54.4).

To find k in $Q = k\theta_m$, a known Q is passed through the instrument by using the *standard mutual inductance* (p. 440). (A capacitor cannot be used because it would cause the circuit to revert to an open one.)

Apart from the comparison of capacitance (left), uses of the b.g. include

(a) measurement of flux changes $\Delta\Phi$ from

$$Q = \frac{\Delta\Phi}{R} \quad (\text{p. 428})$$

(b) measurements on the *Earth's magnetic field* using the *Earth inductor*, and

(c) the plotting of *hysteresis curves*. (p. 462)

54.6 A SIMPLE CURRENT BALANCE

A **current balance** is a device which measures the current in terms of the forces that current-carrying conductors exert on each other. It is an **absolute instrument** in the sense that the current is measured in terms of *mechanical* quantities, and is *directly related to the definition of the ampere*. No electrical measurements are needed, nor prior knowledge of any electrical quantity except the defined value of μ_0 .

There are several versions: fig. 54.5 (overleaf) shows a simple one designed for the student laboratory. Note that the solenoid is in series with, and therefore carries the same current as, the wire.

If the beam is balanced through its centre of gravity

$$mg = BIl$$

in which $B = \mu_0 nI$ if we neglect corrections for the solenoid's finite length.

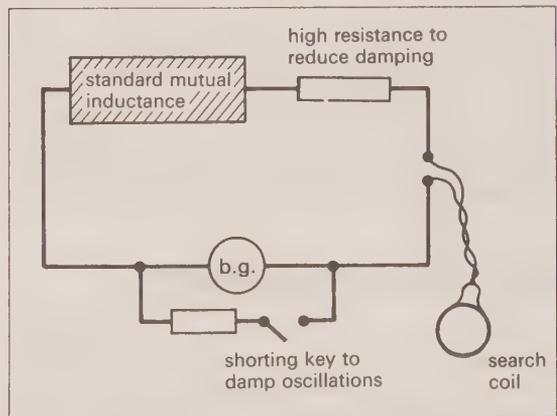


Fig. 54.4. Calibration of a ballistic galvanometer on closed circuit.

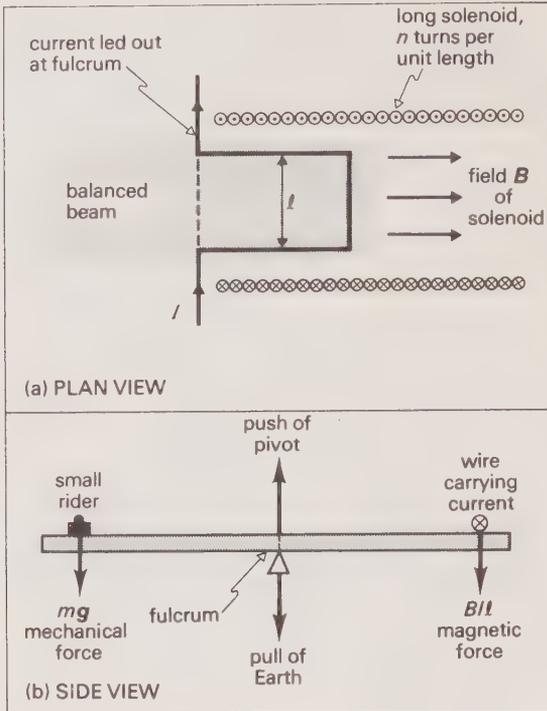


Fig. 54.5. A simple student-laboratory form of current balance.

$$\begin{aligned} \therefore mg &= (\mu_0 n I) l \\ &= (\mu_0 n l) I^2 \end{aligned}$$

Thus I can be calculated from a knowledge of μ_0 and the mechanical quantities m , g , n and l .

Notes

(a) The current balance represents a practical way of realizing the situation from which the ampere is defined (p. 9).

(b) The balance is primarily used to standardize other more convenient current-measuring devices.

(c) Since $mg \propto I^2$, the balance can be used to register the r.m.s. value of an alternating current.

54.7 THE ELECTRODYNAMOMETER

This too, like the current balance, is an absolute instrument. It consists of a small pivoted coil carrying a current which experiences a torque when placed in the magnetic field at the centre of a second larger outer coil carrying the same current (fig. 54.6).

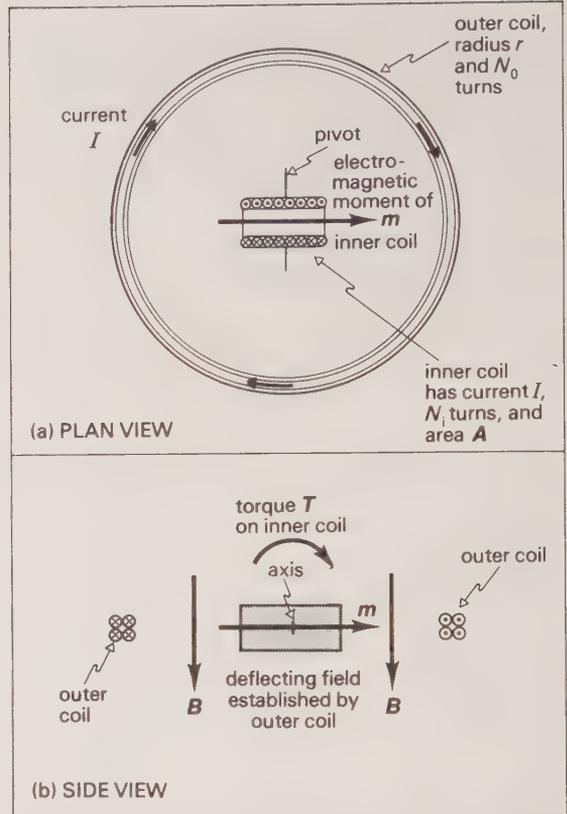


Fig. 54.6. Schematic arrangement of an electro-dynamometer.

The principle is that of the moving-coil galvanometer, but the device is less sensitive because the deflecting field is much smaller than that established by a permanent magnet.

The deflecting torque on the inner coil is given by

$$T = mB \tag{p. 407}$$

in which $m = N_1 I A$ (p. 407)

and $B = \frac{\mu_0 N_0 I}{2r}$ (p. 412)

giving $T = \frac{\mu_0 N_1 N_0 A I^2}{2r}$

The sense of the torque is such as to align the electromagnetic moment m in the same direction as the deflecting field B . Equilibrium can be attained by exerting an equal but opposite elastic torque, or by attaching the inner coil to a pivoted beam which can be loaded.

The notes of 54.6 apply equally here, and again I can be calculated from a knowledge of μ_0 and purely mechanical measurements.

54.8 THE WATTMETER

Principle

The rate of conversion P of electrical energy by the load X in fig. 54.7 is given by

$$P = VI$$

The **wattmeter** combines the function of voltmeter and ammeter to record the *power* directly.

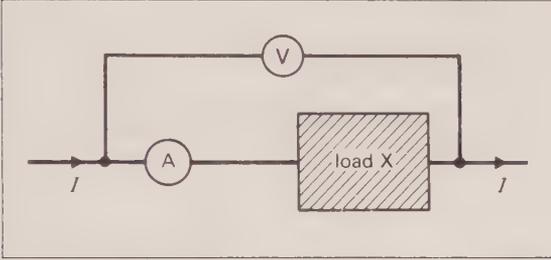


Fig. 54.7. The principle of the wattmeter.

Adaptation of the dynamometer

In fig. 54.8

(a) the *fixed coils* carry the same current I as the load X , and

(b) the *moving coil* carries a current which is proportional to the *potential difference* V across the load.

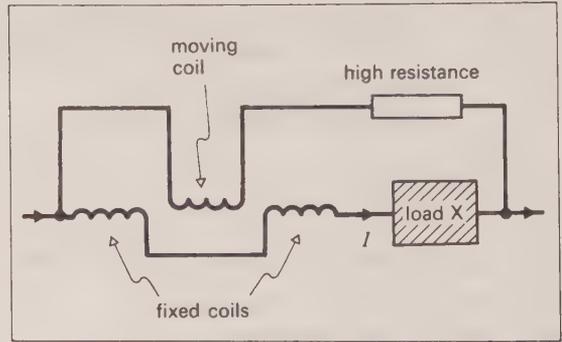


Fig. 54.8. The dynamometer as a wattmeter.

Deflecting couple

$$\propto (\text{fixed coil current}) \times (\text{moving coil current})$$

$$\propto VI$$

$$\propto P, \text{ the instantaneous power supply.}$$

$$\therefore \left(\begin{array}{c} \text{mean} \\ \text{deflection} \end{array} \right) \propto \left(\begin{array}{c} \text{mean} \\ \text{power supply} \end{array} \right)$$

The device can be used to measure the rate of conversion of electrical energy by a device carrying a.c. when V and I may be *out of phase* (p. 452).

The **joulemeter** (or watt-hour meter) incorporates a revolution counter to record *energy* consumption.

55 PRINCIPLES OF ELECTROMAGNETIC INDUCTION

55.1	ELECTROMAGNETIC INDUCTION	424
55.2	MAGNETIC FLUX-LINKAGE $N\Phi$	424
55.3	CALCULATION OF THE MAGNITUDE OF THE INDUCED E.M.F.	424
55.4	THE LAWS OF ELECTROMAGNETIC INDUCTION	426
55.5	FURTHER EXAMPLES OF ELECTROMAGNETIC INDUCTION	427

55.6	THE E.M.F. INDUCED IN A ROTATING COIL	428
55.7	THE E.M.F. INDUCED BY A CHANGING FIELD	428
55.8	THE MAGNITUDE OF AN INDUCED CHARGE	428

55.1 ELECTROMAGNETIC INDUCTION

Fig. 55.1(a) (opposite) shows a conductor being pulled by an external agent at velocity v through a magnetic field B directed into the paper.

Moving charge-carriers constitute a current: in a magnetic field they experience a (magnetic) force. By *Fleming's left-hand motor rule* electrons experience a force *to the left*: because they are *free* they move to the left. (The positive ions of the lattice also experience the force, but are not free to move.)

The effects of their movement are shown in fig. 55.1(b).

If one thinks of the conductor as a battery, the electric field *within it* (from positive to negative terminal) is in a direction opposite to that of the conventional e.m.f.

Notes

(a) The electric field in the conductor is the *result* of the electron accumulation, and *not the cause* of the electron movement.

(b) The charge accumulation stops either when

(i) the relative movement stops, or
 (ii) when the (electrical) forces that result from the charge accumulation are equal in magnitude but opposite in direction to the (magnetic) forces that result from the movement of the conductor.

(c) The e.m.f. results in a p.d. across the ends of the conductor so long as the movement is maintained. It is called an **induced e.m.f.**

(d) If the conductor is moved in the opposite direction, then the polarity of the e.m.f. is also reversed.

55.2 MAGNETIC FLUX-LINKAGE $N\Phi$

The magnetic flux passing through a loop or coil is said to **thread** or **link** the coil.

Suppose a coil has N turns, and that the flux linking *each turn* is Φ . Then we define the **flux-linkage** to be $N\Phi$.

No particular symbol is used in this book.

The unit is the **weber-turn**, but $N\Phi$ has the same dimensions as the flux Φ .

55.3 CALCULATION OF THE MAGNITUDE OF THE INDUCED E.M.F.

Fig. 55.2 shows an external agent exerting a force F so as to pull the rectangular coil at a *steady* velocity v through the field B .

Suppose the e.m.f. \mathcal{E} causes a current I . The (magnetic) force on the wire, by the motor rule,

$$F = BIl \quad \text{to the left}$$

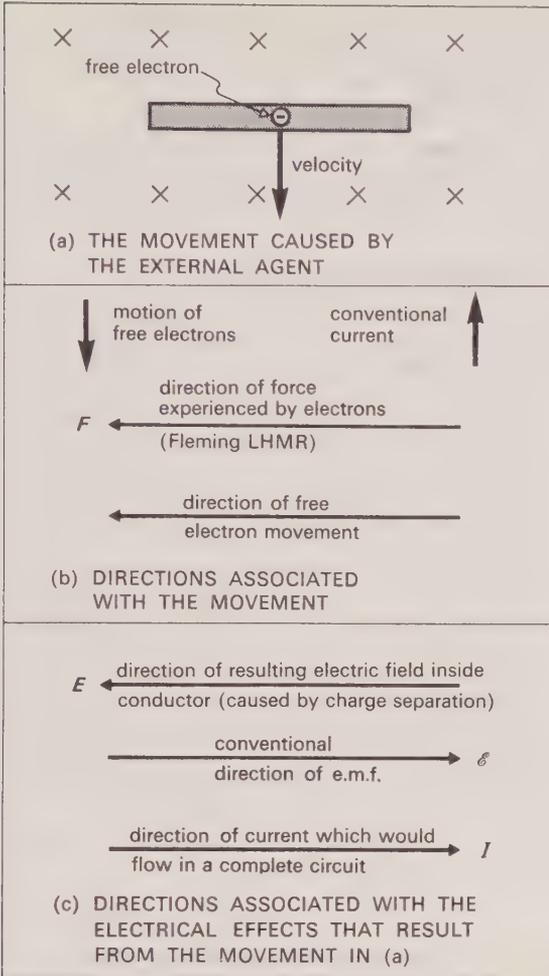


Fig. 55.1. Generation of a motional e.m.f.

The wire is in equilibrium (since its velocity is constant), and so

$$F = BIl \text{ to the right}$$

The rate of working of the agent = Fv

The rate of working of the seat of e.m.f. = $\mathcal{E}I$

By the law of conservation of energy, if no other source (such as the cause of B) provides energy, then these are equal.

$$\mathcal{E}I = (BIl)v$$

$$\mathcal{E} = Blv$$

Notes

(a) If side CD had also been in the field, it would have been the seat of an equal e.m.f., which would have acted anticlockwise in the loop. The induced current would have been zero.

(b) The external agent does work, which is (i) converted to electrical energy, but then (ii) dissipated to internal energy by Joule heating.

The net effect is to increase the internal energy of the wire.

A force calculation of the e.m.f.

Consider the equilibrium situation shown in fig. 55.3. The forces on the free charge $-Q$ are

(a) a magnetic force to the left, of size BQv ,

(b) an electric force to the right, of size EQ .

Putting $E = (\mathcal{E}/l)$ (p. 351) and equating, we have

$$(\mathcal{E}/l)Q = BQv$$

$$\mathcal{E} = Blv$$

This should be compared to the theory of the Hall effect (p. 403).

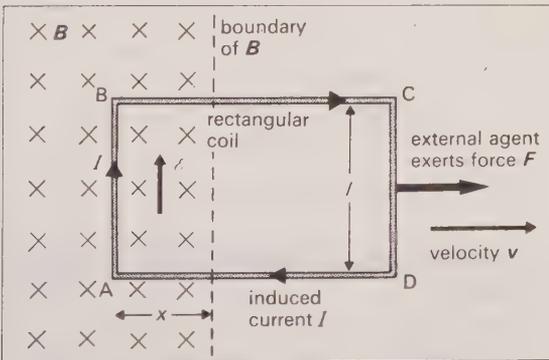


Fig. 55.2. Calculation of the motional e.m.f.

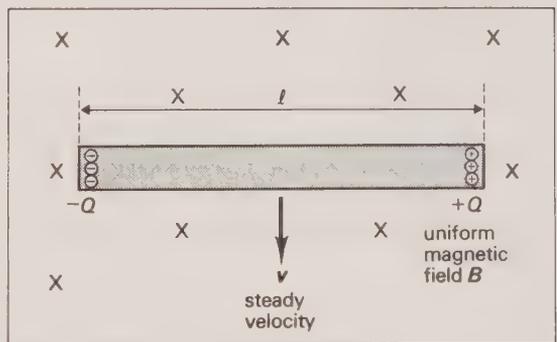


Fig. 55.3. Alternative calculation of the motional e.m.f.

Ways of expressing the induced e.m.f.

(a) The side AB in fig. 55.2 is of length l . In time Δt it sweeps out an area $lv \Delta t$, through which there is a flux density B . The flux $\Delta\Phi$ cut by the conductor in time Δt is given by

$$\Delta\Phi = Blv \Delta t$$

so
$$\mathcal{E} = Blv = \frac{\Delta\Phi}{\Delta t}$$

The induced e.m.f. is proportional to the rate at which the conductor cuts magnetic field lines.

(b) The flux through the coil of fig. 55.2 is given by $\Phi = Blx$. Since there is only one turn, the flux-linkage is also Blx . Then

$$\begin{aligned} \frac{d\Phi}{dt} &= Bl \frac{dx}{dt} \quad (\text{since } B \text{ and } l \text{ are constant}) \\ &= Blv = \mathcal{E} \end{aligned}$$

The induced e.m.f. is equal to the rate of change of flux-linkage.

This result is of more general use, since it can be used when the coil is stationary and the flux-linkage made to change by altering the magnitude of B .

If there are N turns, then the total e.m.f.

$$\mathcal{E} = N \cdot \frac{d\Phi}{dt}$$

55.4 THE LAWS OF ELECTROMAGNETIC INDUCTION**Faraday's First Law:**

When the magnetic flux threading a circuit is changing, an e.m.f. is induced in the circuit.

The magnitude comes from

Faraday's Second Law:

The magnitude of an induced e.m.f. is proportional to the rate of change of flux-linkage.

This law is sometimes expressed in Neumann's Equation

$$\mathcal{E} = -N \cdot \frac{d\Phi}{dt}$$

$$\mathcal{E} [\text{volts}] = -N \cdot \frac{d\Phi [\text{Wb}]}{dt [\text{s}]}$$

On the left $1 \text{ V} = 1 \text{ J C}^{-1}$

On the right $\frac{1 \text{ Wb}}{\text{s}} = \frac{1 \text{ N m}}{\text{A s}} = 1 \text{ J C}^{-1}$

The minus sign is discussed below.

The direction of the induced e.m.f.

This must be such that the law of conservation of energy is obeyed. **Le Chatelier's Principle** is a general statement:

When a constraint is applied to a physical system in equilibrium then, so far as it can, the system will adjust itself so as to oppose the constraint.

For electromagnetic induction this is expressed by

(1) Lenz's Law

The direction of an induced current (if one were to flow) is such that its effect would oppose the change in magnetic flux which gave rise to the current.

This is achieved by exerting forces of some kind, or establishing a magnetic field.

Or (2) Fleming's right-hand dynamo rule

This rule, which is particularly useful for straight conductors, is shown in fig. 55.4.

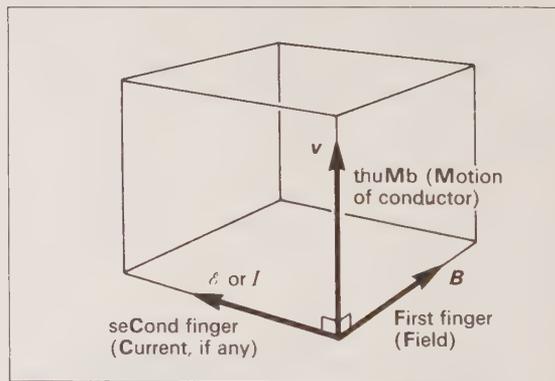


Fig. 55.4. Fleming's right-hand dynamo rule.

► The electric field associated with an induced e.m.f.

Fig. 55.5 shows the analogous situations in which

(a) a moving charge (current) establishes a magnetic field, whose sense and direction are

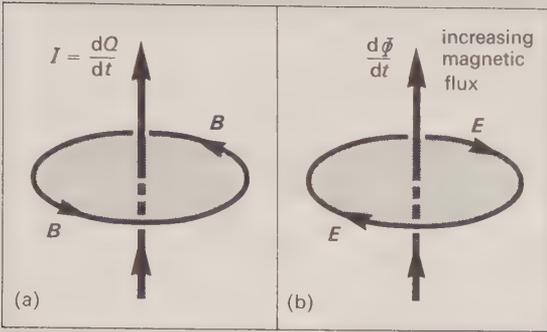


Fig. 55.5. (a) Right-handed rule, and (b) left-handed rule.

given by *Maxwell's right-handed corkscrew rule*, and

(b) a time-varying magnetic flux gives rise to an electric field, whose sense and direction are given by a *left-handed corkscrew rule*.

The *left-handedness* of the field E explains why *Neumann's equation* is given a *minus sign*.

The lines of E in fig. 55.5(b) are continuous: the electric field caused by the changing magnetic flux is *non-conservative*. (Compare with that established by static electric charges, p. 340.)

55.5 FURTHER EXAMPLES OF ELECTROMAGNETIC INDUCTION

In this section we give examples of the conversion of *mechanical energy into electrical energy* ((a) to (d)) or of the conversion of energy stored in a magnetic field into electrical energy (e).

In each example either

- (i) the flux-linkage of a loop is changing, or
- (ii) a conductor cuts across a magnetic field,

and this results in an electric field of the type discussed in section 55.4.

(a) If the magnet of fig. 55.6 is withdrawn, the sense of the induced current in the coil is reversed to preserve the *status quo*. The induced current is zero when their relative speed is zero.

Lenz's Law indicates that approach produces repulsion, and withdrawal produces attraction.

(b) The experiment of fig. 55.6 can be repeated using a loop carrying a current instead of a permanent magnetic dipole. The results are the same.

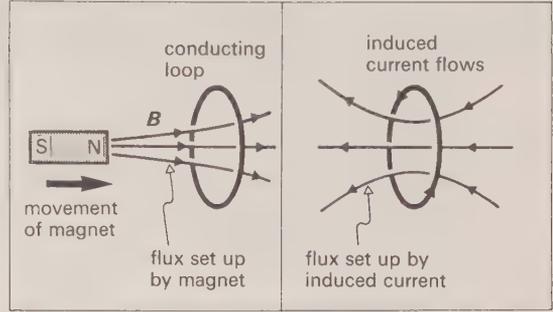


Fig. 55.6. Relative movement between a magnet and a conducting loop.

(c) Rotation of a loop in a magnetic field sets up an induced e.m.f. (and current).

This is the principle of the generator (p. 433). The conducting material of the loop cuts the magnetic field lines.

(d) If a loop is deformed in a magnetic field so that its area is changed, there is a change of flux-linkage, and a consequent e.m.f. An induced current causes forces that oppose further change of area.

(e) In fig. 55.7 the induced current is in the same sense as the original current in the first coil. The decay of magnetic flux is delayed by the temporary current flow in the second coil. The energy to produce this current comes from the energy stored in the magnetic field. This effect (**mutual induction**) and a similar effect (**self-induction**) are discussed in chapter 57.

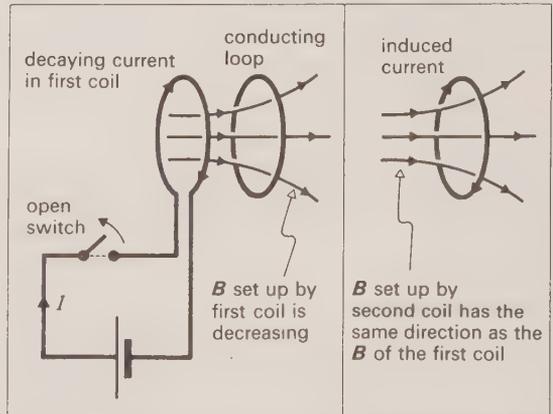


Fig. 55.7. A change of current in one loop induces a current in the other.

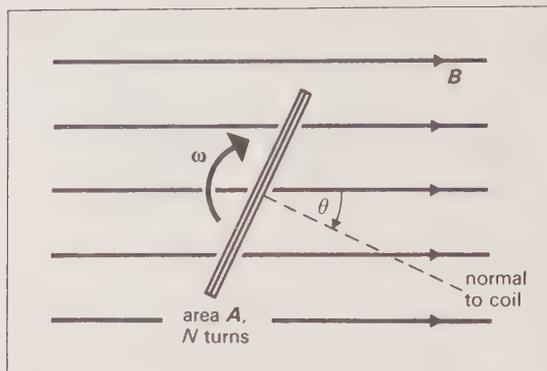


Fig. 55.8. Production of a sinusoidal e.m.f.

55.6 THE E.M.F. INDUCED IN A ROTATING COIL

Consider a conducting coil of N turns each of area A being made to rotate with angular velocity ω in a uniform magnetic field B (fig. 55.8).

The flux through each turn is given by

$$\Phi = BA \cos \theta \quad (\theta = \omega t)$$

The induced e.m.f. for each turn is

$$-\frac{d\Phi}{dt} = \omega AB \sin \omega t$$

For N turns

$$\begin{aligned} \mathcal{E} &= N\omega AB \sin \omega t \\ &= \mathcal{E}_0 \sin \omega t \end{aligned}$$

where the maximum value of \mathcal{E} is $\mathcal{E}_0 = N\omega AB$, since $\sin \omega t$ cannot exceed 1.

The output from a simple generator is a **sinusoidally varying e.m.f.**

When the *plane* of the coil is

- (a) perpendicular to B , $\mathcal{E} = 0$
- (b) parallel to B , $\mathcal{E} = \mathcal{E}_0$

\mathcal{E}_0 can be made larger by increasing N , ω , A and B . The behaviour of a.c. is discussed in chapter 58, p. 443. The dynamo (generator) is discussed on p. 433.

Motor effect

In the generator a coil carrying a current is made to rotate in a magnetic field, and so experiences a torque which (by *Lenz's Law*) opposes the rotation. To produce electrical energy the external agent must expend mechanical energy (do work) against this opposing torque.

55.7 THE E.M.F. INDUCED BY A CHANGING FIELD

Suppose a coil of area A and N turns is placed normal to a changing magnetic field B . For each turn $\Phi = BA$, so the e.m.f. induced across the coil is given by

$$\mathcal{E} = NA \left(\frac{dB}{dt} \right)$$

The a.c. search coil

The e.m.f. induced gives useful information about the field. If the field is established by a sinusoidal a.c. we can write

$$B = B_0 \sin \omega t$$

so

$$dB/dt = \omega B_0 \cos \omega t$$

$$\therefore \mathcal{E}_{\max} \propto \omega B_0$$

The size of the induced e.m.f. can be measured by a c.r.o., and is proportional to the peak value of the field at that point. It is also proportional to the frequency (p. 433).

55.8 THE MAGNITUDE OF AN INDUCED CHARGE

Fig. 55.9 shows a search coil of N turns and area A connected into a circuit of resistance R .

Suppose the coil is totally immersed in the field B of the magnet, and then suddenly snatched away.

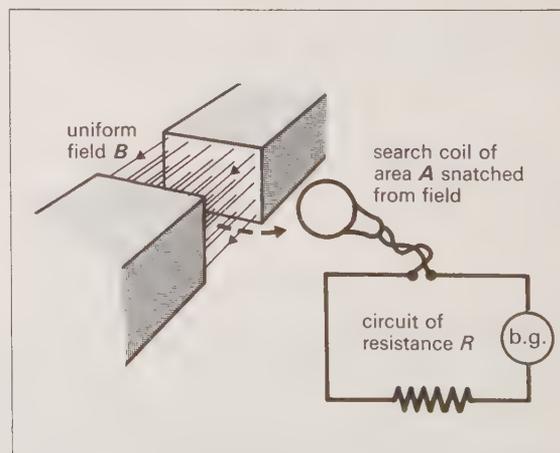


Fig. 55.9. Search coil.

Instantaneously $I = \frac{\mathcal{E}}{R}$

$$= \frac{1}{R} \left[\frac{d}{dt} (N\Phi) \right] \quad (\Phi = BA)$$

$$= \frac{NB}{R} \left(\frac{dA}{dt} \right)$$

(As we are not concerned here with the *sense* of the current, the minus sign has been dropped.)

As the area immersed in the field changes, so current flows in the circuit.

It is seen from fig. 55.10 that the *charge* passing any cross-section of the circuit is *independent of the timing* Δt , being given by

$$Q = \int_0^{\Delta t} I dt$$

$$= \frac{NB}{R} \int \frac{dA}{dt} dt$$

$$= \frac{NBA}{R}$$

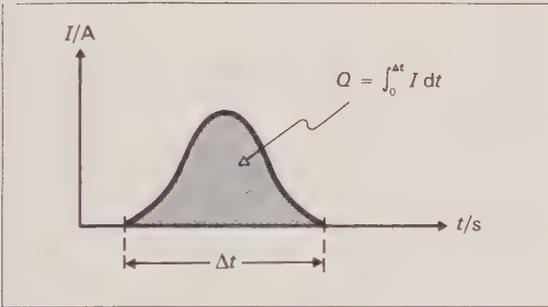


Fig. 55.10. Calculation of induced charge.

If the coil remains in the field, but is rotated through an angle π rad (half a revolution), then

$$Q = \frac{2NBA}{R}$$

More generally

$$Q = \frac{\text{change of flux-linkage}}{R}$$

$$Q[\text{C}] = \frac{N \Delta\Phi[\text{Wb}]}{R[\Omega]}$$

Applications

(a) A search coil used in this way with a ballistic galvanometer can evaluate Q , $\Delta\Phi$ and hence B , the flux density of a magnetic field. (See, for example, p. 421, and the *Rowland Ring* on p. 459.)

(b) The **Earth inductor** consists of a special search coil of a few hundred turns (and each of large area) mounted on a special wooden frame. It can measure the relatively weak magnetic field of the Earth.

If one measures the vertical and horizontal components B_{ver} and B_{hor} separately, one can calculate the **angle of dip** D from

$$\tan D = \frac{B_{\text{ver}}}{B_{\text{hor}}}$$

56 APPLICATIONS OF ELECTROMAGNETIC INDUCTION

56.1	EDDY CURRENTS	430
56.2	THE TRANSFORMER	431
56.3	ABSOLUTE MEASUREMENT OF RESISTANCE (LORENZ)	432
▶ 56.4	ABSOLUTE MEASUREMENT OF RESISTANCE (BELHAM)	433

56.5	THE DYNAMO	433
56.6	THE MOTOR AND BACK E.M.F.	434
56.7	THE ALTERNATOR	435

56.1 EDDY CURRENTS

Any block of conducting material either

- (i) moving in a magnetic field, or
- (ii) placed in a changing magnetic field

experiences induced currents, called **eddy currents**. They flow in whatever paths are dictated by the need to *oppose the change* that produced them (fig. 56.1). They show both heating and magnetic effects.

(A) A heating effect

(a) This can be useful, as in the **induction furnace**. *Joule heating* causes the melting of a conductor placed in the changing field, but insulators remain unaffected.

(b) This can be a nuisance in *iron-cored apparatus*, such as motors, dynamos and transformers. The *Joule heating* ($\propto V^2/R$) is reduced by increasing the resistance of the eddy current path by

- (i) *laminating* the core,
- (ii) using a bundle of iron wires, or
- (iii) using a core packed with iron dust.

At extremely high frequencies one must use **ferrites**, materials of high resistivity whose magnetic properties are in some respects similar to those of iron (p. 461).

(B) A magnetic effect

A moving conductor will experience forces caused by the interaction of the external magnetic field

with that set up by its own eddy currents. These will result in a *braking effect*.

Examples

(a) A copper plate swinging between the poles of a magnet is strongly damped unless it is slotted in the right direction.

(b) *Galvanometer damping* (see also p. 420).

(i) In the *dead-beat* movement the circuit resistance is just low enough to prevent oscillation (so that the induced current is relatively large), but not so low as to give over-damping.

(ii) In the *ballistic* movement the induced current is as small as possible so that the coil swings freely.

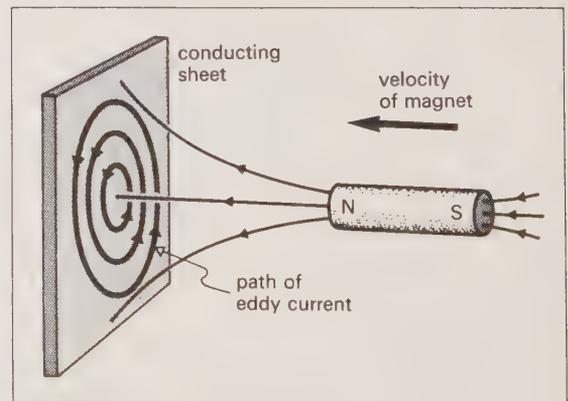


Fig. 56.1. Illustrating the direction of a typical eddy current.

56.2 THE TRANSFORMER

Suppose fig. 56.2 represents an **ideal transformer** in which

- (a) the primary resistance is zero,
- (b) there is no flux leakage, so that we have the same flux through each turn of both the primary and secondary coils, and
- (c) the secondary is on open circuit.

Suppose the current change in the primary changes the flux through the core at the rate $d\Phi/dt$.

Then the back e.m.f. in the primary

$$= -N_p \frac{d\Phi}{dt}$$

and the induced e.m.f. in the secondary

$$\mathcal{E}_s = -N_s \frac{d\Phi}{dt}$$

Applying the loop equation to the primary

$$\mathcal{E}_p + \left(-N_p \frac{d\Phi}{dt}\right) = IR = 0 \quad (R = 0)$$

$$\therefore \mathcal{E}_p = N_p \frac{d\Phi}{dt}$$

from which
$$\frac{\mathcal{E}_p}{\mathcal{E}_s} = -\frac{N_p}{N_s}$$

The minus sign indicates that the two e.m.f.s are in antiphase.

The transformer *steps-up* the applied p.d. if $\mathcal{E}_s > \mathcal{E}_p$.

Secondary not on open circuit

Suppose a *load resistance* R is connected across the secondary, so that the secondary current is I_s , and the primary current becomes I_p . If there is no energy loss

power input = power output

$$\mathcal{E}_p I_p = \mathcal{E}_s I_s$$

So

$$\frac{I_p}{I_s} = \frac{\mathcal{E}_s}{\mathcal{E}_p} = \frac{N_s}{N_p}$$

In a step-up transformer $I_s < I_p$.

In practice a *non-ideal* transformer shows these losses:

- (a) I^2R losses (*Joule heating*) in the copper windings,
- (b) I^2R losses due to eddy currents in the iron core,
- (c) hysteresis losses (the core dissipates energy on repeated magnetization, p. 461), and
- (d) loss due to magnetic leakage.

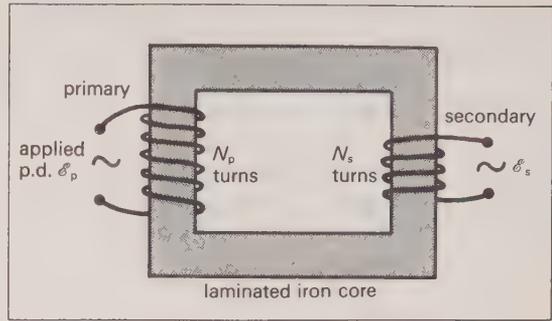


Fig. 56.2. An ideal transformer.

► Reflected resistance

Using the zero-loss equation

$$\frac{I_p}{I_s} = \frac{\mathcal{E}_s}{\mathcal{E}_p} = \frac{N_s}{N_p}$$

and

$$I_s = \frac{\mathcal{E}_s}{R}$$

We have

$$I_p = \left(\frac{N_s}{N_p}\right) \cdot \left(\frac{\mathcal{E}_s}{R}\right) = \frac{\mathcal{E}_p}{R \left(\frac{N_p}{N_s}\right)^2}$$

A resistive load R across the secondary acts in the primary as though it were a resistance $R(N_p/N_s)^2$ which is called a **reflected resistance**.

Demonstration

When in fig. 56.3

- (a) the switch is open, the lamp remains unlit;
- (b) the switch is closed, the resistance in the

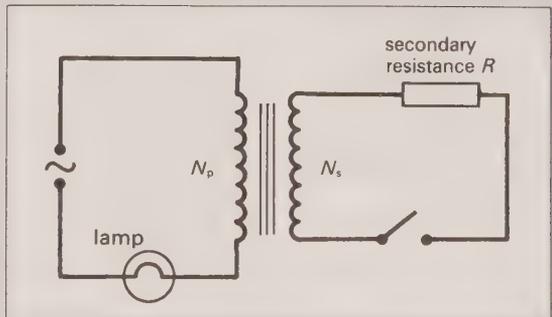


Fig. 56.3. Controlling the primary current by use of the secondary resistance.

secondary *increases* the current through the *primary*, and this causes the lamp to shine.

Impedance matching

The previous result is important when a transformer is used for **impedance matching**. A circuit of output impedance Z_p can be *matched* to a second circuit of input impedance Z_s through a transformer if

$$\frac{Z_p}{Z_s} = \left(\frac{N_p}{N_s}\right)^2$$

A step-up transformer would therefore be used to *match* a low impedance at its input to a high impedance at its output. This impedance matching ensures maximum power transfer from the generator (primary circuit) to the load (secondary circuit).

56.3 ABSOLUTE MEASUREMENT OF RESISTANCE (LORENZ)

This method constitutes a determination of resistance without the need to make any *electrical* measurements. Because no knowledge of V or I is required, this is an **absolute** method.

Principle

Fig. 56.4 shows that there exists an e.m.f. (and hence a p.d.) between O and C when the copper disc is rotated in the magnetic field.

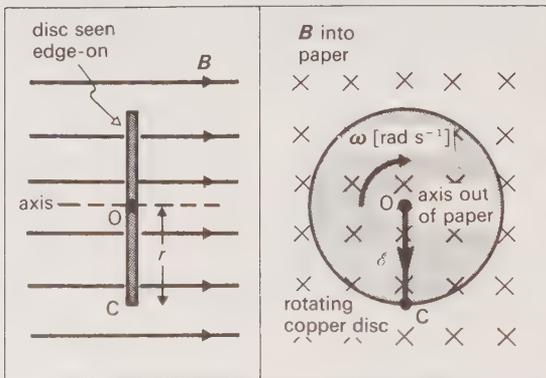


Fig. 56.4. There exists a p.d. between the centre and rim of the rotating disc.

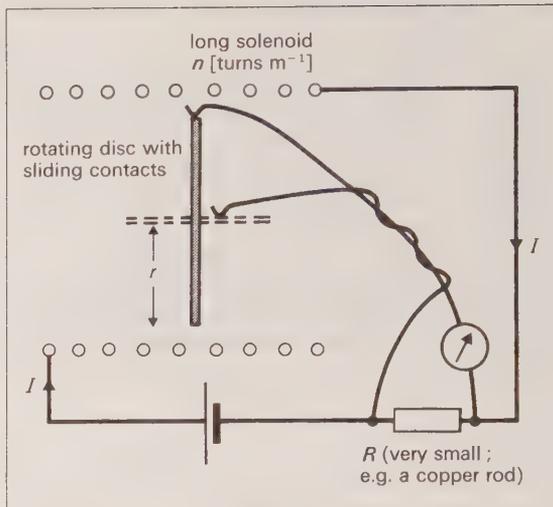


Fig. 56.5 The Lorenz experiment.

Experiment

In fig. 56.5 the induced e.m.f. (which at balance equals the p.d. between the sliding contacts) is changed by altering the angular speed ω of the disc until the galvanoscope shows no deflection. Inside the long solenoid

$$B = \mu_0 n I$$

Any radius of the disc cuts flux at the rate BA in each revolution

$$\therefore \frac{d\Phi}{dt} = \left(\frac{\omega}{2\pi}\right) BA$$

Thus the p.d. across the brushes

$$V_{oc} = \left(\frac{\omega}{2\pi}\right) (\mu_0 n I) (\pi r^2)$$

When ω is correctly adjusted,

$$V_{oc} = IR$$

Equating
$$R = \left(\frac{\mu_0}{2}\right) \omega n r^2$$

since the current has cancelled.

We need to measure r^2 [m²], n [m⁻¹] and ω [rad s⁻¹] so all readings are in terms of *mechanical* quantities.

Precautions

The induced e.m.f. is very small, so

- (a) the Earth's field must be taken into account,
- (b) contact p.d.s must be avoided,
- (c) allowance must be made for thermoelectric

e.m.f.s generated at the sliding contacts by frictional heating, and
 (d) a very sensitive galvanoscope must be used.

► 56.4 ABSOLUTE MEASUREMENT OF RESISTANCE (BELHAM)

Refer to fig. 56.6. Suppose a current $I_0 \sin \omega t$ flows through the long solenoid S and the resistor R . Let A be the solenoid's area of cross-section, n its number of turns per unit length, and N the number of turns on the search coil L . Inside the solenoid

$$B = \mu_0 n I_0 \sin \omega t$$

$$\frac{dB}{dt} = \mu_0 n I_0 \omega \cos \omega t$$

The maximum e.m.f. induced across L

$$\mathcal{E}_{\max} = NA\mu_0 n I_0 \omega$$

and is proportional to ω . If R is non-inductive the maximum p.d. across it is

$$V_{\max} = I_0 R$$

and is independent of ω . The experiment consists of adjusting ω until the detector shows that $\mathcal{E}_{\max} = V_{\max}$.

Then
$$I_0 R = NA\mu_0 n I_0 \omega$$

giving
$$R = NA\mu_0 n \omega$$

Notes

(a) The method is absolute because the current amplitude cancels, and so need not be known.

(b) This is a more accurate experiment in the school laboratory than the *Lorenz* method, since the e.m.f. induced across L can easily be made as much as 1 volt. This enables R to be about 1 ohm, as compared to a few $\mu\Omega$ in the *Lorenz* method.

56.5 THE DYNAMO

The theory of the dynamo is given on p. 428. The induced e.m.f. causes a current to flow in the external circuit to which the dynamo is connected. This current is led away by **slip-rings** which rotate with the coil.

In the *d.c. dynamo* the current is made unidirectional by using a **split-ring commutator**: the output is then a unidirectional pulsating d.c., as shown in fig. 58.1 on p. 442.

Dynamos can be

(a) **shunt-wound**, in which case the **field coils** (which energize the electromagnet) are in parallel with the **armature**, or

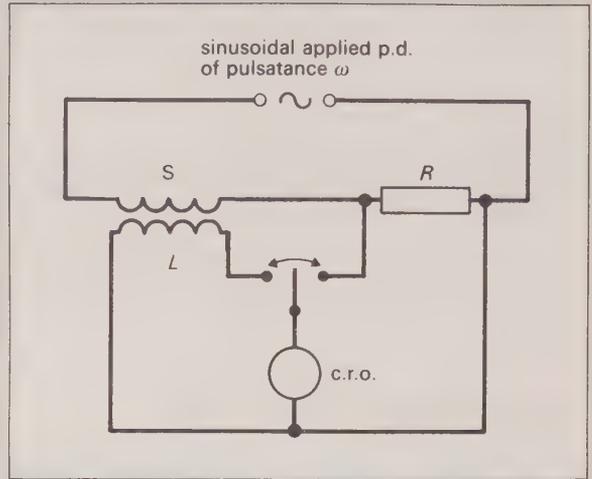


Fig. 56.6. The circuit of Belham's experiment.

(b) **series-wound**, in which case they are connected in series with the armature.

In both types of winding the output p.d. depends on the current taken from the dynamo (the **load**). It can be made nearly constant if the dynamo is **compound-wound**.

The methods of winding are illustrated in fig. 56.7.

In the dynamo, the induced e.m.f. \mathcal{E} is in the same direction as the current I that it drives through the armature and round the circuit. If V is

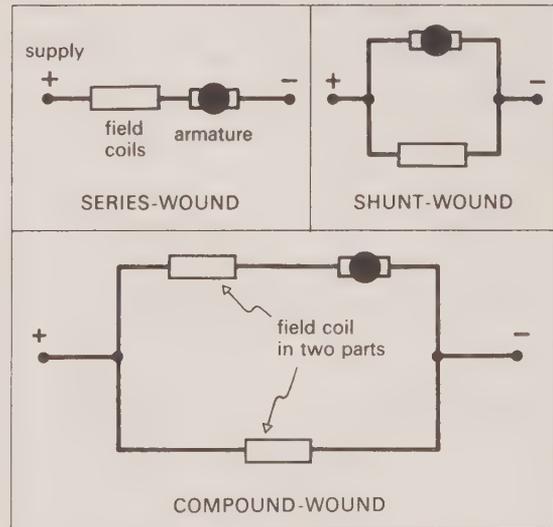


Fig. 56.7. Different ways of winding the field coils and armature.

the p.d. across the armature and R is its *internal resistance*, the loop equation (p. 390) gives

$$\mathcal{E} - V = IR$$

$$\therefore \mathcal{E}I = I^2R + VI$$

$$\left(\begin{array}{c} \text{electrical} \\ \text{power} \\ \text{generated} \end{array} \right) = \left(\begin{array}{c} \text{power dissipated} \\ \text{by Joule heating} \\ \text{in the armature} \end{array} \right) + \left(\begin{array}{c} \text{electrical} \\ \text{power} \\ \text{output} \end{array} \right)$$

The electrical power generated is equal to the *mechanical power input* required to maintain the rotation against the opposing electromagnetic torque.

56.6 THE MOTOR AND BACK E.M.F.

The motor is a device which converts electrical energy to mechanical energy. The principle is the same as that of the moving-coil galvanometer (p. 406), with the important difference that the rotating part (**armature**) consists of

(a) several coils: these experience torques whose average resultant value is steady, and

(b) a laminated soft-iron core, which provides both inertia (to maintain a steady angular speed), and a *radial field* (to provide a steady torque).

Because the armature is far more massive, it is supported on more robust bearings than those of the galvanometer.

D.C. motor

In the **d.c. motor** the current sense is reversed automatically twice in each rotation by the splitting commutator. This makes the sense of the torque constant.

Methods of winding the field coils and the armature are given in fig. 56.7 (p. 433).

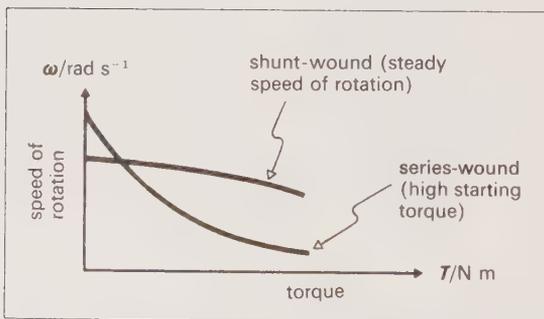


Fig. 56.8. Motor characteristics.

The characteristics of the different windings are given in fig. 56.8, and can be explained in terms of the back e.m.f.

The *series-wound* motor gives a large torque at slow speeds, and is used for pulling heavy loads (e.g. trains).

The *shunt-wound* motor has a steady current in the field coil, and so turns at a steady speed: it is used in situations where greater control is required (e.g. lathes and record-players).

In practice most motors are *compound-wound*, and have some of the advantages of both types.

A.C. motor

In the **a.c. motor** one tries to reverse the sense of current flow in both the armature and the field coil at the same instant, so as to maintain a torque in the same sense.

(a) In the *series-wound* motor the electromagnet changes polarity in phase with the change of current in the armature, as required.

(b) In the *shunt-wound* motor the large self-inductance (p. 436) of the field coil delays the establishment of the magnetic field until after the armature current has reached its maximum, and so such a motor cannot be used.

Back e.m.f. in the motor

An e.m.f. \mathcal{E}_B is induced in the rotating coil which tends to oppose the rotation of the coil.

Suppose V = p.d. applied to the coil

R = coil resistance

I = current through the coil

The loop equation (p. 390) gives

$$V - \mathcal{E}_B = IR$$

$$\therefore VI = I^2R + \mathcal{E}_BI$$

$$\left(\begin{array}{c} \text{power supplied} \\ \text{by Joule heating} \\ \text{in the armature} \end{array} \right) + \left(\begin{array}{c} \text{rate of} \\ \text{working} \\ \text{of motor} \end{array} \right)$$

Suppose the coil has N turns, each of area A , and that it rotates with an angular speed ω in a field B . The torque acting on the coil is given by

$$T = NIAB \quad (\text{p. 407})$$

$$= NAB \left(\frac{V - \mathcal{E}_B}{R} \right)$$

If there were no load and no friction, the motor would accelerate until the resultant torque became zero. This would occur when $V = \mathcal{E}_B$, where $\mathcal{E}_B = N\omega AB$ (p. 428). Note that the motor produces a mechanical power *output* whereas the dynamo

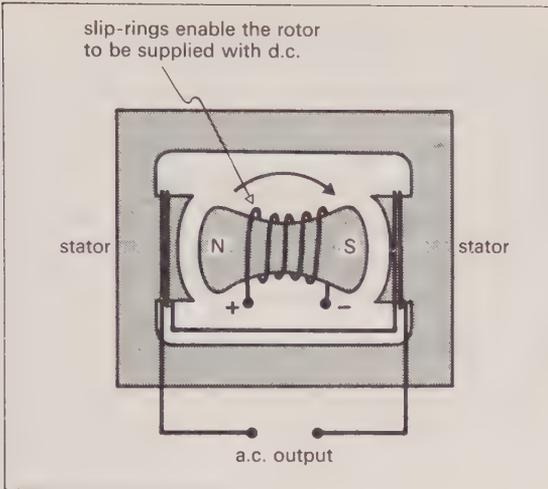


Fig. 56.9. The alternator.

requires a mechanical power *input*. For both devices, the armature resistance R should be as low as possible if high efficiency is to be achieved.

The need for a starting resistance

When the motor is first switched on, $\omega = 0$, so $I_s = V/R$. This means that I_s is large if R is small.

When the motor is running, \mathcal{E}_B increases, and so I decreases to its working value.

To prevent the armature burning out under a higher starting current, it is placed in series with a rheostat, whose resistance is decreased as the motor gathers speed.

56.7 THE ALTERNATOR

The output current of the usual a.c. generator is carried by the slip-rings and brushes. In the **alternator** of fig. 56.9 the slip-rings carry only the relatively small current needed to energize the electromagnet.

This is achieved by rotating the magnet (the **rotor**) relative to the coil in which the current is induced (the **stator**).

57 INDUCTANCE

57.1	DEFINITIONS OF SELF-INDUCTANCE	436
57.2	CALCULATION OF SELF-INDUCTANCE	437
▶ 57.3	THE LR D.C. CIRCUIT	437
▶ 57.4	THE ENERGY STORED BY AN INDUCTOR	438

▶ 57.5	DEFINITIONS OF MUTUAL INDUCTANCE M	439
▶ 57.6	CALCULATION OF MUTUAL INDUCTANCE	440
57.7	INDUCTANCE AND SWITCHES	440

57.1 DEFINITIONS OF SELF-INDUCTANCE

Fig. 57.1 shows a single coil in which the current is being made to increase.

The flux threading the coil increases, and the *change of flux* induces an e.m.f. in the coil itself.

By *Lenz's Law* the e.m.f. opposes the change: it tends to drive a current *against* the current that is *increasing*. If a current is *decreasing*, the induced e.m.f. tends to drive a current *in the same direction* as the original current to preserve the status quo.

For quantitative work we suppose no ferromagnetic materials are present. Then it follows from the *Biot-Savart Law* (p. 410) that $\Phi \propto I$.

We define the self-inductance L in two equivalent ways:

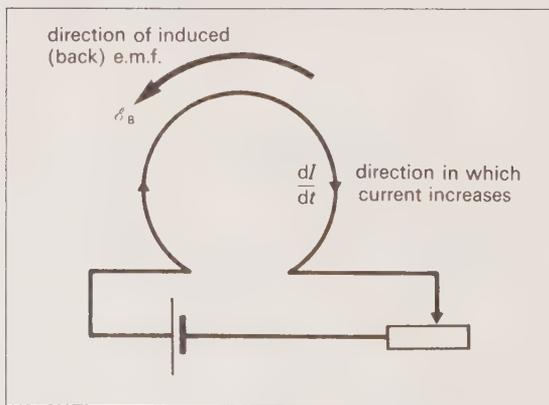


Fig. 57.1. Self-inductance

(A) Definition used to find L

Suppose a current I in a coil of N turns causes a flux Φ to thread each turn.

The **self-inductance** L is defined by the equation

$$N\Phi = LI$$

$$N\Phi[\text{Wb}] = L \left[\frac{\text{Wb}}{\text{A}} \right] I[\text{A}]$$

L is numerically equal to the flux-linkage of a circuit when unit current flows through it.

(B) Definition that describes the behaviour of an inductor in a circuit

Using *Neumann's equation* on p. 426, we can calculate the self-induced e.m.f. \mathcal{E} from

$$\begin{aligned} \mathcal{E} &= -\frac{d}{dt} N\Phi \\ &= -L \frac{dI}{dt} \end{aligned}$$

The **self-inductance** L may also be defined by the equation

$$\mathcal{E} = -L \frac{dI}{dt}$$

$$\mathcal{E}[V] = -L \left[\frac{V s}{A} \right] \frac{dI}{dt} \left[\frac{A}{s} \right]$$

L is numerically equal to the e.m.f. induced in the circuit when the current changes at the rate of 1 ampere in each second.

Notes

(a) The units $Wb A^{-1}$ and $V s A^{-1}$ are equivalent. They occur so often that each is conveniently called a **henry (H)**.

Thus $1 H = 1 V s A^{-1}$. The mH and μH are useful subdivisions.

(b) The permeability of free space μ_0 is measured in $Wb A^{-1} m^{-1}$.

Since $1 H = 1 Wb A^{-1}$, it follows that μ_0 may also be measured in $H m^{-1}$; this is the unit which is usually quoted.

(c) $[L] = [ML^2T^{-2}I^{-2}]$.

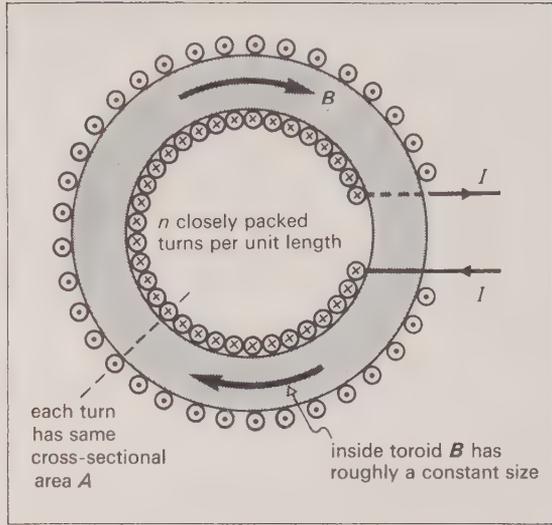


Fig. 57.2. Calculation of L for a toroid.

57.2 CALCULATION OF SELF-INDUCTANCE

Compare the defining equations

$$C = \frac{Q}{V} \quad \text{and} \quad L = \frac{N\Phi}{I}$$

In electrostatics we can calculate C if the geometry allows us to find V for a given Q . In electromagnetism we can find L if the geometry of a circuit allows us to find $N\Phi$ for a given I .

When a current flows in a *toroid*, the magnetic flux is entirely confined within the windings: outside B is self-cancelling. (Compare E inside and outside a parallel-plate capacitor.)

L for a toroid

Consider a finite solenoid of length l (\gg the diameter) bent into the doughnut shape of fig. 57.2. The magnitude of B is nearly constant over the cross-section, and given by

$$B = \mu_0 n I \quad (\text{p. 413})$$

The flux-linkage $N\Phi = nlBA$.

$$\begin{aligned} \text{So} \quad L &= \frac{N\Phi}{I} \\ &= \frac{nl(\mu_0 n I)A}{I} \\ &= \mu_0 n^2 (Al) \\ &= \mu_0 n^2 V \end{aligned}$$

where $V (= Al)$ is the volume of the toroid.

Notes

(a) The inductance of volume V of a long *straight* solenoid is also given by

$$L = \mu_0 n^2 V$$

near its centre.

(b) The inductance is $\propto n^2$. An increase in the number of turns per metre by a factor of k , increases

- (i) the flux through *each* turn by k , and also
- (ii) the number of turns to be linked by k .

The inductance then increases by a factor k^2 . A coil of 1000 turns may have a relatively large self-inductance (\sim mH) even when air-filled.

► 57.3 THE LR D.C. CIRCUIT

(A) Growth of current in the inductor

When the switch of fig. 57.3 (overleaf) is connected across AC, the inductor provides an e.m.f. in the opposite direction to that of the battery. Using the loop equation (p. 389) clockwise,

$$\Sigma \mathcal{E} = \Sigma IR$$

$$\mathcal{E} - L \frac{dI}{dt} = IR$$

We can check by substitution that the differential equation has the solution

$$I = \frac{\mathcal{E}}{R} (1 - e^{-t/(L/R)})$$

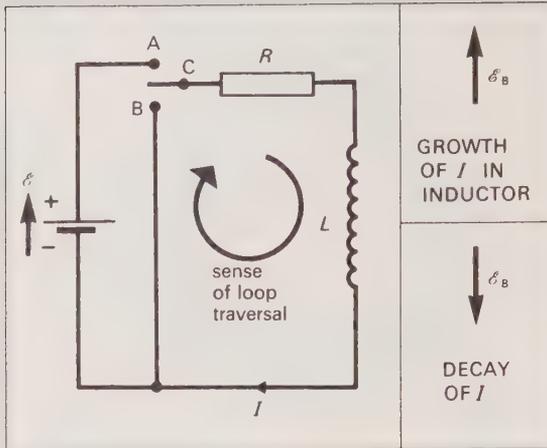


Fig. 57.3. Growth and decay of current in an LR circuit. The direction of the induced (back) e.m.f. is shown on the right.

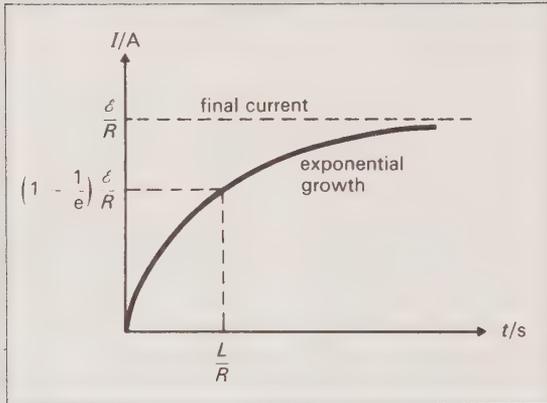


Fig. 57.4. Current growth in an LR circuit.

Fig. 57.4 shows that the growth of current toward its final value \mathcal{E}/R is exponential.

When $t = \left(\frac{L}{R}\right)$, $I = \frac{\mathcal{E}}{R} \left(1 - \frac{1}{e}\right)$

the current is smaller than its final value by $1/e$ or 37%. (L/R) is called the **inductive time constant** (τ_L).

$$\tau_L = \frac{L}{R}$$

$$\tau_L[\text{s}] = \frac{L[\text{V s/A}]}{R[\text{V/A}]}$$

E.g. If $L = 2.0 \text{ H}$ and $R = 100 \Omega$, then

$$\tau_L = \frac{L}{R} = 2.0 \times 10^{-2} \text{ s}$$

(B) Decay of current

After a time $(t \gg L/R)$, connect the switch from AC to BC. The inductor now provides an e.m.f. in the same direction as was the battery's.

In the loop equation $\mathcal{E} = 0$, and dI/dt is *negative*. (The back e.m.f. across the inductor reverses in direction.)

$$\therefore -L \frac{dI}{dt} = IR$$

This differential equation has the solution

$$I = \frac{\mathcal{E}}{R} e^{-t/(L/R)}$$

as can be checked by substitution.

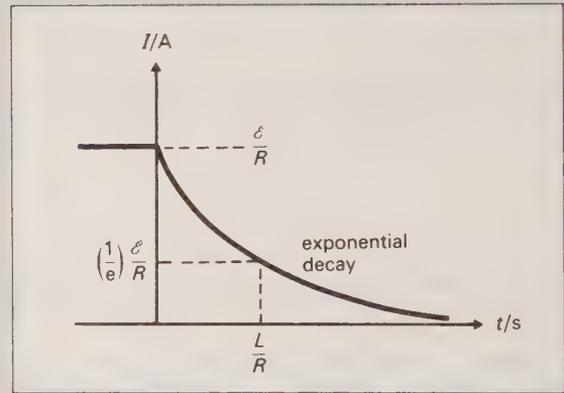


Fig. 57.5. Decay of the current through an inductor.

Fig. 57.5 shows that the current decays exponentially to zero.

When

$$t = \left(\frac{L}{R}\right) = \tau_L$$

the current has fallen to $1/e$ (or 37%) of its former value. This again is the inductive time constant, which has the same value for growth and decay.

► 57.4 THE ENERGY STORED BY AN INDUCTOR

An inductor carrying a current has energy associated with it. When the battery of a circuit is

disconnected, the inductor becomes a seat of e.m.f. as it converts its stored magnetic energy to electrical energy.

We can write the loop equation

$$\Sigma \mathcal{E} = \Sigma V$$

so
$$\mathcal{E} = L \frac{dI}{dt} + IR$$

in which $L(dI/dt)$ is regarded as a potential drop across the inductor.

$$\therefore \mathcal{E}I = LI \frac{dI}{dt} + I^2R$$

(input power from source) = (rate at which electrical energy is dissipated as internal energy)

Using the law of conservation of energy we regard $LI(dI/dt)$ as *absorbed but non-dissipated power*. It represents the rate at which the inductor stores energy in its associated magnetic field.

When the current reaches its final value I_0 the energy W stored is given by

$$\begin{aligned} W &= \int_0^{I_0} \left(LI \frac{dI}{dt} \right) \cdot dt \\ &= \int_0^{I_0} LI \cdot dI = \frac{1}{2} LI_0^2 \end{aligned}$$

$$W = \frac{1}{2} LI_0^2$$

The energy storage in the magnetic field of an inductor is analogous to that in the electric field of a capacitor.

Compare $W = \frac{1}{2} LI^2$
with $W = \frac{1}{2} Q^2/C$.

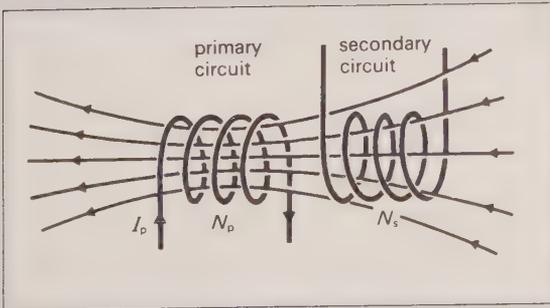


Fig. 57.6. Mutual induction.

► 57.5 DEFINITIONS OF MUTUAL INDUCTANCE M

Fig. 57.6 shows a situation in which a current in the first circuit (the **primary**) causes a magnetic flux to thread the second circuit (the **secondary**). As on p. 436, we suppose no ferromagnetic materials are present.

Suppose a current I_p in the primary circuit causes a flux Φ_s to link each turn of the secondary circuit.

(1) The **mutual inductance** M is defined by the equation

$$N_s \Phi_s = MI_p$$

Using *Neumann's equation* (p. 426) we can calculate the e.m.f. \mathcal{E}_s , induced in the secondary by a changing current in the primary, from

$$\begin{aligned} \mathcal{E}_s &= -\frac{d}{dt}(N_s \Phi_s) \\ &= -M \frac{dI_p}{dt} \end{aligned}$$

(2) The **mutual inductance** M may also be defined by the equation

$$\mathcal{E}_s = -M \frac{dI_p}{dt}$$

Notes

(a) It can be proved that the same value is obtained for M if one considers the flux threading the primary when a current flows in the secondary.

So
$$\begin{aligned} M &= \frac{N_s \Phi_s}{I_p} = \frac{N_p \Phi_p}{I_s} \\ &= -\frac{\mathcal{E}_s}{dI_p/dt} = -\frac{\mathcal{E}_p}{dI_s/dt} \end{aligned}$$

(b) The two definitions for M , like those for L , are equivalent, and show that M is measured in the unit **henry**.

Expressed in words

(1) M is numerically equal to the flux-linkage in one circuit when unit current flows through the other.

We use this definition to calculate M .

(2) M is numerically equal to the e.m.f. induced

in one circuit when the current changes in the other at the rate of 1 ampere in each second.

We use this definition to describe the mutual behaviour of two circuits.

► 57.6 CALCULATION OF MUTUAL INDUCTANCE

The situation and symbols are defined in fig. 57.7.

The flux density B inside the long solenoid is given by

$$B = \mu_0 \left(\frac{N_p}{l} \right) I_p$$

The flux at the centre = BA

$$= \mu_0 A \left(\frac{N_p}{l} \right) I_p$$

$$= \Phi_s$$

since *all* this flux links each turn of the secondary.

From the definition

$$\begin{aligned} M &= \frac{N_s \Phi_s}{I_p} \\ &= \frac{N_s [\mu_0 A (N_p / l) I_p]}{I_p} \\ &= \frac{\mu_0 A N_p N_s}{l} \end{aligned}$$

For an air-cored coil it is likely that $\Phi_s < BA$ because of magnetic flux leakage.

The idea of mutual inductance is used in the transformer (p. 431).

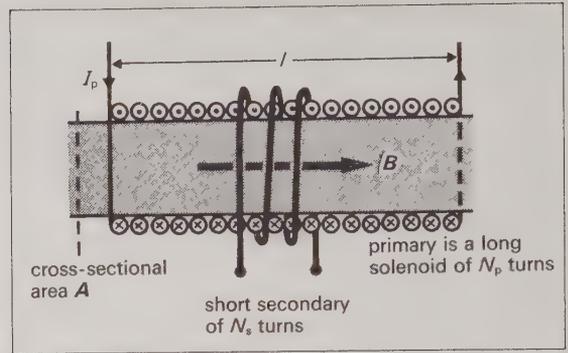


Fig. 57.7. Calculation of M for a short coil wound round a long solenoid.

57.7 INDUCTANCE AND SWITCHES

When an inductive circuit is switched on, the induced e.m.f. opposes the applied p.d. and delays the growth of the current to its final value. On switching off, the current drops rapidly to zero so that there is a high rate of change of flux linked with the inductor. The peak value of the pulse of induced e.m.f. may be several orders of magnitude greater than the original applied p.d. Once the circuit has been broken, the contacts must move apart quickly to reduce the danger of arcing between them. The design of switches is therefore very important to avoid the risk of fire or of melting the switch contacts. Domestic light switches use spring-loaded switch arms. Larger switch systems use jets of compressed air to blow out the arc, or they have the contacts mounted in a tank of non-inflammable oil which suppresses the arc.

58 ALTERNATING CURRENT

58.1	ALTERNATING CURRENT AND E.M.F.	441
58.2	AVERAGE VALUE OF A SINUSOIDALLY VARYING CURRENT	442
58.3	THE EFFECTIVE VALUE OF SINUSOIDALLY VARYING CURRENT	442
58.4	METHODS OF MEASURING ALTERNATING QUANTITIES	442
58.5	THE EFFECTS OF ALTERNATING CURRENT	443
58.6	IMPEDANCE	444
58.7	REPRESENTATION OF A SINUSOIDAL QUANTITY BY ROTATING VECTOR	445
▶ 58.8	ADDITION OF TWO SINUSOIDAL QUANTITIES OF THE SAME FREQUENCY	445
58.9	A PURELY RESISTIVE CIRCUIT	445

58.10	A PURELY INDUCTIVE CIRCUIT	446
58.11	A PURELY CAPACITATIVE CIRCUIT	448
58.12	SUMMARY OF CIRCUIT ELEMENTS TAKEN SEPARATELY	448
▶ 58.13	R, L AND C IN SERIES	449
▶ 58.14	R, L AND C IN PARALLEL	450
▶ 58.15	POWER IN A.C. CIRCUITS	451
▶ 58.16	RESONANCE	452
▶ 58.17	ELECTRICAL OSCILLATIONS	453
▶ 58.18	AN ANALOGUE BETWEEN MECHANICAL AND ELECTRICAL OSCILLATIONS	454
58.19	RECTIFICATION OF A.C.	455
58.20	SMOOTHING CIRCUITS	456
58.21	THE THREE-PHASE SYSTEM	458

58.1 ALTERNATING CURRENT AND E.M.F.

An *alternating* current or e.m.f. is one whose magnitude and direction vary periodically with time.

The simplest a.c. is one which varies sinusoidally (in a simple harmonic manner) with time. It is of great importance

(a) because it can be produced by rotating a coil in a *uniform* magnetic field (p. 427), and

(b) because *any* periodic variation is equivalent to a combination of sinusoidal variations of appropriate frequencies, amplitudes and phase differences.

The *frequency* f of alternation, *angular frequency* (**pulsatance**) ω and *period* T are related by

$$\omega = 2\pi f = \frac{2\pi}{T}$$

Typical values for f are:

British mains	50 Hz (50 s ⁻¹)
Radio frequencies	10 ⁵ to 2 × 10 ¹⁰ Hz
(Visible light	~10 ¹⁵ Hz)

Symbols

Throughout this chapter, this convention is adopted:

	Current	e.m.f.
instantaneous values	I	\mathcal{E}
root mean square (r.m.s.) values	$I_{r.m.s.}$	$\mathcal{E}_{r.m.s.}$
peak values	I_0	\mathcal{E}_0

Phase, phase difference, lag and lead

If $I = 0$ when $t = 0$, we write

$$I_1 = I_0 \sin \omega t$$

A second current I_2 may be given by

$$I_2 = I_0 \sin (\omega t + \phi)$$

The two currents have the same pulsantance ω , and the same amplitude I_0 (**peak** or maximum value), but differ in **phase**.

$$\left. \begin{array}{l} I_1 \text{ has a phase } \omega t \\ I_2 \text{ has a phase } (\omega t + \phi) \end{array} \right\}$$

The two currents are **out of phase**: their **phase difference** (or **shift**) is ϕ [rad], which corresponds to a time interval of ϕ/ω [seconds]. (See p. 103.)

I_2 leads I_1 by the angle ϕ ,
 I_1 lags I_2 by the angle ϕ .

Two *different* alternating quantities (such as p.d. and current) may also show a phase difference when they vary with the same frequency. Examples appear on p. 446, p. 448 and p. 449.

58.2 AVERAGE VALUE OF A SINUSOIDALLY VARYING CURRENT

By definition the *average value* of any periodic function f over its period T is given by

$$f_{av} = \frac{\int_0^T f \cdot dt}{T} \quad (\text{p. 23})$$

Consider $I = I_0 \sin \omega t$

(A) Whole cycle

By the symmetry of the sine curve, the average current $\langle I \rangle = 0$ over a whole number of cycles.

(B) Half cycle

$$\begin{aligned} \langle I \rangle &= \frac{1}{T/2} \int_0^{T/2} I_0 \sin \omega t \cdot dt & T &= \frac{2\pi}{\omega} \\ &= \frac{1}{\pi/\omega} \int_0^{\pi/\omega} I_0 \sin \omega t \cdot dt & \frac{T}{2} &= \frac{\pi}{\omega} \\ &= \frac{I_0}{\pi/\omega} \left[-\frac{\cos \omega t}{\omega} \right]_0^{\pi/\omega} = \frac{I_0}{\pi/2} \end{aligned}$$

Example. The pulsating d.c. of fig. 58.1 has an average value of $I_0/(\pi/2)$ over a whole number of half cycles.

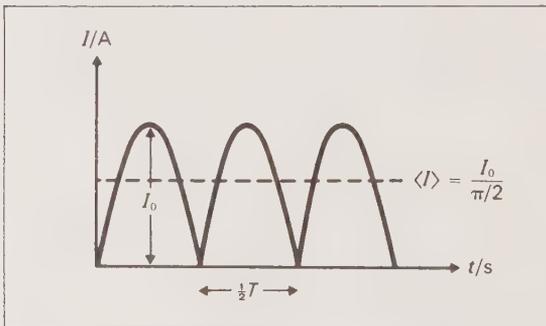


Fig. 58.1. Average value of full-wave rectification.

58.3 THE EFFECTIVE VALUE OF SINUSOIDALLY VARYING CURRENT

The **effective value** of an a.c. is that steady current which would convert electrical energy to other forms at the same rate as the a.c. (It is sometimes called the **virtual current**.)

The *instantaneous* rate of conversion

$$I^2 R = I_0^2 R \cdot \sin^2 \omega t$$

The *average* rate of conversion

$$\begin{aligned} \langle I_0^2 R \cdot \sin^2 \omega t \rangle &= I_0^2 R \cdot \langle \sin^2 \omega t \rangle \\ &= \frac{1}{2} I_0^2 R \\ &= \left(\frac{I_0}{\sqrt{2}} \right)^2 R \end{aligned}$$

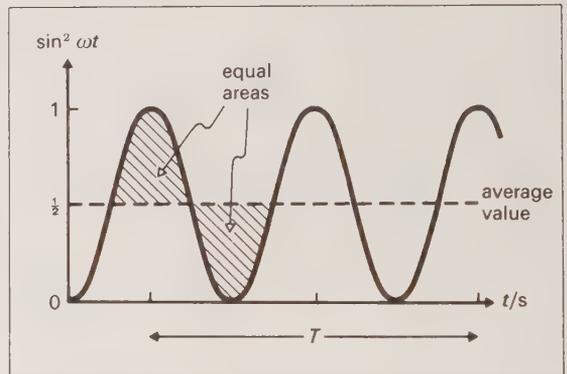


Fig. 58.2. Average value of the $\sin^2 \omega t$ curve.

Thus when the resistance is purely ohmic the effective value of current

$$= \sqrt{(\text{mean value of } I^2)}$$

and is called the **root mean square current**.

$$I_{r.m.s.} = \frac{I_0}{\sqrt{2}}$$

Example. For the British a.c. mains, we have

$$\mathcal{E} = \mathcal{E}_0 \sin \omega t$$

in which $\mathcal{E}_{r.m.s.} = 240 \text{ V}$

$$\begin{aligned} \text{so} \quad \mathcal{E}_0 &= 240\sqrt{2} \text{ V} \\ &= 339 \text{ V} \end{aligned}$$

$\mathcal{E}_{r.m.s.}$ is usually quoted.

58.4 METHODS OF MEASURING ALTERNATING QUANTITIES

(A) Suppose an instrument gives a deflection $\propto I$

Then if it has been calibrated by d.c., and a.c. is passed through it, it may read

- (a) zero (e.g. an unadapted moving-coil galvanometer),
- (b) $\frac{1}{2} I_0 / (\pi/2)$, if half-wave rectification is used,
- (c) $I_0 / (\pi/2)$, if full-wave rectification is used (such as that in the bridge circuit of fig. 54.2).

(B) Suppose the instrument gives a deflection $\propto I^2$

Then it may read

- (d) $I_{r.m.s.} = I_0 / \sqrt{2}$ if it has high inertia, and the frequency is high,
- (e) I_0 if it has low inertia, and the frequency is low,
- (f) a varying value between (d) and (e) if it tries to follow the variation, but is unable to keep up.

As a general rule, meters designed for a.c. use will read *root mean square values*. Examples include the current balance, electro-dynamometer, hot-wire and thermocouple ammeter, and moving-iron instruments.

The diode voltmeter and c.r.o. both record *peak values*.

(B) *Chemical effect*

The observed effect depends on the frequency. At 50 Hz

- (a) dilute $H_2SO_4 \rightarrow (2H_2 + O_2)$, a gaseous mixture at each platinum electrode, but
- (b) $CuSO_4$ gives zero *net* effect with Cu electrodes.

At higher frequencies the fact that an ion has a much greater mass and therefore smaller mobility than an electron may prevent sufficient movement to show any effect.

An a.c. can give an electric shock, i.e. does produce a *physiological effect*.

(C) *Heating effect*

Joule heating is not dependent on the direction of the current. A filament lamp may show 100 brightnesses per second, since the *power* has twice the frequency of the current. The heat capacity of the filament controls the variation of brightness, but the filament may give a *stroboscopic effect*. (A discharge lamp has quicker response.)

The thermocouple ammeter

This uses the heating effect of the current, and is more satisfactory than the hot-wire ammeter.

The thermocouple (fig. 58.3) has the *net* effect of converting the a.c. to a d.c. which is measured by a sensitive moving-coil galvanometer.

58.5 THE EFFECTS OF ALTERNATING CURRENT

An alternating current differs from a direct current only in the *varying* magnitude and direction of the drift velocity imposed on the charge-carriers' random movement. The effects of a.c. are fundamentally the same as those of d.c., but this may not be apparent.

(A) *Magnetic effect*

The a.c. gives a time-varying value of B , which

- (a) causes torques and forces of varying size to be exerted on coils and magnetic dipoles: when $\omega = 2\pi \times 50 \text{ rad s}^{-1}$, they may not respond significantly because of their inertia.
- (b) is part of an electromagnetic wave (the charge-carriers are accelerated).

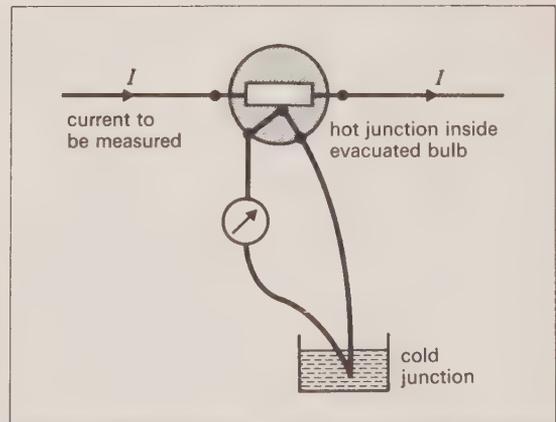


Fig. 58.3. A thermocouple ammeter.

58.6 IMPEDANCE

For d.c. we define *resistance* R by $R = V/I$. Suppose that for an a.c. the maximum value of the applied p.d. across a device is V_0 , and that the maximum current through the device is I_0 .

The **impedance** Z of the device is defined by the equation

$$Z = \frac{V_0}{I_0} = \frac{V_{r.m.s.}}{I_{r.m.s.}}$$

$$Z [\Omega] = \frac{V_{r.m.s.} [V]}{I_{r.m.s.} [A]}$$

Notes

(a) Z plays the same role in an a.c. circuit as R does in a d.c. circuit: it is the property of a component or a circuit that opposes the passage of a.c., and (together with V_0) determines its maximum value.

(b) Because V_0 and I_0 may occur at different instants of time, it is *not* true (in general) to say $Z = V/I$, where V and I are instantaneous values.

(c) The value of Z depends upon

- (i) the frequency of the a.c., and
 - (ii) the values of R , L and C .
- (d) $Z \geq R$ in series circuits.

Demonstration of impedance

In the circuit of fig. 58.4, $V_{r.m.s.}$ is kept constant and $I_{r.m.s.}$ noted by a milliammeter.

Action	Qualitative observation	Quantitative deduction
increase ω	for R , $I_{r.m.s.} = \text{constant}$	R does not depend on ω^*
increase ω	for L , $I_{r.m.s.}$ decreases	$Z \propto \omega$
increase ω	for C , $I_{r.m.s.}$ increases	$Z \propto \frac{1}{\omega}$

*In practice R may be found to depend on ω because of the **skin effect** (when an a.c. is largely confined to surface layers), and because of various dissipative effects in inductive coils.

The deductions are derived theoretically in this chapter.

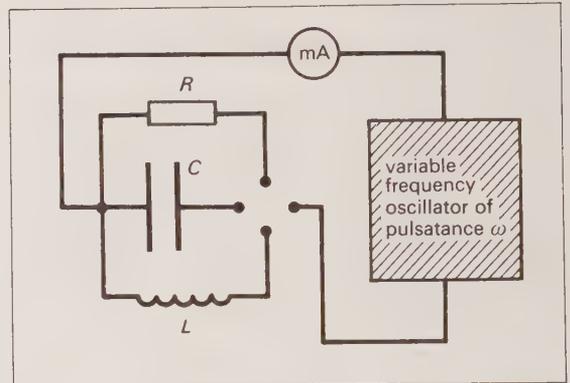


Fig. 58.4. Demonstration that the impedance of a device depends upon ω .

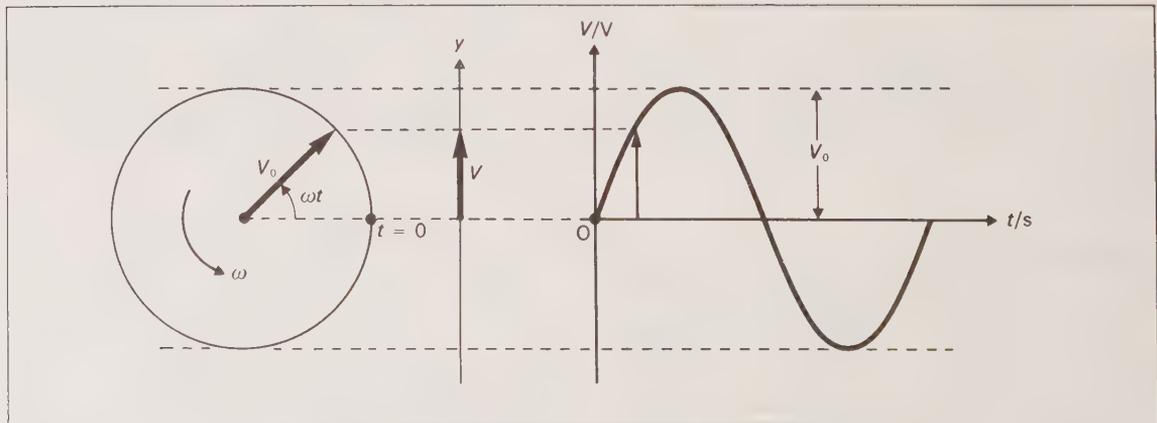


Fig. 58.5. $V = V_0 \sin \omega t$ represented by a rotating vector.

58.7 REPRESENTATION OF A SINUSOIDAL QUANTITY BY ROTATING VECTOR

Suppose we have a vector of magnitude V_0 rotating in an anticlockwise sense at constant angular speed ω (fig. 58.5). It is referred to as a **phasor**.

The y -component of the vector at any instant has magnitude $V = V_0 \sin \omega t$.

Thus we can represent a sinusoidal quantity by the projection on to a fixed line of a rotating vector.

By convention the anticlockwise sense of rotation is taken to be *positive*.

► 58.8 ADDITION OF TWO SINUSOIDAL QUANTITIES OF THE SAME FREQUENCY

Method

(1) Represent the two quantities by rotating vectors.

(2) Use the fact that

$$\left(\begin{array}{l} \text{projection of vector} \\ \text{sum of two vectors} \end{array} \right) = \left(\begin{array}{l} \text{sum of projections} \\ \text{of two vectors} \end{array} \right)$$

This is shown in fig. 58.6.

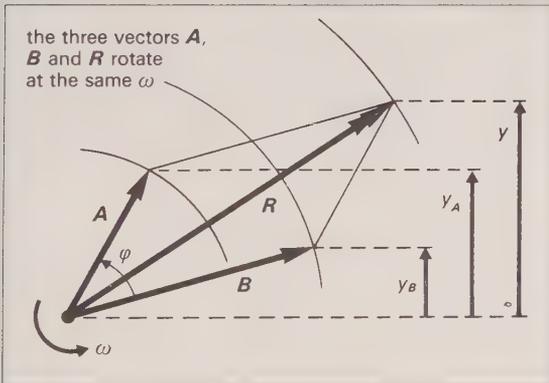


Fig. 58.6. Addition of sinusoidal quantities.

y is the projection of R , where

(a) R = vector sum of A and B , and

(b) $y = y_A + y_B$
 $= A \sin \omega t + B \sin (\omega t - \phi)$.

The **phase lag** ϕ of y_B behind y_A corresponds to a **time lag** Δt , where

$$\frac{\Delta t}{T} = \frac{\phi}{2\pi}$$

so
$$\Delta t = \left(\frac{\phi}{2\pi} \right) \cdot T = \frac{\phi}{\omega}$$

Δt is fixed because A and B have the same angular frequency ω .

Notes

(a) The resultant y has the same frequency as y_A and y_B , but reaches its maximum at an instant different from either y_A or y_B (i.e. shows a *phase shift*).

(b) In general the magnitude of R is not equal to the sum of the magnitudes of A and B .

(c) We apply these ideas to add p.d.s across components in a.c. circuits. These p.d.s have the same frequency but are *out of phase*.

Important Note

In sections 58.9, 58.10, 58.11 and 58.12, the current is described by

$$I = I_0 \sin \omega t$$

The phase of the applied p.d. is chosen appropriately to make this so.

58.9 A PURELY RESISTIVE CIRCUIT

(A) Phase relation for I and V

Let applied p.d. $V = V_0 \sin \omega t$.

Using the definition

$$R = \frac{V}{I}$$

we have
$$I = \frac{V}{R} = \frac{V_0}{R} \sin \omega t$$

$$= I_0 \sin \omega t$$

Fig. 58.7 (overleaf) shows

(a) I and V have the same frequency,

(b) I and V are *in phase*.

$$R = \frac{V_0}{I_0}$$

(B) Power

Instantaneous power dissipation

$$P = VI$$

$$= (V_0 \sin \omega t) \times (I_0 \sin \omega t)$$

The power oscillates at an angular frequency 2ω , since

$$\sin^2 \omega t = \frac{1}{2}(1 - \cos 2\omega t)$$

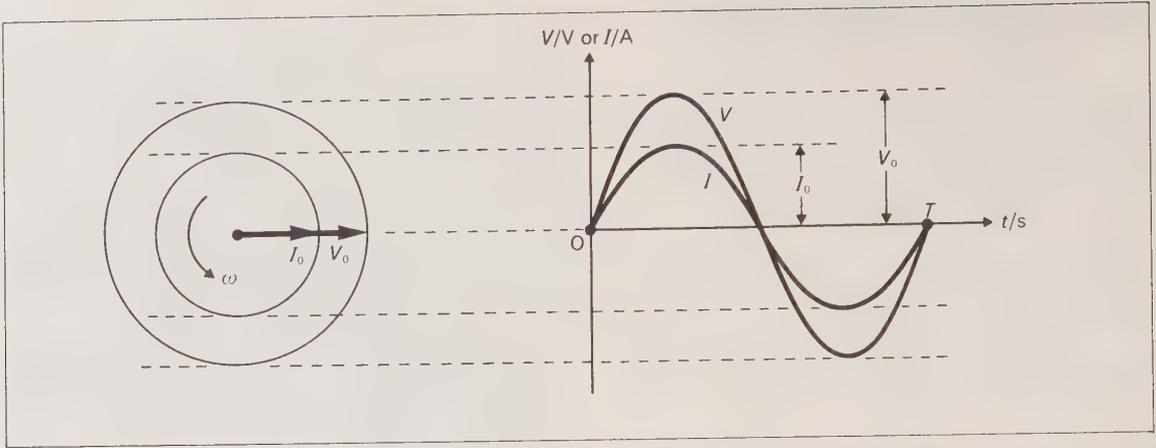


Fig. 58.7. V and I in a purely resistive a.c. circuit.

Average power dissipation

$$\langle P \rangle = V_{\text{r.m.s.}} I_{\text{r.m.s.}}$$

which is always positive (fig. 58.2).

The resistor causes a *non-reversible* conversion of electrical energy to thermal energy.

58.10 A PURELY INDUCTIVE CIRCUIT

(A) Reactance and phase relation for I and V

Since we have assumed $R = 0$ in fig. 58.8, if we apply the loop equation $\sum \mathcal{E} = \sum IR$ in a clockwise sense,

$$V_0 \cos \omega t - L \frac{dI}{dt} = 0$$

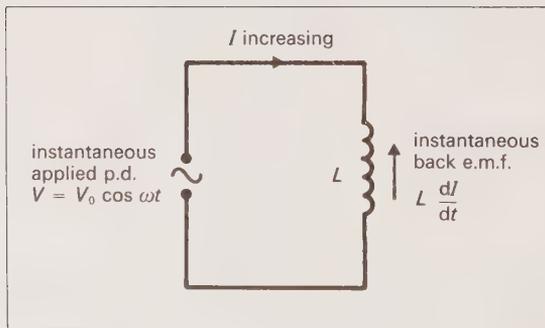


Fig. 58.8. A purely inductive circuit.

$$\frac{dI}{dt} = \frac{V_0}{L} \cos \omega t$$

Integrating $I = \left(\frac{V_0}{\omega L} \right) \cdot \sin \omega t$
 $= I_0 \sin \omega t$

Fig. 58.9 (opposite) shows

- (a) I and V have the same frequency,
- (b) the current lags behind the applied p.d. with a *phase lag* of $\pi/2$ rad: the current reaches its maximum value $T/4$ later in time.

We define the **inductive reactance** X_L by the equation

$$X_L = \frac{V_0}{I_0} = \frac{V_{\text{r.m.s.}}}{I_{\text{r.m.s.}}}$$

Since (above) $I_0 = \frac{V_0}{\omega L}$

$$X_L = \omega L$$

$$X_L [\Omega] = \omega \left[\frac{1}{s} \right] \cdot L \left[\frac{V}{A/s} \right]$$

ωL controls the current magnitude in the same way as resistance does. The inductive reactance is increased by

- (a) increasing ω , and
- (b) increasing L .

See the *choke effect* (p. 458).

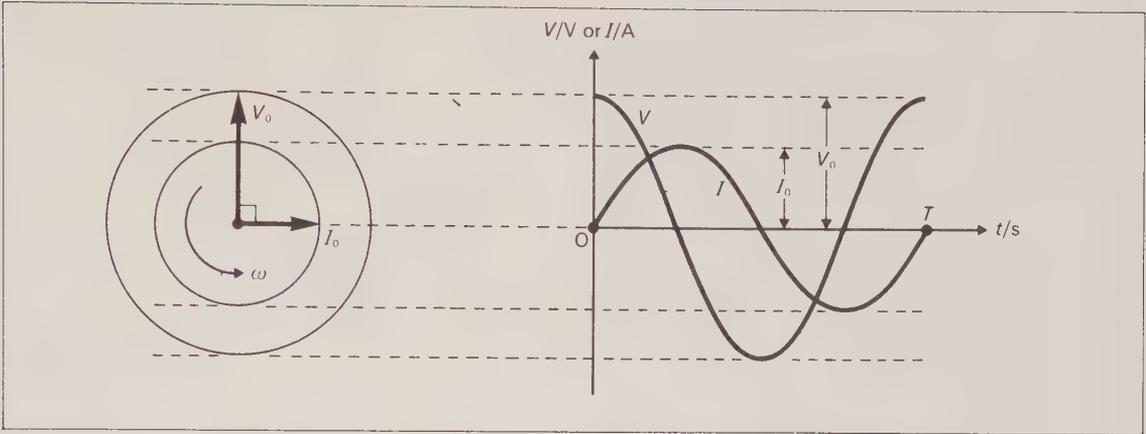


Fig. 58.9. V and I in an inductive circuit. (The symbol \curvearrowright for L is helpful for remembering the vector diagram.)

(B) Power

Instantaneous power conversion

$$P = VI$$

$$= (V_0 \cos \omega t) \times (I_0 \sin \omega t)$$

The power oscillates at an angular frequency 2ω , as shown in fig. 58.10.

The *average power* (rate of energy conversion) $\langle P \rangle$ over a cycle is shown to be zero. Over a whole number of cycles there is no *net* energy transfer from the seat of e.m.f. to the magnetic field associated with the inductor. (In practice energy is always dissipated because R is finite.)

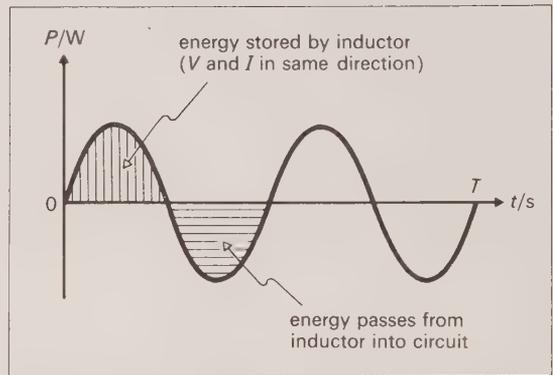


Fig. 58.10. Power of an inductive circuit.

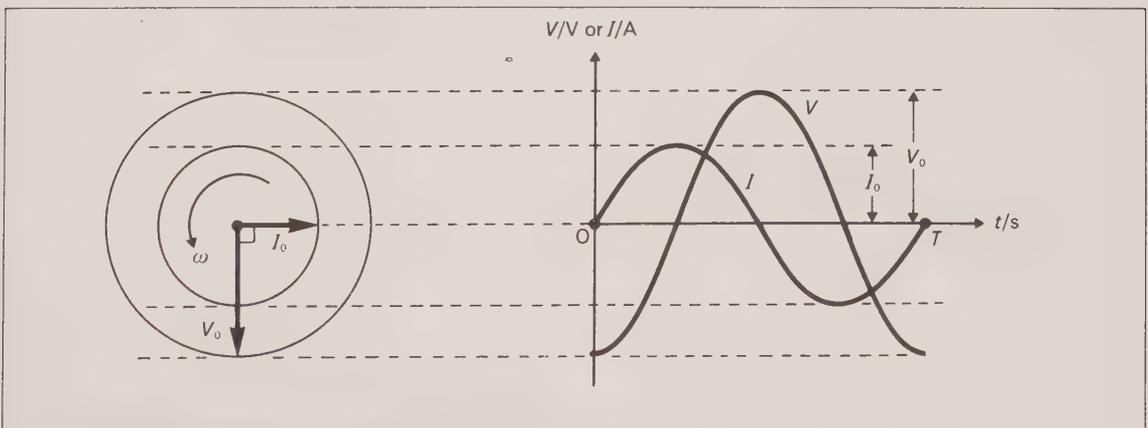


Fig. 58.11. V and I in a capacitive circuit.

58.11 A PURELY CAPACITATIVE CIRCUIT

(A) Reactance and phase relation for I and V

For a capacitor $V = Q/C$ defines C .
The instantaneous current

$$I = \frac{dQ}{dt} = C \frac{dV}{dt}$$

We choose the applied p.d. $V = -V_0 \cos \omega t$ (for the reason on p. 445),

so
$$\frac{dV}{dt} = \omega V_0 \sin \omega t$$

and
$$I = (\omega C) V_0 \sin \omega t$$

$$= I_0 \sin \omega t$$

Fig. 58.11 (p. 447) shows
(a) I and V have the same frequency,
(b) the current leads the applied p.d. with a phase lead of $\pi/2$ rad: the current reaches its maximum value $T/4$ earlier in time.

Note that there is no real current flow *through the capacitor*, but the oscillation of charge through the circuit continually changes the sign of the charge on the capacitor plates.

We define the **capacitive reactance** X_C by the equation

$$X_C = \frac{V_0}{I_0} = \frac{V_{r.m.s.}}{I_{r.m.s.}}$$

Since (above)

$$I_0 = \frac{V_0}{(1/\omega C)}$$

$$X_C = \frac{1}{\omega C}$$

$$X_C [\Omega] = \frac{1}{\omega [1/s] C [A s/V]}$$

The capacitive reactance is increased by

- (a) reducing ω , and
- (b) reducing C .

(B) Power

Instantaneous power conversion

$$P = VI$$

$$= (-V_0 \cos \omega t) \times (I_0 \sin \omega t)$$

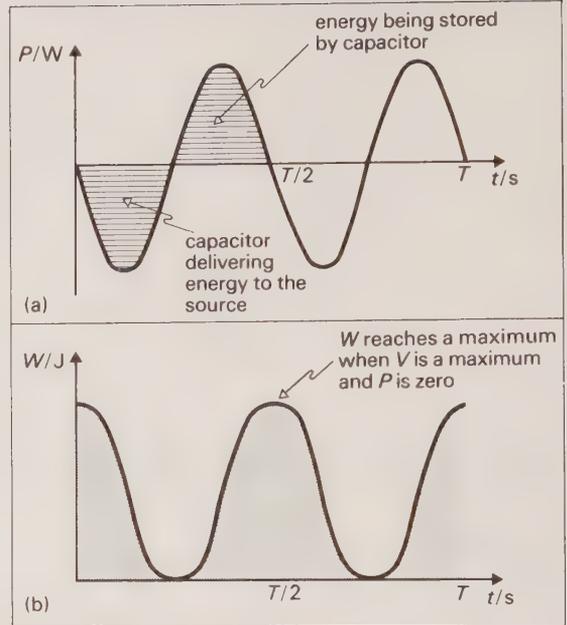


Fig. 58.12. (a) The power, and (b) the energy storage in a capacitive circuit.

The power oscillates at an angular frequency 2ω , as shown in fig. 58.12.

The *average power* (rate of energy conversion) $\langle P \rangle$ over a cycle is zero, since the capacitor absorbs energy when the field between its plates is increasing, but returns it to the source as the field collapses. (In practice energy is always dissipated because R is finite.)

58.12 SUMMARY OF CIRCUIT ELEMENTS TAKEN SEPARATELY

Property of device	Resistance or reactance X	$V-I$ phase relation	Instantaneous energy storage
R	R	in phase	nil
L	$X_L = \omega L$	V leads I by $\pi/2$ rad	$\frac{1}{2} LI^2$
C	$X_C = 1/(\omega C)$	V lags I by $\pi/2$ rad	$\frac{1}{2} CV^2$

This phase shift is independent of the frequency.

Fig. 58.13 (opposite) shows the limiting values of ωL and $1/\omega C$ as $\omega \rightarrow 0$, and $\omega \rightarrow \infty$.

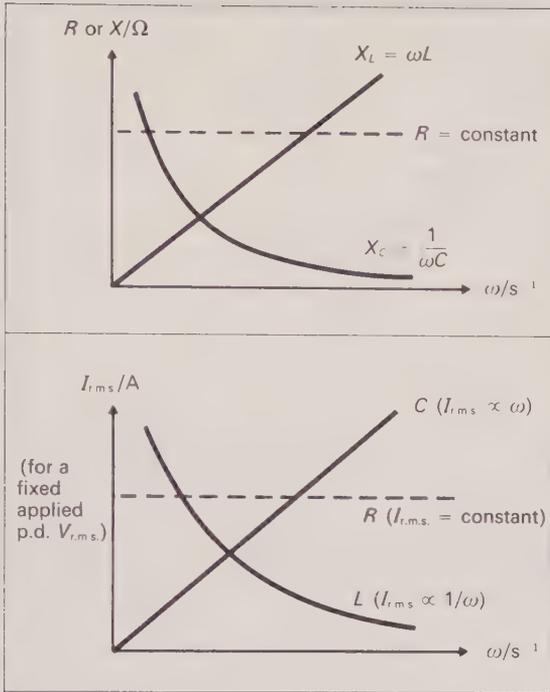


Fig. 58.13. The effect of the pulsance ω on reactance and current.

► 58.13 R, L AND C IN SERIES

Statement of the problem

Refer to fig. 58.14. Note that $V = V_0 \sin \phi$ when $t = 0$.

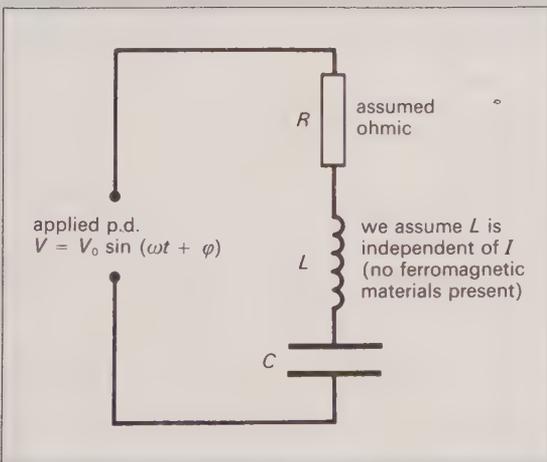


Fig. 58.14. The RLC series circuit.

We want to calculate

- (1) ϕ (and Z) in terms of L, C and R (and ω),
- (2) the net rate of transfer of energy from the source to the circuit (the power dissipation).

Principles

(a) The law of charge conservation tells us that the instantaneous value of I through R, L, C and the source must be the same.

(b) The law of energy conservation gives us the loop equation. At any instant $V = V_R + V_L + V_C$.

Analysis by differential equation

The loop equation (above) can be written

$$V_0 \sin(\omega t + \phi) = IR + L \frac{dI}{dt} + \frac{Q}{C}$$

$$= L \frac{d^2Q}{dt^2} + R \frac{dQ}{dt} + \frac{Q}{C}$$

since $I = \frac{dQ}{dt}$.

The solution of this equation gives Q , and hence I , as a function of t . We will not adopt this method.

Solution by rotating vector method

Since V_R is in phase with I , we can relate the phases of V_R, V_L and V_C as in fig. 58.15 (overleaf).

By choice $V_R = 0$ when $t = 0$. To evaluate

$$V = V_R + V_L + V_C$$

we can either

- (a) add their instantaneous values, or
- (b) add $V_{R,0} + V_{L,0} + V_{C,0}$ by vector addition and calculate the projection of their resultant.

This is done in fig. 58.16 (overleaf).

The phase shift between V_0 and I_0 is given by

$$\tan \phi = \frac{V_{X,0}}{V_{R,0}} = \frac{V_{L,0} - V_{C,0}}{V_{R,0}}$$

The relationship between V and I is given in fig. 58.17 (overleaf). This shows how

- (a) I and V have the same frequency,
- (b) the applied p.d. leads the current (in this example) by a phase angle ϕ . ϕ is negative when $V_{C,0} > V_{L,0}$.

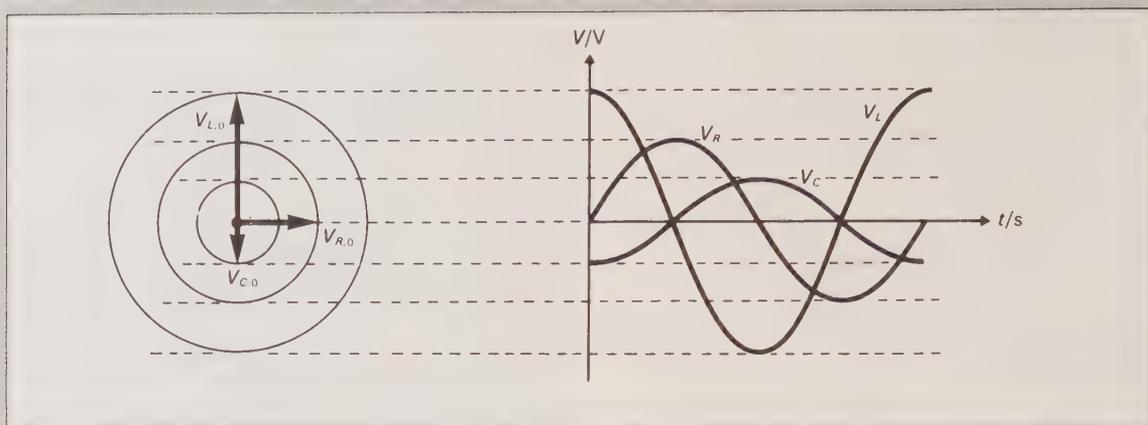


Fig. 58.15. The variation with time of the p.d.s across the three components.

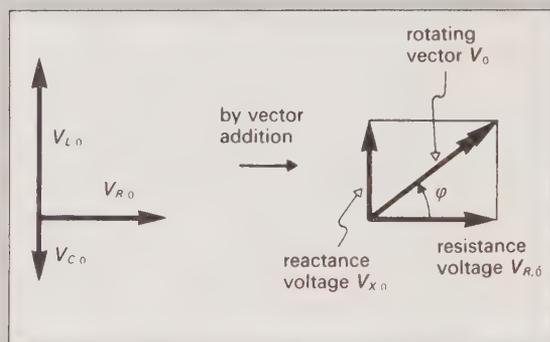


Fig. 58.16. Vector addition of maximum p.d.s.

Summary diagram

Fig. 58.16 shows that

$$V_0 = \sqrt{[V_{R,0}^2 + (V_{L,0} - V_{C,0})^2]}$$

Dividing each side by I_0 , we have

$$Z = \frac{V_0}{I_0} = \sqrt{[R^2 + (X_L - X_C)^2]}$$

(see fig. 58.18 on p. 451).

Remember that current is common to all elements of the *series* circuit.

To investigate

- (a) an RC circuit, let $L \rightarrow 0$,
- (b) an LR circuit, let $C \rightarrow \infty$, or
- (c) an LC circuit, let $R \rightarrow 0$.

Power dissipation is discussed on p. 451.

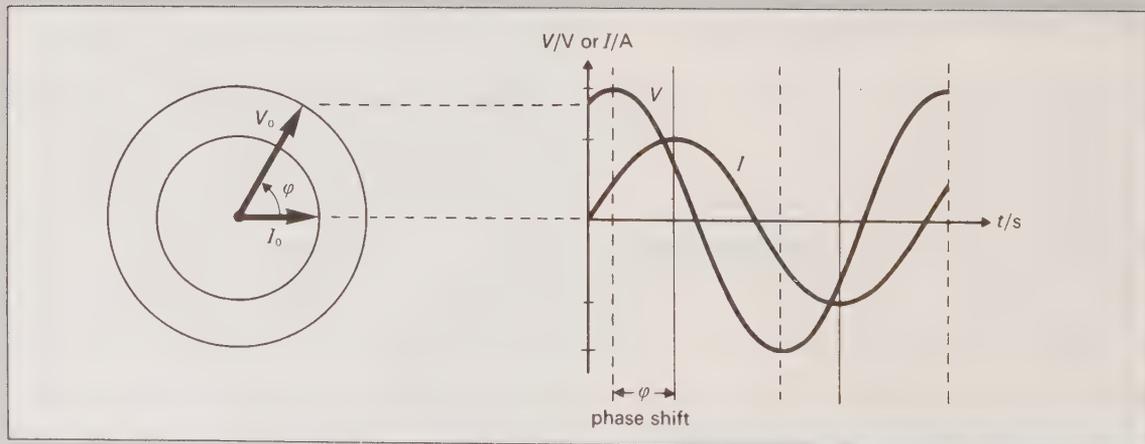


Fig. 58.17. V and I in an RLC series circuit.

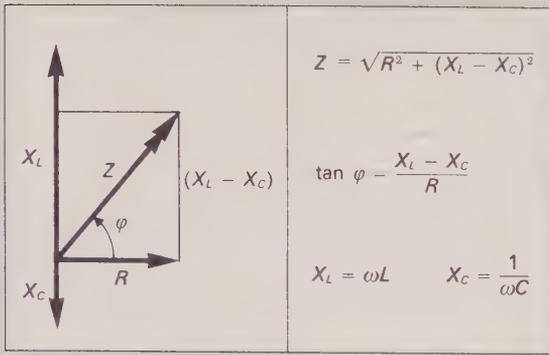


Fig. 58.18. How to calculate impedance in an RLC series circuit.

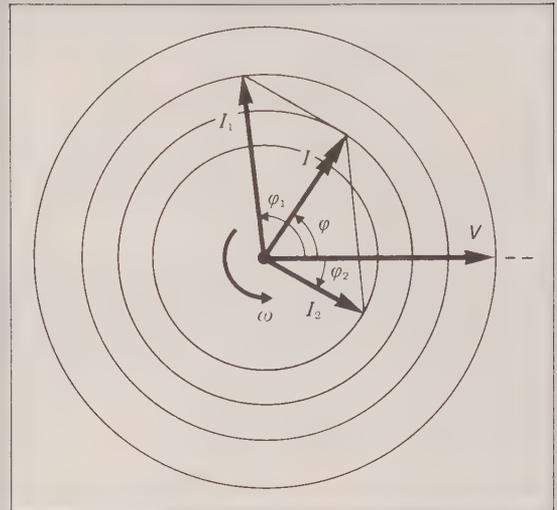


Fig. 58.20. Rotating vector diagram for the circuit of fig. 58.19.

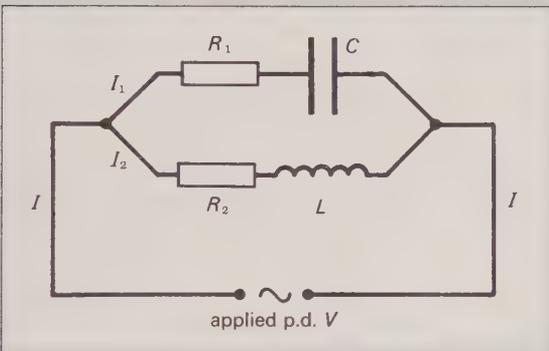


Fig. 58.19. An RLC parallel circuit.

► 58.14 R, L AND C IN PARALLEL

The parallel circuit in fig. 58.19 is different from a series circuit in that

- (1) the instantaneous p.d. V across both branches is the same,
- (2) the instantaneous current I is equal to the vector sum of the instantaneous currents I_1 and I_2 , because
- (3) the currents I_1 and I_2 are out of phase (I_1 leads I_2).

The equivalent impedance of the circuit can be found by using the rotating vector method. From fig. 58.20 it can be seen that:

(a) I_1 leads V by ϕ_1 where

$$\phi_1 = \arctan \frac{X_C}{R_1}$$

(b) I_2 lags V by ϕ_2 where

$$\phi_2 = \arctan \frac{X_L}{R_2}$$

(c) I leads V by ϕ .

The equivalent impedance of the circuit is given by $Z = V/I$; where V is known and I must be calculated.

The evaluation of I involves the calculation of I_1 and I_2 , their phase angles, and some trigonometry.

The final expression for the equivalent impedance will obviously be complicated and is not given here, although it can be readily derived from fig. 58.20.

► 58.15 POWER IN A.C. CIRCUITS

The RLC series circuit

Instantaneous power

Refer to fig. 58.21 (overleaf).

$$\begin{aligned} P &= VI \\ &= (V_0 \sin(\omega t + \phi)) \times (I_0 \sin \omega t) \\ &= V_0 I_0 \sin \omega t (\sin \omega t \cos \phi + \cos \omega t \sin \phi) \\ &= V_0 I_0 \cos \phi \sin^2 \omega t + V_0 I_0 \sin \phi \sin \omega t \cos \omega t \end{aligned}$$

This is shown in fig. 58.22 (overleaf).

The average power $\langle P \rangle$ over one cycle

$$\begin{aligned} \langle P \rangle &= \langle V_0 I_0 \cos \phi \sin^2 \omega t \rangle + 0 \\ &= \frac{1}{2} V_0 I_0 \cos \phi \\ &= V_{\text{r.m.s.}} I_{\text{r.m.s.}} \cos \phi \quad \left(\frac{V_{\text{r.m.s.}}}{I_{\text{r.m.s.}}} = Z \right) \\ &= I_{\text{r.m.s.}}^2 Z \cos \phi \quad (R = Z \cos \phi) \\ &= I_{\text{r.m.s.}}^2 R \end{aligned}$$

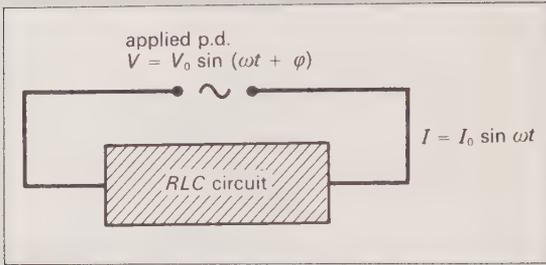


Fig. 58.21. Calculation of power in the RLC series circuit.

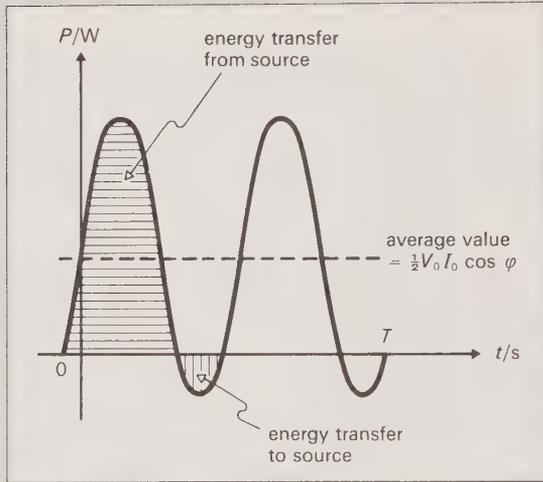


Fig. 58.22. Power of an RLC series circuit.

$$\langle P \rangle = I_{r.m.s.}^2 Z \cos \phi = I_{r.m.s.}^2 R$$

We define the **power factor** of an RLC series circuit to be

$$\frac{\text{average power input}}{V_{r.m.s.} I_{r.m.s.}}$$

Thus the power factor = $\cos \phi = \frac{R}{Z}$

Notes

(a) When $\phi = \pm \pi/2$ rad, the net energy conversion is zero, and the current is said to be **wattless**.

(b) The resistance R is solely responsible for the net conversion of electrical energy, which it does at the rate

$$I_{r.m.s.}^2 R = \left(\frac{I_0}{\sqrt{2}} \right)^2 R$$

(c) A square wave consists of rectangular pulses with a **mark-space ratio** of 1.

$$\left(\text{mark-space ratio} = \frac{\text{duration of pulse}}{\text{time between successive pulses}} \right)$$

This means that the average power delivered by a square wave over one cycle is $I_0^2 R/2$ where I_0 is the current value during a pulse. If negative current pulses of equal amplitude fill the time intervals between positive pulses, the average power is always $I_0^2 R$.

Measurement of a.c. power

(a) For the **ammeter/voltmeter method** measure $V_{r.m.s.}$ and $I_{r.m.s.}$. Then $V_{r.m.s.} I_{r.m.s.}$ is the **apparent power**, and $V_{r.m.s.} I_{r.m.s.} \cos \phi$ the **real power**.

(b) In the **dynamometer wattmeter** (p. 423) the instantaneous couple on the moving coil

$$\propto VI \text{ (instantaneous values)}$$

and is always in the same sense.

The reading is proportional to the **real power** because the power factor $\cos \phi$ is taken into account automatically.

► **58.16 RESONANCE**

(A) RLC series circuit

$$I_0 = \frac{V_0}{Z} = \frac{V_0}{\sqrt{\left[R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2 \right]}}$$

For a fixed value of V_0 , I_0 reaches a maximum value when $\omega L = 1/(\omega C)$ and thus when V and I are **in phase**.

The occurrence of a maximum current at $\omega_0 = 1/\sqrt{LC}$ is known as **current resonance** (cf. *velocity resonance* in a mechanical system).

Charge resonance corresponds to mechanical *amplitude resonance*. The frequency of charge resonance, that of current resonance and the natural frequency of the circuit approach each other as R tends to zero (fig. 58.23, opposite).

(B) RLC parallel circuit

Resonance occurs at the frequency at which the current and potential difference are in phase (fig. 58.24(b)).

In fig. 58.24(a), the current is the same in each branch: one current leads V while the other lags V by the same angle.

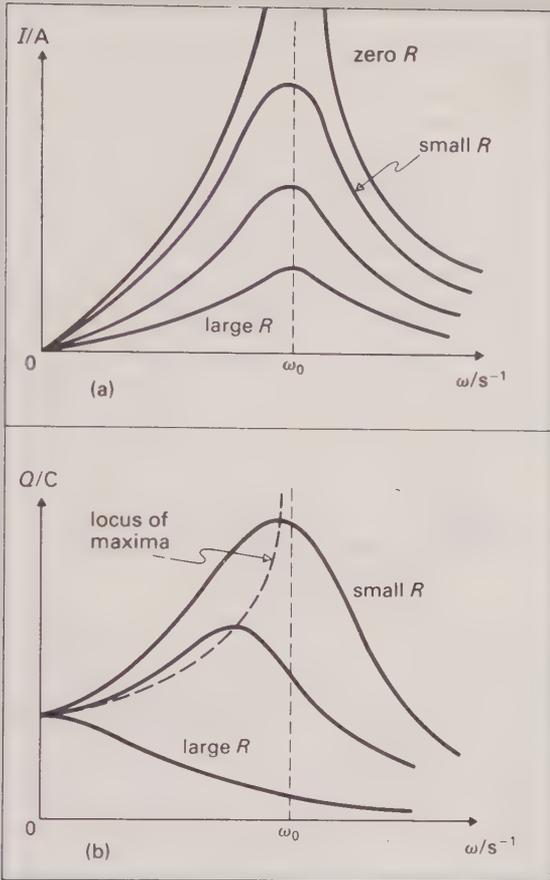


Fig. 58.23. Variation with frequency of (a) current and (b) charge in the RLC series circuit. (Compare with fig. 11.5.)

If X_L and X_C are much greater than R , the phase angle is large and therefore I is very small. Thus the equivalent impedance of the circuit is much greater than that of either branch.

At resonance (fig. 58.25, overleaf), there is maximum equivalent impedance and minimum current.

Use of resonant circuits

(a) The sharp maximum response of a series tuned circuit makes it suitable for selecting a small band of frequencies from a wide range.

(b) The parallel circuit has a sharp maximum impedance at resonance and in situations where the amplification of a circuit depends on an impedance, the parallel tuned circuit gives high amplification over only a small range of frequencies.

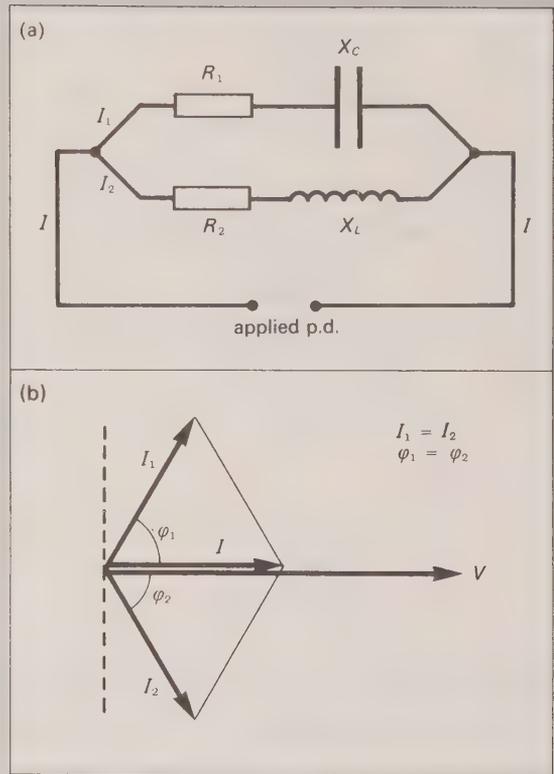


Fig. 58.24. (a) A parallel circuit, and (b) its corresponding vector diagram for resonance.

► 58.17 ELECTRICAL OSCILLATIONS

An electronic oscillator produces an output signal which continuously repeats the same pattern (e.g. sine wave, square wave, saw-tooth, etc.) of current and p.d. variations with time. Applications include

- (a) sine-wave oscillators which provide the carriers for transmission of sound and picture information, and which are also used for reception,
- (b) induction and dielectric heating in industry, and
- (c) timing and triggering in computers and oscilloscopes.

The LC oscillatory circuit

Fig. 58.26 (overleaf) shows the type of circuit used for radio-frequency generation. The circuit tends

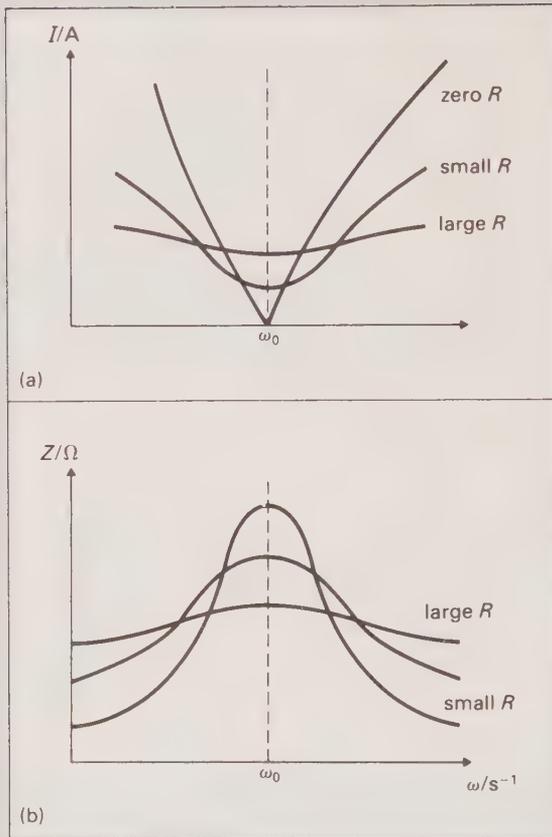


Fig. 58.25. Variation with frequency of (a) current and (b) impedance, in the parallel RLC circuit of fig. 58.24.

to oscillate at its resonant frequency when a pulse of energy is put into it. The energy stored in the electric field of the charged capacitor is converted into energy stored in the magnetic field of the coil. During build-up a back e.m.f. is induced, and this causes the capacitor to be recharged.

When $X_L = X_C$, this oscillatory process occurs at a frequency f_0 , where

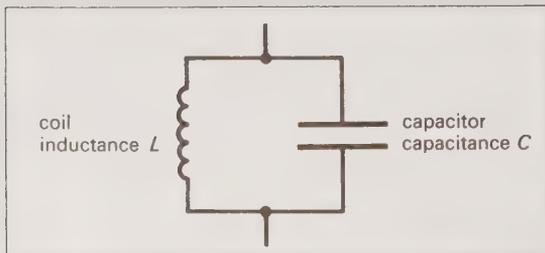


Fig. 58.26. The parallel LC resonant circuit.

$$2\pi f_0 L = \frac{1}{2\pi f_0 C}$$

so

$$f_0 = \frac{1}{2\pi\sqrt{LC}}$$

Notes

- (a) For f_0 to be high, L and C must be small.
- (b) To alter f_0 , either L or C may be varied.

These oscillations will be damped by the resistive losses in the coil and wiring, and by dielectric losses in the capacitor. (Cf. damping by frictional forces, p. 93.)

The oscillations can be maintained by using **feedback** which continuously puts energy into the circuit in phase with its natural oscillations (p. 486).

► 58.18 AN ANALOGUE BETWEEN MECHANICAL AND ELECTRICAL OSCILLATIONS

Analogues are models which utilize some (supposed) similarity between physical processes. They can be misleading if used incautiously, and some give solutions which are qualitative rather than quantitative.

For *direct current* useful analogues exist using concepts of

- (a) *heat flow* (p. 239), and also
- (b) *water flow*.

The latter is particularly useful for an elementary introduction, since the ideas of water pumps, hydrostatic pressure differences and volume-flow rates of (visible) water are easier to visualize than those of sources of e.m.f., electric p.d. and rates of flow of (invisible) electric charge.

A *mechanical* system is found most useful for visualizing *electrical oscillations*, and fig. 58.27 and the table on p. 456 outline one such system. (Other self-consistent systems exist: one, for example, compares F and I rather than F and V .)

Forced damped oscillations

Fig. 58.28 (opposite) introduces two new factors.

(a) A *resistance* in the electrical circuit causes energy dissipation. $R = V/I$ defines R , and has a mechanical analogue $\rho = F/\dot{x}$, in which ρ is a measure of frictional or viscous drag.

(b) An *external agent* (periodic force or alternating p.d.) is required to supply power if the amplitude of the system is to be maintained. Dur-

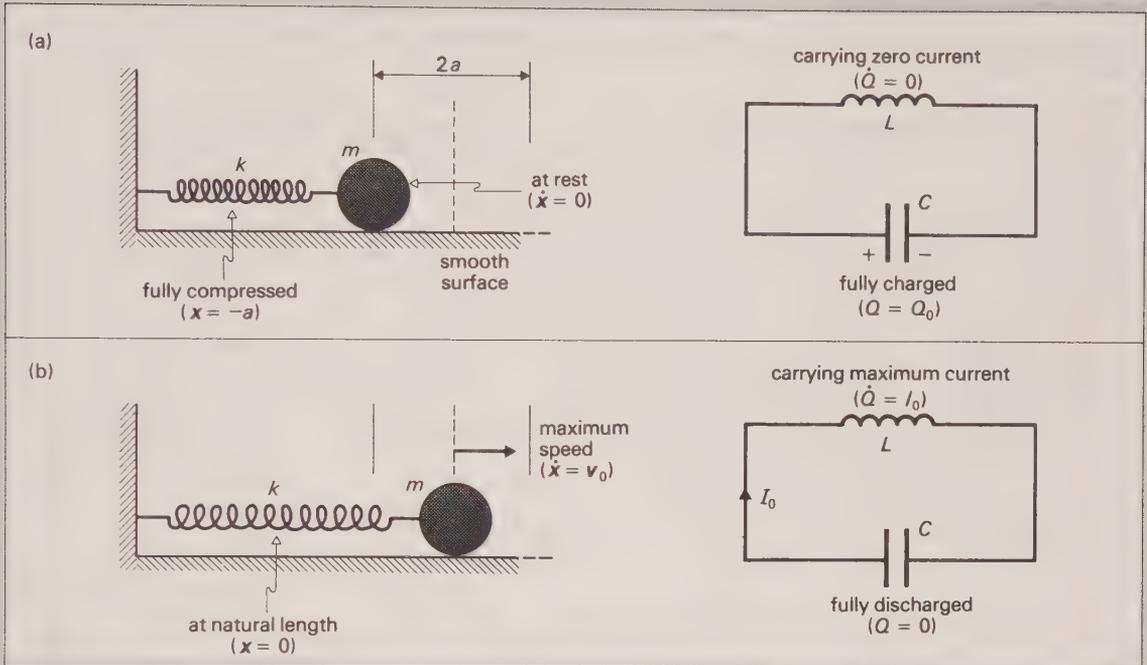


Fig. 58.27. Examples of undamped electrical and mechanical oscillatory systems showing (a) maximum displacement, and (b) zero displacement.

ing each cycle an amount of energy is provided to replace that which has been converted into internal energy.

Each system has a Q -factor, a measure of the selectivity of the resonant system, defined by

$$Q = \frac{2\pi (\text{total energy of system})}{(\text{energy supplied to the system during one cycle})}$$

58.19 RECTIFICATION OF A.C.

The need for a.c.: Power loss in transmission is reduced by using large p.d.s (p. 386). Such p.d.s are obtained using a transformer (p. 431) which needs a *varying* current.

One needs d.c. for

- (a) electronic apparatus (such as the c.r.o. and transistors), and
- (b) some electrolytic processes.

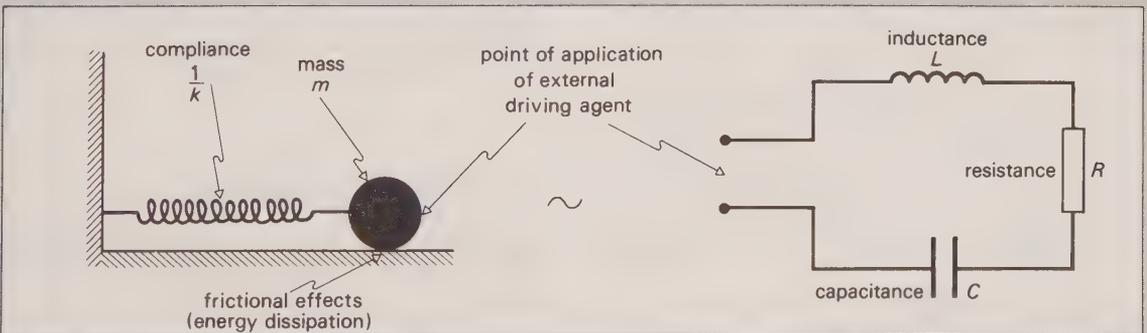


Fig. 58.28. Analogues for forced damped oscillations.

Undamped free oscillations

To compare the two systems shown in fig. 58.27 (p. 455) we can use the analogous quantities and ideas in the table below.

Mechanical		Electrical	
mass	m	L	inductance
displacement	x	Q	charge
velocity	$v = \dot{x}$	$I = \dot{Q}$	current
acceleration	\ddot{x}	$\ddot{Q} = \frac{dI}{dt}$	rate of change of current
force (<i>Newton II</i>)	$F = m\ddot{x}$	$V = L \frac{dI}{dt} = L \frac{d^2Q}{dt^2}$	p.d. (following from definition of L)
elastance (stiffness or spring constant)	$k = F/x$	$1/C = V/Q$	
compliance	$1/k = x/F$	$C = Q/V$	definition of capacitance
elastic p.e.	$\frac{1}{2} kx^2$	$\frac{1}{2} (Q^2/C)$	electric field energy
kinetic energy	$\frac{1}{2} m\dot{x}^2$	$\frac{1}{2} LI^2$	magnetic field energy
natural frequency of oscillation	$\frac{1}{2\pi} \sqrt{\frac{k}{m}}$	$\frac{1}{2\pi} \sqrt{\frac{1}{LC}}$	natural frequency of oscillation

Rectification is the process by which one converts a.c. to d.c. It is achieved by a device which has a characteristic similar to fig. 58.29.

A **rectifier** impedes current flow in one direction far more than in the reverse direction.

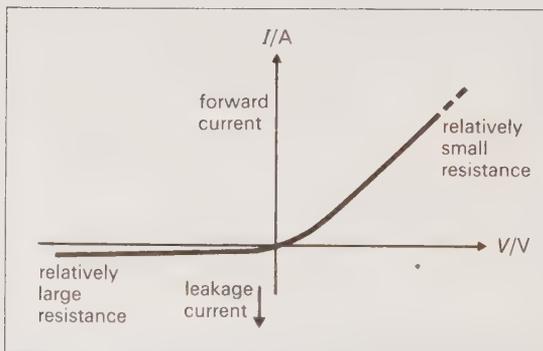


Fig. 58.29. A typical rectifier characteristic.

Examples include

- (a) the metal rectifier, and
- (b) the semiconductor junction diode (p. 476).

Suppose we have a sinusoidal input to a rectifier. Fig. 58.30 (opposite) shows the circuit and output waveform for (a) **half-wave** and (b) **full-wave** rectification. The latter is also given by the *rectifier bridge circuit* of fig. 54.2 on p. 420.

58.20 SMOOTHING CIRCUITS

Rectification gives a *unidirectional* but *pulsating* p.d. or current. To achieve a *steady* p.d. or current we use

- (1) a **reservoir capacitor**, whose action in smoothing the pulsating d.c. is explained by fig. 58.31(a) and (b) (p. 458), together with

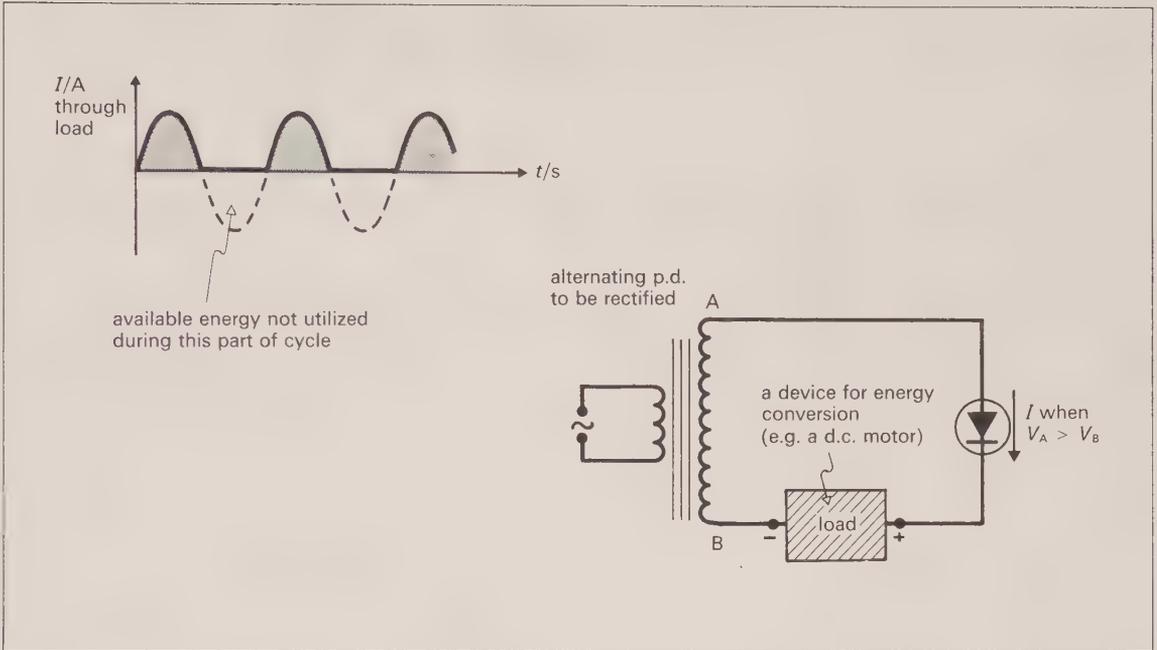


Fig. 58.30(a). Half-wave rectification.

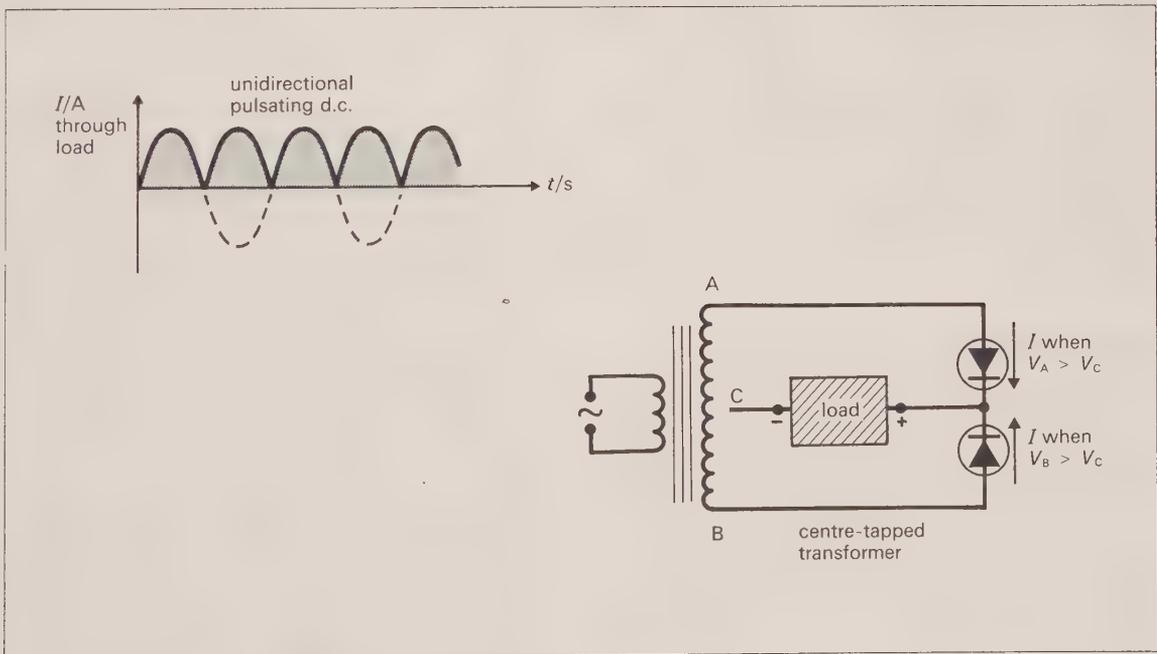


Fig. 58.30(b). Full-wave rectification.

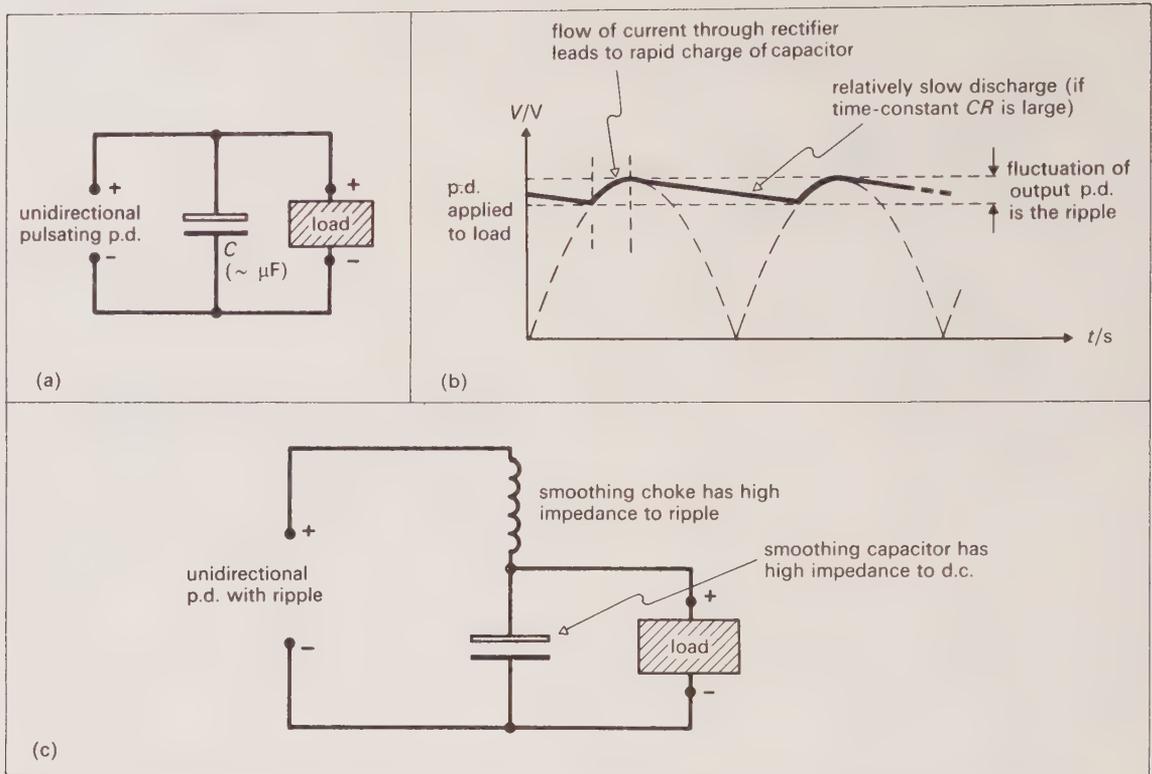


Fig. 58.31. Smoothing circuits: (a) reservoir capacitor, (b) action of the reservoir capacitor, (c) the smoothing filter suppressing the ripple.

(2) a **smoothing filter**: this consists of a *smoothing choke* and *smoothing capacitor* (fig. 58.31(c)), which reduce the effect of the *ripple*.

A **power pack** comprises a rectifier, reservoir capacitor and smoothing filter built into a single unit.

58.21 THE THREE-PHASE SYSTEM

Fig. 56.9 showed a *two-pole single-phase* alternator. If a similar two-pole rotor rotates within three separate sets of stator coils placed symmetrically on pole pieces, *three-phase* alternating current is generated in the *two-pole three-phase* alternator. Each pair of stator coils is connected to a separate distribution line, but all three share a common return wire (the neutral wire in the domestic supply). This three-phase system is used by the National Grid, not only because the alternator

action is smoother and more efficient, but because transmission is more economic for the following reason. The currents in the three circuits have phase differences of 120° and current-time graphs show that, if the three currents are the same size, the instantaneous sum is always zero. Even when the currents are not exactly equal, the resultant will be small and so the common return wire can be narrower than the three main cables. A single-phase system would need six main cables. Industrial installations are usually supplied with all three phases and they would use three-phase transformers and motors. The domestic supply is single-phase and the three currents in the main transmission cables are kept roughly equal by having approximately the same number of houses in a given locality on each phase. If a three-phase supply is required in the home for *synchronous* or *induction* motors, a three-phase alternator is needed or else *phase-splitting* circuits must be devised using capacitors, inductors and resistors.

59 MAGNETIC PROPERTIES OF MATTER

59.1	THE ROWLAND ANCHOR RING	459
59.2	DEFINITIONS	459
59.3	EXPLANATION OF MAGNETISM BY EQUIVALENT SURFACE CURRENTS	460
59.4	MAGNETIZATION CURVES AND HYSTERESIS LOOPS	461
59.5	PLOTTING A HYSTERESIS LOOP	462

▶ 59.6	CLASSIFICATION OF MAGNETIC MATERIALS	463
▶ 59.7	EXPLANATION OF MAGNETIC BEHAVIOUR	463
▶ 59.8	THE MAGNETIC CIRCUIT	464
▶ 59.9	AMPÈRE'S LAW MODIFIED	465

59.1 THE ROWLAND ANCHOR RING

It was explained on p. 429 that if the flux through a search coil of N turns changes by $\Delta\Phi$, the charge Q that flows past a cross-section of the circuit is given by

$$Q = \frac{N \Delta\Phi}{R}$$

where R is the resistance of the circuit.

If we use a given search coil with a particular galvanometer, the throw

$$\begin{aligned} \theta &\propto Q \\ &\propto \Delta\Phi \end{aligned}$$

In fig. 59.1. the toroidal coil is wound onto a **Rowland anchor ring**. Because B is effectively constant over the cross-section of the toroid, the flux Φ through the search coil $\propto B$, the average flux density over a cross-section.

If the switch is opened or closed, then

the ballistic throw $\theta \propto B$, the flux density.

To investigate the magnetic properties of a material, we measure

(a) B_0 , the flux density using a toroid filled with air (which approximates to a vacuum), and

(b) B , the flux density when the toroid is filled by the material.

Materials are called

- (i) **diamagnetic** if $B < B_0$, and
- (ii) **paramagnetic** if $B > B_0$. For some materials $B \gg B_0$; they are called **ferromagnetic**.

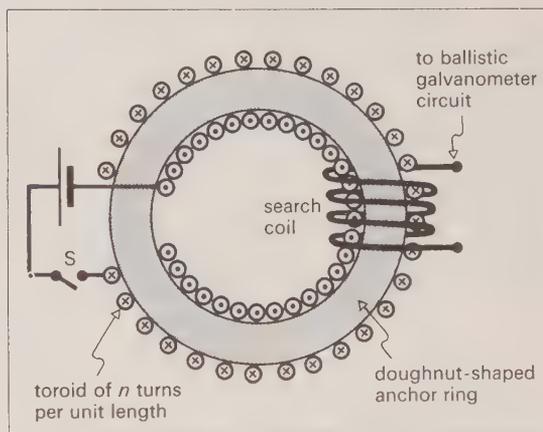


Fig. 59.1. The Rowland anchor ring.

59.2 DEFINITIONS

We write $B = B_0 + B_M$, where

- B = the resultant flux density,
- B_0 = the flux density when the toroid is empty (i.e. that set up by the current alone),
- B_M = the additional flux density set up by the material of the core. (B_M is negative for a diamagnetic material.)

Dividing by B_0

$$\frac{B}{B_0} = 1 + \frac{B_M}{B_0}$$

or

$$\mu_r = 1 + \chi_m$$

(a) We have defined the **relative permeability** μ_r by the equation

$$\mu_r = \frac{B}{B_0}$$

For a ferromagnetic material μ_r is not a constant, but depends on B_0 . It can be considered as the permeability of a material relative to that of a vacuum.

Thus

$$\mu_r = \frac{\mu}{\mu_0}$$

where μ is called the **absolute permeability**. μ_r is a pure number.

(b) We have put $B_M/B_0 = \chi_m$, the **magnetic susceptibility**, although it is not defined in this way in more advanced work. It is a parameter which measures the response of the material to the external magnetic field. χ_m is a pure number.

Since $B_0 = \mu_0 nI$, where I is the current in the toroid, we can measure corresponding values of B

and B_0 by using different values of I , and hence evaluate μ_r for *ferromagnetic materials*. For other materials

$$|\chi_m| < 10^{-3}$$

and the method is not suitable in practice.

59.3 EXPLANATION OF MAGNETISM BY EQUIVALENT SURFACE CURRENTS

In a first study of magnetism, the facts of permanent magnetism are sometimes explained by the concept of the *elementary magnetic dipole*.

We replace that idea here by the concept of an *elementary coil circuit* (fig. 59.2).

The resultant **Amperian surface currents** set up a flux density which is superimposed on B_0 . The equation

$$B = B_0 + B_M$$

has an equivalent in

$$I = I_r + I_M$$

where I = effective current responsible for the resultant B ,

I_r = real current (through the windings) responsible for B_0 , and

I_M = the surface current on the magnetized body responsible for B_M .

Using this idea we explain the effects of a permanent magnetic dipole in terms of surface currents similar to those carried by a solenoid.

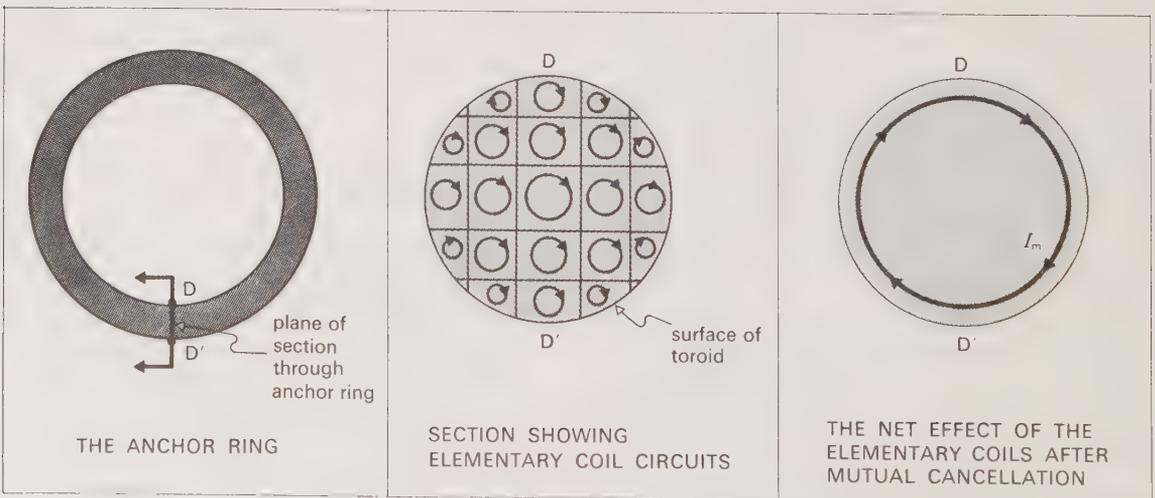


Fig. 59.2. Amperian surface currents.

59.4 MAGNETIZATION CURVES AND HYSTERESIS LOOPS

The curve which shows B plotted against B_0 for a ferromagnetic material is called a **hysteresis loop** (fig. 59.3).

The line ab represents the **magnetization curve**. The different parts of the curve represent the following:

Part of curve	Operation
a to b	increase of toroid current I , using an unmagnetized ferromagnetic core
b to c	I is reduced to zero
c to d	I is reversed in direction, and its magnitude increased to a maximum
d to e	I is again reduced to zero
e to b	I is again increased to its forward maximum

The hysteresis loop shows the following features:

(i) The lack of retracability is called **hysteresis** ('lagging behind').

(ii) At the points c and e the core is magnetized, even though $I = 0$. This exemplifies **permanent magnetism**.

(iii) ac is a measure of the **remanence** or **retentivity** of the material.

(iv) aw shows the reversed value of B_0 required to reduce B to zero after the specimen has been **saturated** in the reverse direction.

$aw \propto$ **coercive force**. Coercive force has the dimensions of B_0/μ_0 .

(v) The area under the loop is proportional to the energy dissipated as internal energy in unit volume of the material when it is taken once round the cycle.

(vi) The loop shows that $\mu_r (= B/B_0)$ is *not* a constant.

(vii) At saturation the hysteresis loop has a gradient of 1, but the different scales for B and B_0 disguise this.

A specimen which is withdrawn along the axis of a solenoid carrying an alternating current is taken through a succession of ever-decreasing hysteresis cycles, and is thus demagnetized.

Requirements of magnetic materials

(a) **Permanent magnets** are made of **hard** magnetic material which has a large remanence (to make the magnet 'strong'), and a large coercive

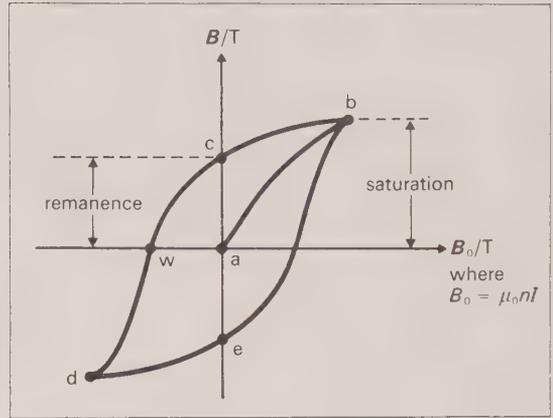


Fig. 59.3. A typical hysteresis loop.

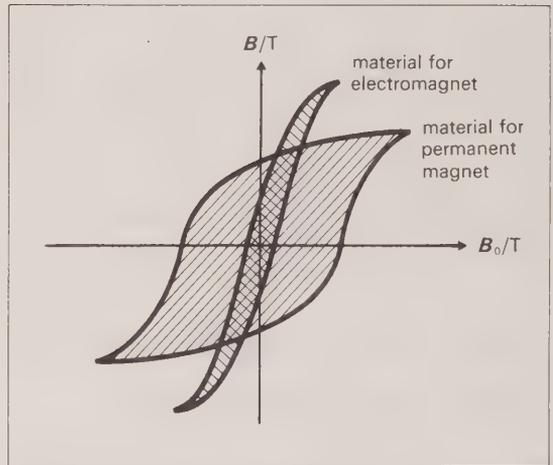


Fig. 59.4. Hysteresis loops for different types of ferromagnetic material. (Note that the scale for B is perhaps one hundred times that for B_0 .)

force (to prevent easy demagnetization by stray fields) (fig. 59.4). Such materials can be made from

(i) **alloys** (e.g. *Ticonal* and *Alcomax*), or
 (ii) **ceramic materials** (e.g. *Magnadur*) which have the mechanical properties of (say) china, the magnetic properties of alloys, and yet high resistivity. (See *ferrites* on p. 430.)

(b) **Transformer cores** have a very narrow loop to prevent overheating by **hysteresis loss**.

(c) An **electromagnet** is made of a **soft** magnetic material which has a small remanence and a small coercive force. It can thus be made a 'strong' but temporary magnet. An **electromagnetic relay** is a switch operated when a relatively small current

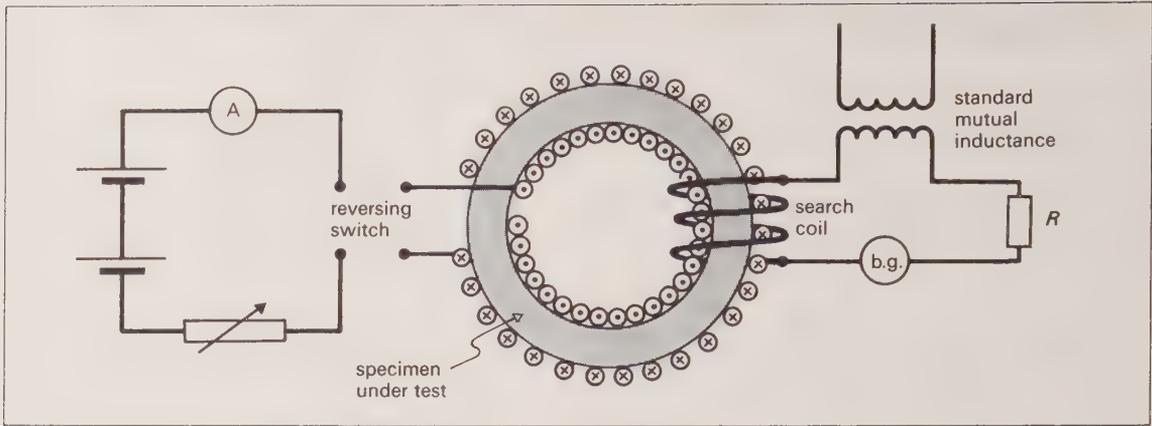


Fig. 59.5. Circuit for using a ballistic galvanometer to plot a hysteresis loop.

flows through a coil wound on a soft-iron core so that a pivoted armature is attracted to the coil. This is used to operate contacts activating another circuit in which a larger current may be required to run a device such as a motor.

(d) An inductor (such as a choke) is often provided with a laminated core of high permeability to increase its self-inductance.

p.d. is applied across the X-plates, and gives a deflection in the x -direction $\propto B_0$.

Notes

- (a) I_{max} must be large enough to produce saturation.
- (b) If the conditions are not changed, the method can be used to compare the loops of different materials.
- (c) This is an example of coordinate representation on the oscilloscope (p. 90).

59.5 PLOTTING A HYSTERESIS LOOP

(A) Using a ballistic galvanometer

Use the circuit of fig. 59.5.

- (1) Using a pre-determined toroid current I , throw the reversing switch, and note the galvanometer throw θ .
- (2) Plot $\theta (\propto B)$ along the y -axis, and $I (\propto B_0)$ along the x -axis.

The procedure is slow and complex because of the necessity of taking the ring through the hysteresis cycle systematically.

(B) Using a cathode ray oscilloscope

This simple quick method can be used with an oscilloscope whose mu-metal screen can be removed to enable an external magnetic field to deflect the electron beam (fig. 59.6).

- (1) The solenoid current I magnetizes the specimen. The B that results causes a deflection of the electron beam in the y -direction (according to the motor rule), which is proportional to B .
- (2) The p.d. across $R \propto I$, which is $\propto B_0$. This

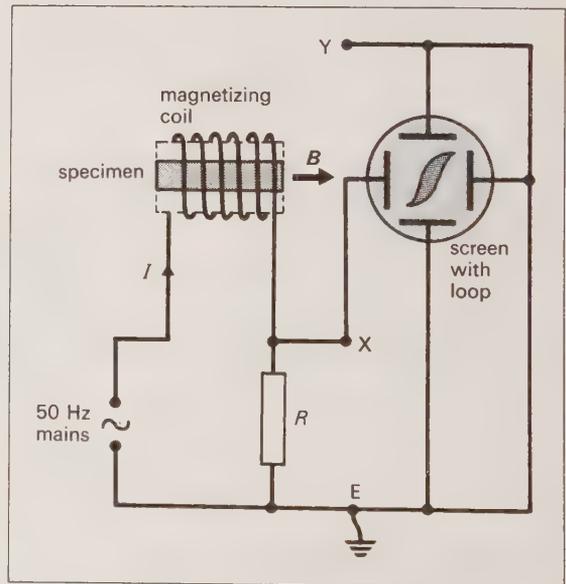


Fig. 59.6. Circuit for using a cathode ray oscilloscope to display a hysteresis loop.

► 59.6 CLASSIFICATION OF MAGNETIC MATERIALS

All substances display some response to a magnetic field.

(A) Diamagnetic substances

(a) When freely suspended in a uniform magnetic field B_0 , they align themselves *across* the field. (B_0 is the flux density *before* the specimen is placed in the field.)

(b) They become magnetized in a direction opposite to that of B_0 . This means that B_M (p. 459) is *negative*, and inside the specimen $B < B_0$.

(c) We draw the field lines as though they avoided passing through the specimen: it experiences a repulsion from the stronger part of the field to the weaker (fig. 59.7(a)).

(d) Bismuth and copper are diamagnetic.

(e) Diamagnetics align themselves parallel to the field if it is non-uniform.

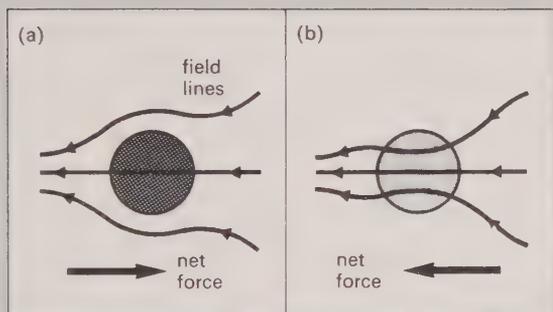


Fig. 59.7. Types of magnetic material: (a) diamagnetic, (b) paramagnetic.

(B) Paramagnetic substances

(a) When freely suspended in a uniform magnetic field B_0 , they align themselves *parallel* to the field.

(b) They become magnetized in the same direction as B_0 . B_M is *positive*.

(c) The field lines are drawn as though being attracted into the specimen (as with a magnetic dipole): it experiences an attraction toward the stronger part of the field (fig. 59.7(b)).

(d) Platinum and aluminium are paramagnetic.

(C) Ferromagnetic substances

These display paramagnetic behaviour, but much more strongly. They may be physically moved to the stronger part of the field.

► 59.7 EXPLANATION OF MAGNETIC BEHAVIOUR

When matter is subject to an external magnetic field, there are two basic effects:

(a) the electronic motion is *distorted*, and

(b) any permanent magnetic dipoles become *partially oriented*.

(a) The **distortion effect** shows the response of the moving charges to the magnetic force. The movement that results from the distortion establishes (according to *Lenz's Law*) a magnetic field in *opposition* to the applied field. The magnetic field in the specimen is *reduced*, corresponding to diamagnetism.

(b) The **orientation effect** will be shown by materials whose atoms have, before the field is applied, an electromagnetic moment. Such atoms include those whose **electron spins** are not paired. The external field B_M supports B_0 , so the field in the specimen is *increased*, corresponding to paramagnetism.

All substances have diamagnetic properties, which are nearly independent of temperature. Except at high temperatures, they may be swamped by paramagnetic properties (if present). Paramagnetism decreases with rising temperature, but some metals show a *weak* paramagnetism which is temperature-independent.

Ferromagnetism

Quantum mechanics explains how, in some substances, the unpaired electron spins become **coupled**, so that $\sim 10^{10}$ atoms may combine to form a **domain** (of linear dimension $\sim 10^{-7}$ m) which has a large electromagnetic moment. When the domains are randomly arranged, the specimen as a whole is unmagnetized.

When an external magnetic field is applied to a specimen, the domains that are suitably oriented grow in preference to others. In soft magnetic specimens the domain boundaries (*Bloch walls*) move easily, and weak fields can produce alignment. In hard specimens the boundaries are more restricted, and greater fields are needed for magnetization.

Evidence for domains

(a) **Bitter photographs**, which show visible surface effects,

Summary of magnetic properties

Effect	χ_m	μ_r	Origin (section 59.7)
diamagnetism	about -10^{-5}	<1	distortion, according to <i>Lenz's Law</i>
paramagnetism	about $+10^{-3}$	>1	orientation of magnetic dipoles in material
ferromagnetism	depend on B_0 , but can be very large ($\sim 10^{+3}$)		magnetic domains

(b) **Barkhausen effect**, in which the jumpy orientation of domains is demonstrated by clicks in a loudspeaker, and

(c) **magnetostriction**, the change of body size on magnetization.

The Curie temperature

When the temperature exceeds the **Curie temperature** the coupling effect disappears, and ferromagnetic materials become paramagnetic.

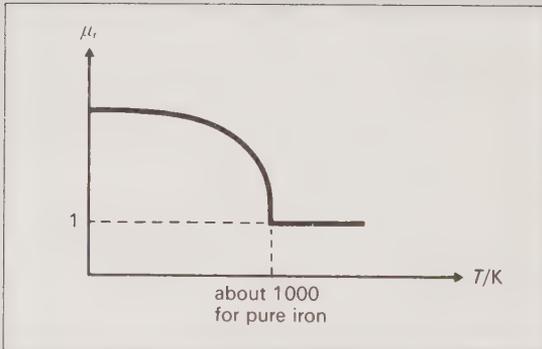


Fig. 59.8. To illustrate the Curie temperature. (μ_r is plotted for a given value of B_0 .)

Fig. 59.8 shows how μ_r decreases as T increases. The Curie temperature depends upon both B_0 , and the material.

► 59.8 THE MAGNETIC CIRCUIT

A magnetic circuit consists of a closed loop of magnetic flux. There is a useful analogy between (i) the electric circuit equation and (ii) a magnetic circuit equation.

(i) (electromotive force) = (current) \times (total circuit resistance)

$$\mathcal{E} = I \times \Sigma \left(\frac{l}{\sigma A} \right)$$

(ii) (magnetomotive force) = (magnetic flux) \times (total circuit reluctance R_m)

$$F_m = \Phi \times \Sigma \left(\frac{l}{\mu A} \right)$$

The **magnetomotive force** is the source of magnetic effects. For example a toroid of N turns each of which carries a current I has an m.m.f. of NI . The unit of m.m.f. is the ampere-(turn). (See below.)

Notes

(a) The concept of a magnetic circuit is not very useful if there is appreciable flux leakage. (Electric charge leakage does not occur in a well-insulated electric circuit.)

(b) The unit of **reluctance** R_m is that of $l/\mu A$, i.e. A Wb^{-1} , which is consistent with that of F_m being A or A-turns .

(c) There are no loose ends to either flux lines or electric current paths. Charged particles flow round an electric circuit whereas nothing flows in a magnetic circuit, although some atoms become reoriented.

(d) μ is a function of I , B and hence Φ in ferromagnetic cores; that is they show non-linear behaviour. On the other hand σ does not vary with I in an ohmic resistor unless the temperature changes. The range of values of σ in various materials is many times greater than that of μ .

(e) If a part of a magnetic circuit does not contain ferromagnetic material then the reluctance of this part is virtually constant, i.e. independent of flux density. (This statement is analogous to *Ohm's law*.)

(f) Flux can be made to pass easily through air and free space, though any air gap effectively controls the total reluctance of a circuit. No mean-

ing can be attached to the terms 'a good magnetic insulator' or 'a magnetic short circuit'.

(g) *Reluctances in series and parallel* can often be combined in the same way as resistances, as long as care is taken to establish the boundaries between different components.

(h) Current density $J = I/A$ appears as an analogy to magnetic flux density $B = \Phi/A$.

(i) When *Ampère's Law* is applied round a closed loop (i.e. a magnetic circuit) which encloses N turns of wire, we obtain

$$\frac{1}{\mu_0} \oint B_0 \cdot dl = \Phi \Sigma \frac{l}{\mu A} = NI$$

From this we deduce that $F_m = NI$, which was implied above.

► 59.9 AMPÈRE'S LAW MODIFIED

(This section is not entirely rigorous. Its aim is to introduce the magnetic field vector H .)

On p. 415 *Ampère's Law* was quoted by the equation

$$\oint B \cdot dl = \mu_0 I$$

which applies in a vacuum.

When magnetic material is present, we write

$$\oint B \cdot dl = \mu_0 (I_r + I_M)$$

in which I_M was introduced on p. 460.

To rewrite the law in a simpler form, we can proceed as follows.

The equation

$$B = B_0 + B_M$$

can be rewritten

$$B = \mu_0 H + \mu_0 M$$

In this equation, note the following:

(a) It is convenient to write $B_0 = \mu_0 H$ in a vacuum. H is called the **magnetic field strength**, and is a *vector* quantity. It has the unit $A \text{ m}^{-1}$. (It is sometimes called the *magnetizing force*.)

(b) M stands for the **magnetization**, which is defined as the electromagnetic moment per unit volume. It is also a *vector*, and has unit

$$\frac{A \text{ m}^2}{\text{m}^3} = A \text{ m}^{-1}$$

Rearrangement of the last equation gives

$$H = \frac{B - \mu_0 M}{\mu_0}$$

which may be regarded as a more formal *defining equation for H*.

For dia- and paramagnetic materials we have the empirical equations

$$B = \mu_r \mu_0 H = \mu H$$

in which μ_r is a constant.

For ferromagnetic materials μ_r is *not* a constant.

H is not a new quantity of fundamental significance, but it enables us to rewrite *Ampère's Law* in a simpler form. It becomes

$$\oint H \cdot dl = I$$

where I represents the sum of the *true currents* (as opposed to *Amperian currents*) enclosed by the loop.

A proper study of the vectors B , H and M is beyond the scope of this book: in this section we merely indicate that H is a useful quantity.

XI ELECTRONIC, ATOMIC AND NUCLEAR PHYSICS

60	THE FREE ELECTRON	468
61	ELECTRONICS	476
62	THE PHOTOELECTRIC EFFECT	503
63	X-RAYS AND THE ATOM	507
64	RADIOACTIVITY	518

60 THE FREE ELECTRON

60.1 SOURCES OF ELECTRONS	468
60.2 PROPERTIES OF CATHODE RAYS	469
60.3 PRINCIPLES OF ELECTRON DEFLECTION	469
60.4 THOMSON'S MEASUREMENT OF THE ELECTRON'S SPECIFIC CHARGE	470

60.5 MILLIKAN'S MEASUREMENT OF THE ELECTRON'S CHARGE	471
60.6 THERMIONIC EMISSION	473
60.7 CONSTRUCTION OF THE CATHODE RAY OSCILLOSCOPE	473
60.8 USES OF THE C.R.O.	475

60.1 SOURCES OF ELECTRONS

Small negatively charged particles can be obtained in the following ways:

(a) By applying a potential difference of a few kilovolts across a gas maintained at a pressure of about 10 Pa. The tube containing the gas is called a **discharge tube**. The particles seem to come from the cathode, and are called **cathode rays** (see below).

(b) By heating a conductor. **Thermions** are emitted from the surface. The process is called **thermionic emission** (p. 473).

(c) By irradiating a metal surface with electromagnetic radiation of appropriate frequency. **Photoelectrons** are emitted from the surface. The process is called **photoelectric emission** (p. 503).

(d) From certain radioactive nuclei. β^- -**particles** are emitted (p. 523). The process is called **β^- -emission**.

(e) By imposing an intense electric field ($\sim 10^8 \text{ V m}^{-1}$) at the surface of a metal. The process is called **field emission**.

(f) By bombarding a metal surface with primary (incident) electrons. The process is called **secondary emission**, and the particles are called **secondary electrons**.

Their identification

The **specific charge** q of a particle is its ratio

$$\frac{\text{charge}}{\text{mass}}$$

and can be measured by experiments like *Thomson's* (p. 470). Such experiments indicate that *all*

these negative particles have the same specific charge, and all are believed to be electrons.

Properties

The electron has the following properties:

(a) It carries a *charge*

$$e = -1.6 \times 10^{-19} \text{ C}$$

(*Millikan's experiment*—p. 471). This is the fundamental quantum of charge.

(b) It has a *specific charge* q given by

$$q = \frac{e}{m_e} = -1.76 \times 10^{11} \text{ C kg}^{-1}$$

(c) It has a *rest mass* $9.1 \times 10^{-31} \text{ kg}$, which can be deduced from e and e/m_e , but which cannot be measured directly.

(d) It is common to all matter.

► Relativistic mass

Because an electron has such a small mass, it is easily accelerated by (electrical) forces to speeds approaching that of light in a vacuum. Its mass m is then given by

$$m = \frac{m_e}{\sqrt{[1 - (v/c)^2]}}$$

where v is the speed of the electron relative to the observer measuring the mass. (An equation of the above form applies to any moving particle, not just an electron.)

Unless otherwise stated, in this book we consider the **rest mass** m_e , rather than the relativistic mass.

60.2 PROPERTIES OF CATHODE RAYS

Although cathode rays were first observed in a discharge tube, it is convenient to demonstrate their properties using thermions, electrons emitted from a heated cathode. See the table below.

Observation	Deduction
(a) They cast a shadow of an obstacle (such as a <i>Maltese Cross</i>) on the wall of the tube.	(a) They travel in nearly straight paths despite the transverse force (their weight) acting on them, and so are moving <i>very fast</i> .
(b) They can, indirectly, transfer their energy so as to set a small paddle wheel into rotation.	(b) They probably consist of fast moving <i>particles</i> , whose k.e. is being converted to rotational energy.
(c) They can pass through aluminium foil.	(c) They must, if charged, have a very small charge (and hence (from e/m_e), a very small (sub-atomic) mass).
(d) They are deflected by a magnetic field in a direction opposite to that in which positive particles with the same velocity would be.	(d) According to <i>Fleming's</i> left-hand motor rule, they are <i>negatively</i> charged.
(e) In an electric field they are deflected towards a positive plate.	(e) This confirms that their charge is negative.
(f) They can be collected in a metal cylinder. An electroscope then shows that the cylinder acquires a negative charge (<i>Perrin's</i> experiment).	(f) This associates them with the negative charge produced in electrostatics experiments. (The rays were deflected into the cylinder by a magnetic field, to check that the effect was not caused by any electromagnetic radiation emitted by the hot cathode.)
(g) They affect a photographic plate.	(g) This confirms that they have energy.

60.3 PRINCIPLES OF ELECTRON DEFLECTION

Care must be taken not to confuse the following situations:

(A) Magnetic field B only

In fig. 60.1(a) overleaf the electron velocity v is perpendicular to B . Hence $F = Bev$, and F is perpendicular to the plane containing B and v .

Since F is perpendicular to v , the speed remains constant, and the path is *circular*.

Using $F = ma$

we have $Bev = \frac{m_e v^2}{r}$

where v^2/r is the electron's centripetal acceleration.

(B) Electric field E only

In fig. 60.1(b), the electron moves in a *parabolic* path. Refer to p. 341.

(C) E and B perpendicular (crossed)

Using the fields shown in fig. 60.1, the magnetic force is in the positive y -direction, and the electric force in the negative y -direction. Their magnitudes can be made equal, so that the trajectory becomes a straight line.

Then $Bev = eE$

so $v = \frac{E}{B}$

$$v \left[\frac{\text{m}}{\text{s}} \right] = \frac{E [\text{N/C}]}{B [\text{N s/C m}]}$$

The method is used to *calculate the speed of the electrons* in a beam. It depends for its action on the

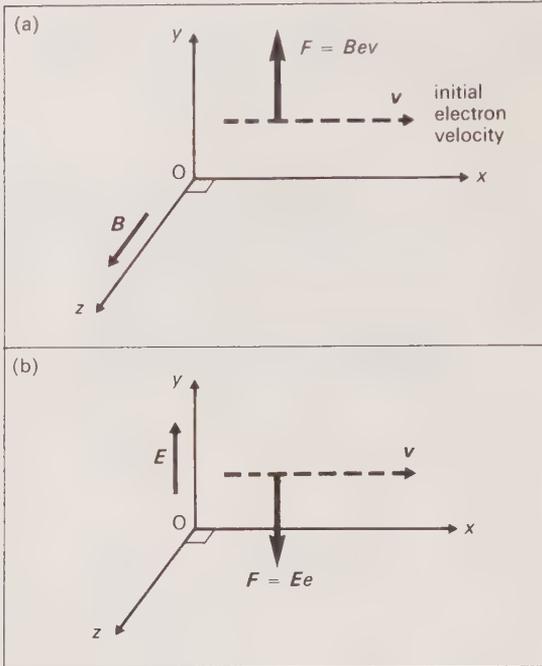


Fig. 60.1. The initial force that acts on an electron moving through uniform fields of B and E .

fact that the electric force is independent of the electron speed, while the magnetic force is proportional to that speed.

Application

Measurement of (e/m_e) for electrons.

(D) E and B parallel

Suppose we have a situation in which E and B are both parallel to Oy . Then the electric and magnetic forces are perpendicular, and the electrons in the beam experience simultaneous deflections

- (a) along yO (electric) and
- (b) along zO (magnetic).

A screen placed some distance along Ox , and parallel to the $y-z$ plane would show a parabolic trace if the electrons were travelling at different velocities.

Application

Measurement of specific charge q for positive ions (p. 509).

60.4 THOMSON'S MEASUREMENT OF THE ELECTRON'S SPECIFIC CHARGE

Fig. 60.2 shows a modernized version of Thomson's apparatus.

Procedure

- (1) With E and B both zero, note position O of the spot on the screen.
- (2) Using a fixed value of B (alone), deflect the spot to a measured position S .

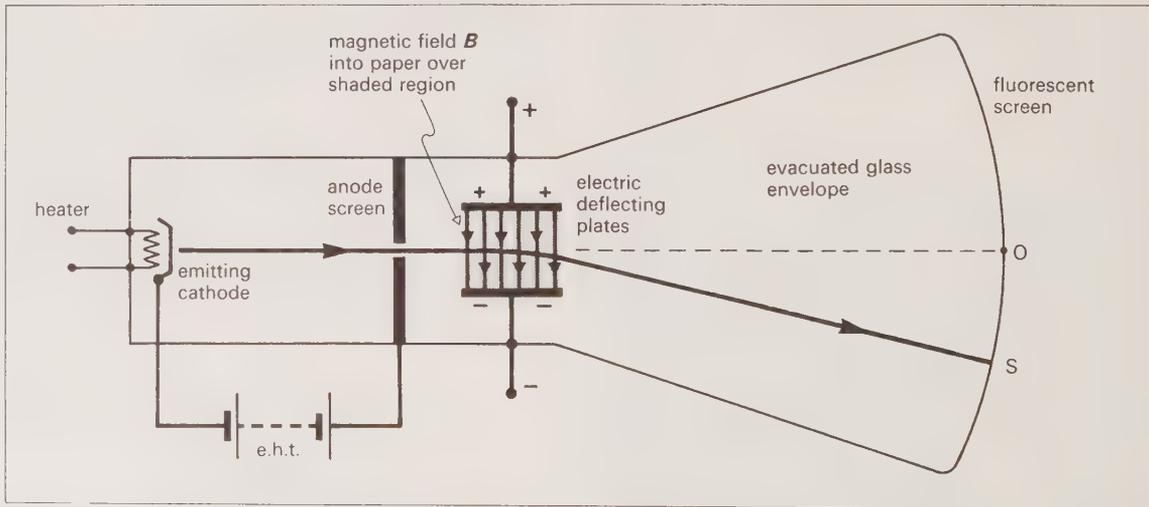


Fig. 60.2. A heated cathode method for measuring e/m_e .

(3) Adjust the value of E so that the spot returns to its undeflected position O .

B can be established using a *Helmholtz coil* arrangement (p. 413).

Calculation

$$Bev = \frac{m_e v^2}{r} \quad (\text{p. 402})$$

so
$$\frac{e}{m_e} = \frac{v}{Br}$$

in which

- (a) v is found from $v = E/B$ (p. 469), and
- (b) r is found as follows.

Measurement of r

From fig. 60.3

- (a) By similar triangles

$$\frac{d}{D} = \frac{l/2}{L}$$

- (b) By the intersecting chords theorem (p. 23)

$$d(2r - d) = l^2$$

d is calculated from (a), and substitution into (b) gives r .

Notes

(a) *Thomson's result* (1897) was 30% inaccurate because not all the electrons travelled at the same speed, and because the exact extent of the fields

was not well defined. *Dunnington's method* (1933) is more accurate.

(b) The experiment

(i) confirms that the cathode ray speed is less than that of electromagnetic waves (i.e. that the rays, if waves, are not *electromagnetic waves*),

(ii) gives a value for e/m_e about 2000 times that found for H^+ from electrolysis.

Millikan's oil-drop measurement of the elementary charge (below) confirms that this is because the electron has a smaller mass (rather than a larger charge).

60.5 MILLIKAN'S MEASUREMENT OF THE ELECTRON'S CHARGE

Procedure

Refer to fig. 60.4 (overleaf).

(1) Earth both plates, and observe the fall of a particular drop under gravity. It acquires a steady downward terminal velocity v_d .

(2) Charge the drop by a burst of X-rays. Apply an electric field E by connecting an e.h.t. battery across the plates. Suppose that the same drop, because it experiences an electric force upwards, acquires a new (upward) terminal velocity v_u .

Calculation

(1) Since the drop is in equilibrium, the resultant force acting on it is zero.

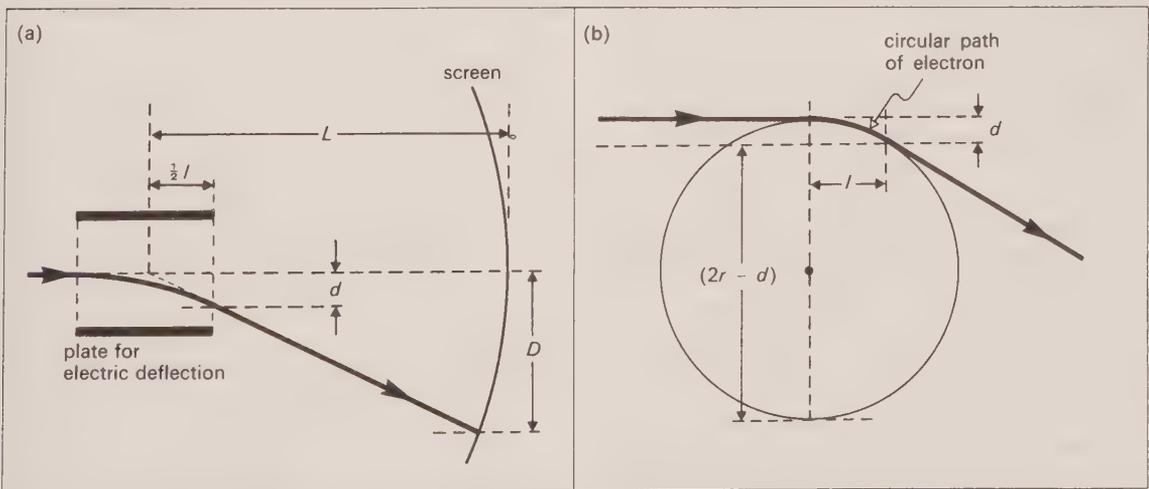


Fig. 60.3. Geometry for measurement of r , radius of circular path caused by magnetic field. (a) Dimensions of the apparatus, and (b) the electron path.

whole number; no fraction of e has ever been observed.

60.6 THERMIONIC EMISSION

Electrons are free to migrate within the boundary of a metal, but if an electron tries to leave the surface, it experiences an attractive force from the resultant positive charge it leaves behind.

When an electron is removed from a metal, an external agent does work against that attractive force.

The **work function energy** W is the least energy that must be supplied to remove an electron from the surface of the metal.

The **work function potential** Φ is defined by

$$W = \Phi e$$

where e is the magnitude of the electronic charge. (Refer also to the photoelectric effect, p. 505.)

Examples

Metal	Φ/V	$W/10^{-19} \text{ J}$
caesium	1.8	2.9
nickel	5.0	8.0
tungsten	4.5	7.2

The low value of Φ (and hence W) for caesium explains why it is used in the photoelectric emission cell (p. 506).

Space charge

When a metal is heated, its electrons may acquire sufficient thermal energy to leave the surface. This process is **thermionic emission**. The greater the temperature, the greater the rate of emission from unit area of surface. There is a good analogy with *latent heat* and *evaporation*.

As with evaporation, a dynamic equilibrium is attained when the region around the metal (the **space charge**) returns electrons to the surface as fast as they leave it.

An **emission current** flows when a positive plate (an **anode**) withdraws electrons from the space charge. The metal then becomes a **cathode**. This is the principle of the **thermionic diode valve**.

Types of cathode

Because thermionic emission occurs close to a metal's m.p., only some metals are suitable.

(a) A thoriated tungsten filament has $\Phi = 2.6 \text{ V}$, and emits copiously at 1900 K.

(b) A metal coated with barium oxide has $\Phi = 1 \text{ V}$, and can be operated at about 1300 K.

(c) As heating is frequently achieved by using a.c., the **indirectly heated cathode** is used: it has the advantage that its surface is at a *uniform potential*.

60.7 CONSTRUCTION OF THE CATHODE RAY OSCILLOSCOPE

Refer to fig. 60.6 (overleaf).

The oscilloscope has three basic components.

(A) The electron gun

This consists of

(a) the indirectly heated *cathode* to emit electrons,

(b) the *grid*, whose potential relative to the cathode controls the electric field at the cathode, and so the number of electrons passing through the grid per second: the grid is the *brightness control*,

(c) the *electron lens*, made up of

(i) a *focussing anode*, and

(ii) an *accelerating anode*.

The potential of the focussing anode is altered by the *focus control*.

Typical potentials for the gun of a small tube might be

(a) cathode -800 V ,

(b) grid -800 V to -850 V ,

(c) (i) focussing anode -500 V to -600 V ,

(ii) accelerating anode zero (earthed).

The figures represent p.d. relative to the accelerating anode.

The change of k.e. of an accelerated electron can be found from

$$eV_{CA} = \frac{1}{2}m_e v^2 - \frac{1}{2}m_e u^2$$

where it is accelerated from a speed u to speed v because of the p.d. V_{CA} between anode and cathode. Usually $\frac{1}{2}m_e u^2 \ll \frac{1}{2}m_e v^2$, and can be considered negligible.

A mu-metal screen round the tube protects it from stray magnetic fields. The accelerating anode and the graphite coating are at earth potential so that earthed objects and people near the screen do not affect the electron beam.

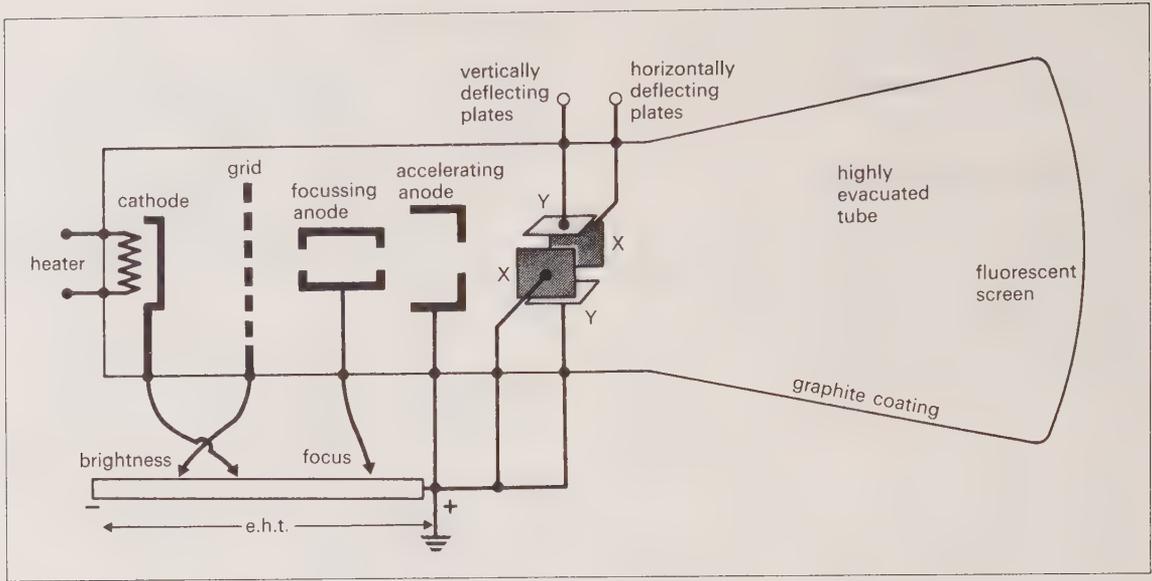


Fig. 60.6. The basic elements of a cathode ray oscilloscope.

(B) The deflecting system

This consists of two pairs of parallel plates arranged so that their respective electric fields produce horizontal and vertical deflection of the electron beam.

(a) The *X-shift* and *Y-shift* controls change the undisplaced location of the spot on the screen.

(b) The *sensitivity* control changes the displacement produced by a given change of p.d. applied to the input (p. 341).

One of each pair of the deflecting plates is at the same (earth) potential as the accelerating anode.

(C) The display system

This consists of a screen coated with a phosphor such as zinc sulphide, whose molecules absorb energy from the incident electron beam, and re-radiate it as visible light. The re-radiation is

- (a) *fluorescence* if there is no afterglow, and
- (b) *phosphorescence* if the effect continues for a short while after impact.

The screen potential remains constant because the electrons return to earth through the internal graphite coating.

The time-base

A time-base consists of a system which applies a saw-tooth p.d. of the type in fig. 60.7 across the X-plates.

Along AB the spot traverses the screen at a steady finite speed. Along BC it flies back (almost instantaneously) to the starting point.

The time period T can be changed by coarse and fine controls on the front of the c.r.o.

The time-base is frequently *synchronized* with the signal applied to the Y-plates so that each linear sweep starts at the same point on the input waveform thereby giving a steady trace. The trace is *suppressed* during the flyback part of the saw-tooth p.d.

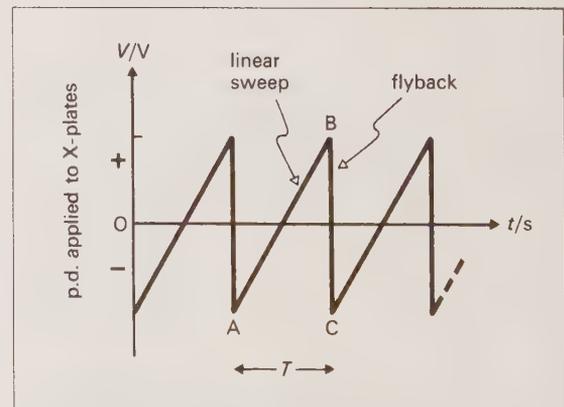


Fig. 60.7. The ideal (theoretical) time-base p.d.

60.8 USES OF THE C.R.O.

The c.r.o. is most versatile. The following are among the uses to which it is put.

(A) Use as a voltmeter

(a) D.C. Connect the p.d. to be measured across the X- or Y-plates, without using the time-base. The spot deflection can be calibrated in mm/V by using a known p.d. The c.r.o. has a very high impedance, and is not damaged by overloading (which merely causes the spot to be deflected off the screen).

(b) A.C. Connect the p.d. to be measured as in (a). The spot deflection represents $2 \times (\text{peak voltage})$, because the inertia of the electron beam is negligible.

The c.r.o. can be used as a voltmeter over a wide range of frequencies. (Cf. the moving-coil galvanometer which reads zero at high frequency.)

A d.c. or a.c. *current* may be calculated from the p.d. across a resistor of *standard* resistance.

(B) The display of waveforms

Connect the signal to be displayed to the Y-plates. Fig. 60.8 gives two examples. The fly-back trace is usually suppressed.

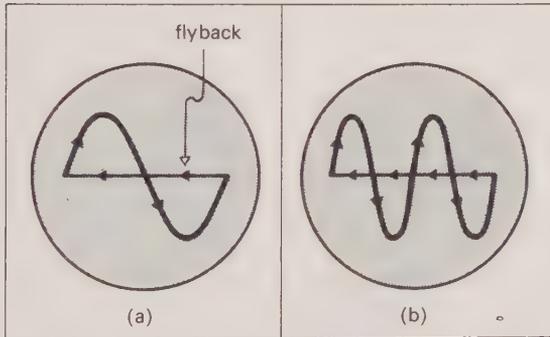


Fig. 60.8. British a.c. mains waveform (50 Hz).

(a) Time-base frequency 50 Hz,

(b) time-base frequency 25 Hz.

This method can be used to display

- (a) half-wave rectification (p. 457),
- (b) full-wave rectification (p. 457),
- (c) the effects of smoothing and filtering (p. 458),
- (d) the waveforms of musical notes.

The c.r.o. is an example of an *analogue system* in that the varying p.d. across the Y-plates can represent other physical quantities such as loudness of sound, blood pressure, magnetic field strength. The *linear* time-base means that the c.r.o. can be regarded as a *graph plotter* to display the variation of a quantity with time, e.g. the charge on the plates of a discharging capacitor. Screens which continue to glow after the electron beam has passed are very effective for this purpose.

(C) Measurement of time intervals

A calibrated time-base is marked with a number of *sweep speeds* so that the horizontal separation (measured in mm or number of divisions) of two signals on the screen can be used to calculate a time interval. This is particularly useful for the measurement of very small time intervals such as the time for an electromagnetic wave pulse to travel along a coaxial cable.

(D) The display of phase relationships

Suppose V_x and V_y are sinusoidal p.d.s of the same frequency. Then if they are of equal amplitude, and connected to the X- and Y-plates, one would see

- (a) a straight line at 45° to the axis if they are in phase,
- (b) a circle if the phase difference is $\pi/2$ rad,
- (c) an ellipse for other phase differences.

The method is useful for demonstrating phase relationships in a.c. circuits (p. 448), and the theory is discussed on p. 103.

Double-beam oscilloscopes have two electron beams and two pairs of Y-plates so that two alternating p.d.s can be observed on the screen at the same time.

(E) To compare frequencies

Suppose the two p.d.s whose frequencies are to be compared are connected across the deflecting plates.

When the frequency ratio is appropriate a steady pattern (called a *Lissajous figure*) appears on the screen. The form of the pattern enables the ratio to be calculated (see p. 90).

(F) To display hysteresis loops

(See p. 462.)

61 ELECTRONICS

61.1 THE PRINCIPLE OF THE p-n JUNCTION DIODE	476	61.10 THE TRANSISTOR AS AN AMPLIFIER	485
61.2 CHARACTERISTIC OF THE p-n JUNCTION DIODE	478	61.11 THE TRANSISTOR AS AN OSCILLATOR	486
61.3 INTRODUCTION TO TRANSISTORS	479	61.12 THE TRANSISTOR AS A SWITCH	487
61.4 THE JUNCTION GATE FIELD EFFECT TRANSISTOR (J.F.E.T.)	481	61.13 LOGIC GATES	487
61.5 CHARACTERISTICS OF n-CHANNEL J.F.E.T.S	481	61.14 CONTROL SYSTEMS	489
61.6 THE INSULATED GATE FIELD EFFECT TRANSISTOR (I.G.F.E.T.)	482	61.15 BINARY ADDITION	490
61.7 CHARACTERISTICS OF p-CHANNEL I.G.F.E.T.S	483	61.16 PULSES AND MULTIVIBRATORS	491
61.8 THE BIPOLAR OR JUNCTION TRANSISTOR (B.J.T.)	483	61.17 BINARY COUNTING	493
61.9 CHARACTERISTICS OF THE JUNCTION TRANSISTOR	484	61.18 THE OPERATIONAL AMPLIFIER	495
		61.19 ANALOGUE AND DIGITAL ELECTRONICS	500
		61.20 BRITISH STANDARD SYMBOLS	502

61.1 THE PRINCIPLE OF THE p-n JUNCTION DIODE

The whole p-n junction (fig. 61.1) must be part of a single continuous crystal of semiconductor material to avoid the high rates of recombination that would otherwise occur at the junction of two separate crystals. For protection the junction diode is enclosed in a light-tight capsule.

Its action

(A) *Zero bias (no applied field)* (fig. 61.2)

Notes

(a) The thermal diffusion of holes into the n-region and electrons into the p-region causes a net migration of charge across the boundary or junction.

(b) The resulting p.d. is called the **contact**

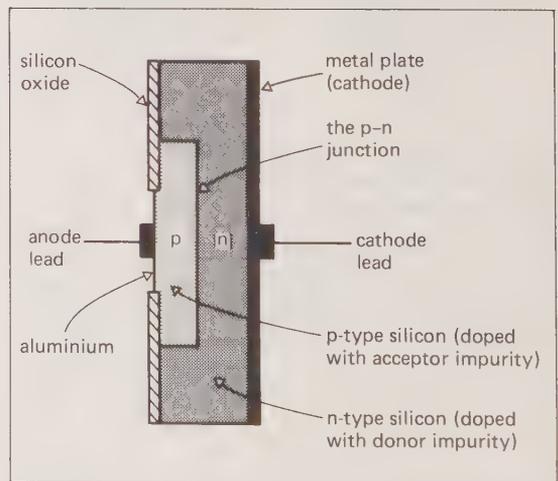


Fig. 61.1. Construction of a silicon planar diode.

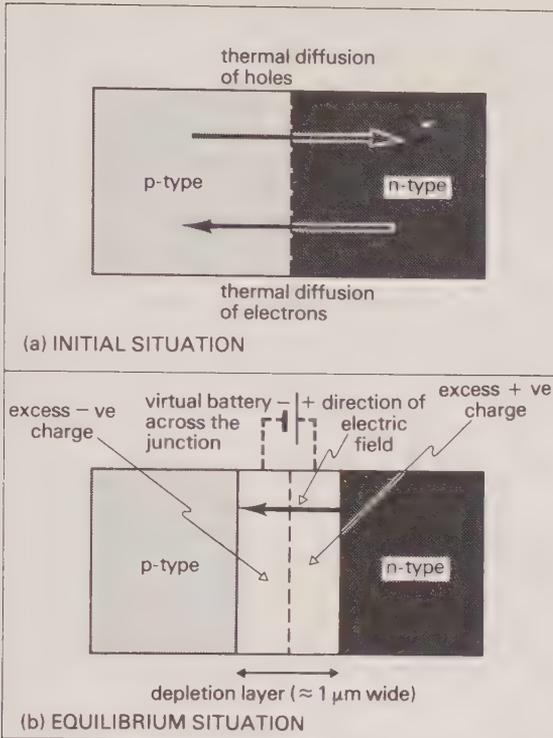


Fig. 61.2. The junction diode with zero bias.

potential, and represents a *barrier* of a fraction of a volt. The resulting field tends to prevent further diffusion of charge-carriers.

(c) The height of the barrier (the p.d.) is determined by the width of the **depletion layer** or **transition region**. In this region the total carrier number density is less than one thousandth of its value in either the p- or n-regions. This depletion is not due to recombination. It is because the sum of the densities of the holes and electrons is least when the densities of each are the same.

(d) The size of the contact potential decreases as the temperature rises. It also decreases with increasing density of doping impurities.

(e) No current would flow if the two regions were connected by a wire, since equal and opposite contact potentials would be set up at the two metal/semiconductor contacts.

(f) The p-n junction behaves like a capacitor of capacitance equal to

(permittivity) \times (junction area) \div (junction width)

(B) Reverse bias (a field applied) (fig. 61.3)

Notes

(a) Reverse bias means that the p-region is made negative w.r.t. the n-region. The zero-bias

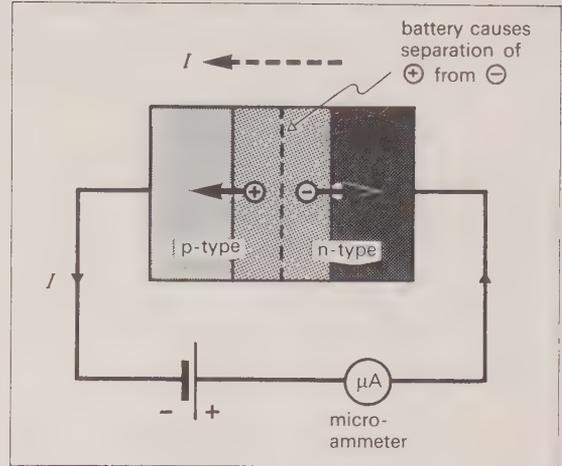


Fig. 61.3. The junction diode with reverse bias.

contact potential across the boundary is increased (approximately in proportion to the square root of the applied p.d.) and the flow of majority carriers is further inhibited.

(b) The electron number density in the p-region and the hole density in the n-region are both greatly reduced. Therefore the minority-carrier densities are also reduced on either side of the depletion layer.

(c) The typical current flow for a small diode is a few μA . It is due to the *minority* carriers being pushed across the junction, and constitutes a **leakage current**.

(d) As the junction width is increased, the capacitance is decreased. A **point-contact diode** has a very small area, resulting in a very small capacitance, which means that it has high impedance to a.c. in the reverse direction ($X_c = 1/2\pi fC$). It is therefore suitable for microwave applications such as high-frequency radio signal detection, which is a process similar to rectification. Germanium is preferred to silicon because lower radio signal p.d.s start it conducting in the forward direction.

(C) Forward bias (fig. 61.4, overleaf)

Notes

(a) Forward bias means that the p-region is made positive relative to the n-region. The zero-bias p.d. across the junction is overcome. Charge-carriers of both signs can move and recombine at the junction. The width of the depletion layer is reduced.

(b) The minority-carrier densities have increased relative to their equilibrium values on either side of the depletion layer.

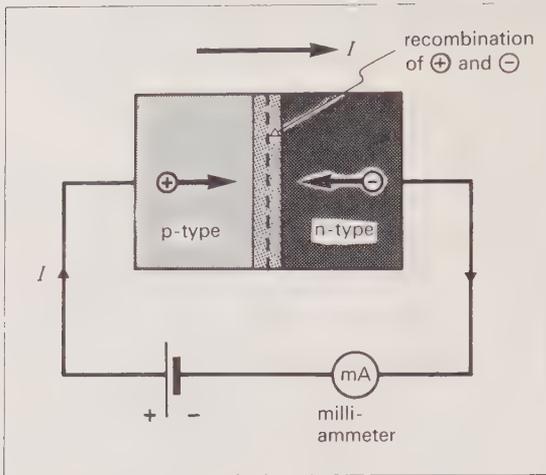


Fig. 61.4. The junction diode with forward bias.

(c) The typical current flow for a small diode is a few mA, but can be made much larger. It is due to the *majority* carriers being pushed across the junction.

(d) The circuit symbol for the junction diode is shown in fig. 61.5.

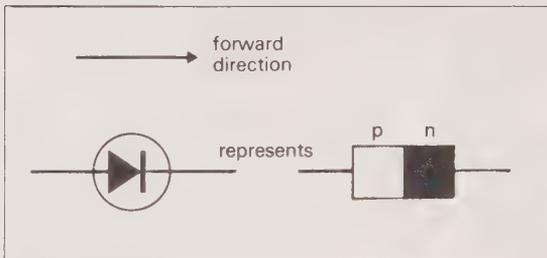


Fig. 61.5. The circuit symbol for the junction diode.

61.2 CHARACTERISTIC OF THE p-n JUNCTION DIODE

Circuits (fig. 61.6)

Notes

(a) The graphs of fig. 61.7 hold provided the temperature is kept below about 100°C for germanium and 200°C for silicon. If either the current or the surrounding air temperature causes the temperature of the junction diode to increase beyond these values, significantly more charge-carriers are generated. This increases the current thereby producing further heating, and the effect

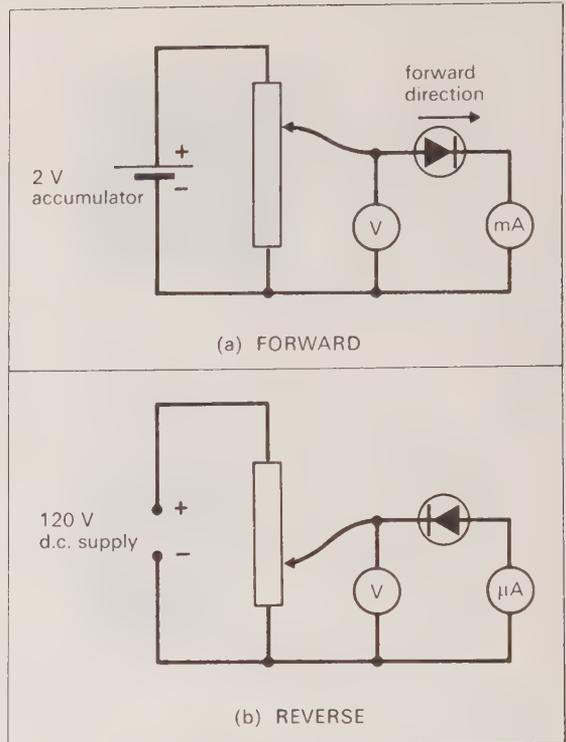


Fig. 61.6. Circuits for plotting junction diode characteristics.

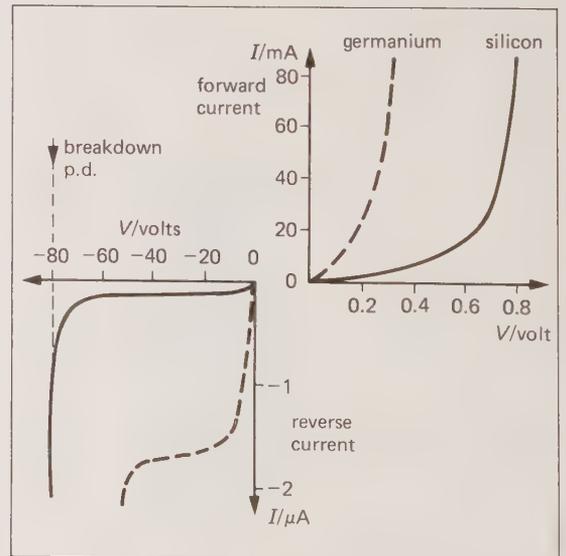


Fig. 61.7. The characteristic of the junction diode. (Note the different current scales.)

escalates until the diode is destroyed by **thermal runaway**. High-current diodes are protected by using **heat sinks** of appropriate **thermal resistance** (p. 238). These possess aluminium cooling fins to dissipate thermal energy at the required rate.

(b) The leakage current due to minority carriers remains small for quite a large range of reverse p.d.s.

(c) The junction diode can be used for *rectification* of alternating current because the diode is alternately on forward and reverse bias. Practical rectifiers are made from silicon so that their leakage currents are very small, which means that their conversion from a.c. to d.c. is efficient. Circuits for half-wave and full-wave rectification are shown in fig. 54.2 (p. 420) and fig. 58.30 (p. 457).

(d) The device has a *maximum peak inverse voltage* (higher for silicon than for germanium): a greater reverse electric field breaks down the insulation of the depletion layer. This leads to a rapid increase in current and may result in damage to the crystal structure.

(i) The **Zener effect** occurs when the applied electric field pulls electrons directly out of their bonds (i.e. from the valence band to the conduction band), giving a sharply increased reverse current.

(ii) The **avalanche effect** occurs when high-energy electrons passing through the crystal release further electrons from the bonds by collisions. These extra electrons in turn dislodge others causing an *avalanche*.

Some specialized junction diodes

(a) Zener diode (or voltage-regulator diode)

This is a silicon junction diode which maintains a steady p.d. from a power supply whose p.d. may change. It is important because many electronic circuits require constant p.d.s. Damage due to over-heating caused by the Zener-effect current is avoided by connecting a resistor in series with a Zener diode (fig. 61.8). This limits the maximum current and leads to a constant p.d. across the diode for a wide range of reverse currents. The impurity concentration near the junction can be adjusted to provide a well-defined *breakdown* (or *Zener*) p.d., so that a varying supply p.d. can be *stabilized* at the specified value. The Zener diode *regulates* the supply p.d. as long as V is greater than V_Z .

(b) Photodiode

This diode consists of a reverse-biased p-n junction with a transparent window, and it produces a significant *photocurrent* (p/current) when illuminated. When the depletion layer is exposed to light of suitable frequency, electron-hole pairs are

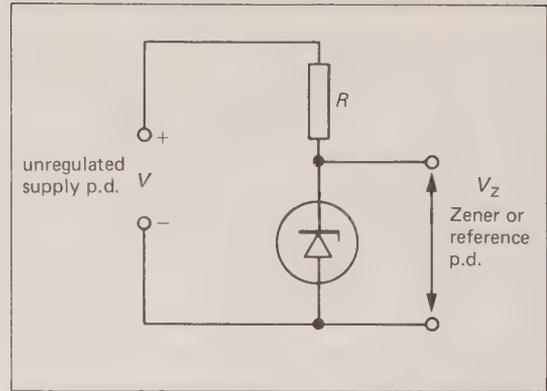


Fig. 61.8. Regulation of a supply p.d.

produced. These extra charge-carriers produce a photocurrent which is superimposed on the reverse current and which is proportional to the light intensity falling on the junction. This effect is used in counters to 'read' holes in punched cards and tapes, in photometers to measure light intensity, and in receivers in fibre-optics systems. In a solid-state particle detector, ionizing radiation of various types can create the extra electron-hole pairs in the depletion layer (p. 476).

(c) Light-emitting diode (LED)

This diode emits light as a result of the recombination of excess electron-hole pairs in the depletion layer of a forward-biased p-n junction. Each recombination releases a quantum of energy, and these quanta are usually absorbed as internal energy. In materials such as gallium arsenide phosphide, many of these quanta are emitted as light, which is released from the LED because the p-n junction is formed very near to the surface. The colour (*frequency*) of the light is determined by the *band-gap energy* and therefore by the exact composition of the material used. At present, red, yellow and green LEDs are available. They are used in electronic indicator lamps and in optical fibre transmitters. Their use in calculator and digital watch displays has been taken over by LCDs (**liquid crystal displays**). A liquid crystal is an organic substance which looks like a liquid but which experiences some degree of order in the molecular arrangement when a p.d. is applied across it. This alters the light-reflecting properties of the liquid crystal so that it appears dark wherever the p.d. is applied. The numbers show up as dark segments against a bright background. This display needs very little electrical energy to operate it but it must be illuminated by daylight or artificial light.

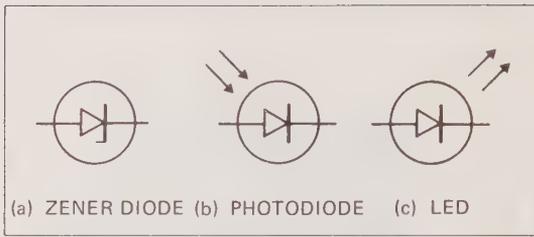


Fig. 61.9. Some specialized diode symbols.

The circuit symbols for the above specialized diodes are shown in fig. 61.9.

61.3 INTRODUCTION TO TRANSISTORS

Their uses

Both field effect and bipolar junction transistors are described in this chapter and, although the student can often regard them as alternatives, their particular properties (compared in the table on the right) must be considered in practical applications.

For the general-purpose transmission of electrical signals, bipolar transistors are preferred because of their greater amplification, higher frequency response and higher power capability. Unipolar transistors are often more suitable in voice-frequency amplifiers with their very small control current and in certain radio receivers where they can reduce interference and signal distortion. In logic or computing circuits the choice is not obvious and requires detailed investigation: unipolar types involve only one diffusion to make both source and drain, and are popular in integrated circuits in computer memories; whereas bipolar types are frequently more appropriate

Transistor type	f.e.t. (unipolar)	b.j.t. (bipolar)
Input impedance/ Ω	10^9 to 10^{18}	1 to 10^6
Effect of incident radiation	negligible	damage caused
Effect of high temperatures	negligible	damage caused
Quantity amplified directly	p.d.	current

owing to their higher amplification and lower operating power requirement.

The distinction between them

(a) In **unipolar** transistors the current is carried by the drift of majority carriers of a single type, and is controlled by the electric fields within the semi-conducting materials. Types include

- (i) the junction gate field effect transistor (j.f.e.t.) (61.4), and
- (ii) the insulated gate field effect transistor (i.g.f.e.t.) (61.6).

Unipolar transistors are often referred to as *majority-carrier devices*.

(b) In **bipolar** transistors (such as the junction transistor b.j.t. of 61.8) the current is carried by the drift of both electrons and positive holes; they are often referred to as *minority-carrier devices*.

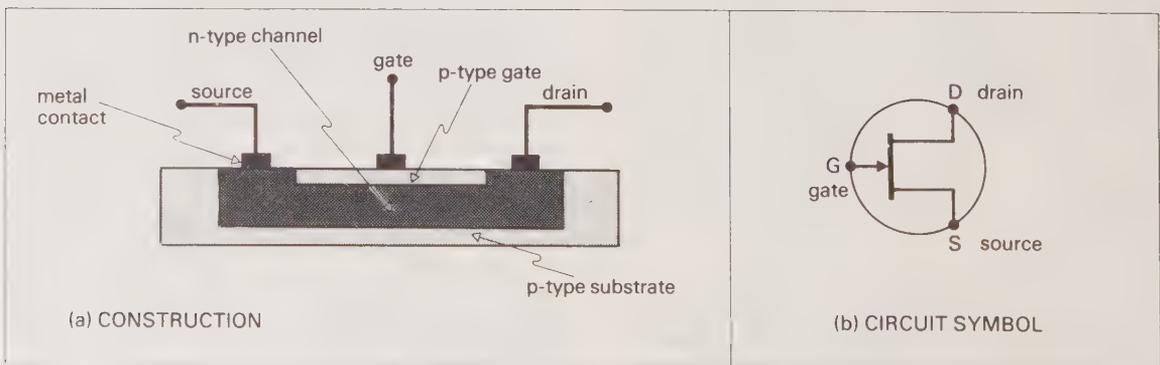


Fig. 61.10. The construction of an n-channel j.f.e.t. and its circuit symbol.

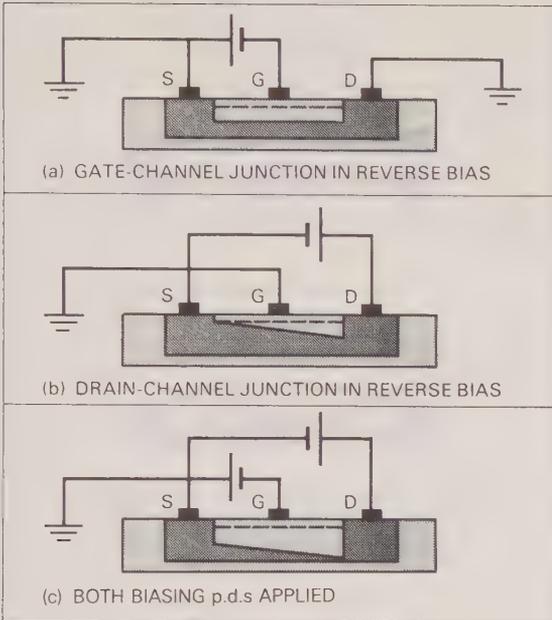


Fig. 61.11. Bias p.d.s in an n-channel j.f.e.t.

61.4 THE JUNCTION GATE FIELD EFFECT TRANSISTOR (J.F.E.T.)

Figs. 61.10 and 61.11 show the construction and bias p.d.s of an n-channel j.f.e.t.

Notes

(a) A p-channel j.f.e.t. has an n-type gate, a p-type channel and an n-type substrate. The circuit symbol has the arrow in the opposite direction.

(b) The current in the channel is controlled by reverse-biasing the p-n junctions formed by the gate-source and the gate-drain interfaces. When reverse-biased, the depletion layer widens, leaving a narrower channel.

(c) Electrons flow through the n-channel when the drain is made positive with respect to the source. For a p-channel j.f.e.t. the polarities of V_{DS} and V_{GS} would be reversed and holes would flow through the channel.

(d) The potential drop along the channel causes a variation in the reverse bias so that the gate becomes deeper at the drain end than at the source end.

(e) At **pinch-off** the gate reaches across the channel making it very narrow. The drain current I_D does not now show any significant increase with increase of V_{DS} . This is because any increase in I_D predicted from an increase in V_{DS} is balanced owing to the narrowing of the channel. If V_{GS} has a small negative value, pinch-off will occur at a lower value of V_{DS} and of I_D .

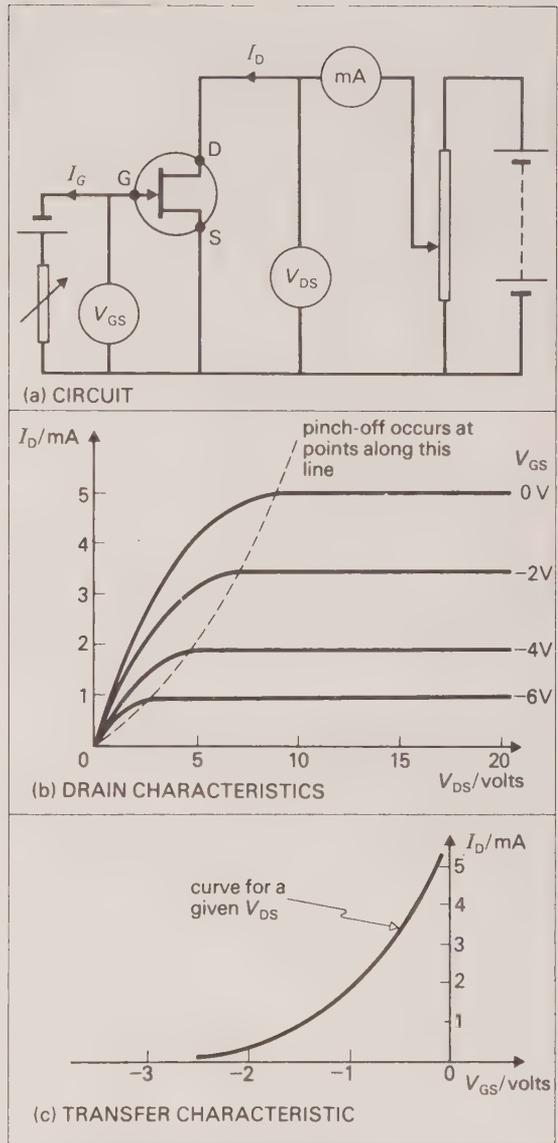


Fig. 61.12. Characteristics for an n-channel j.f.e.t.

61.5 CHARACTERISTICS OF n-CHANNEL J.F.E.T.S

The circuit with the **drain and transfer characteristics** that can be obtained for an n-channel j.f.e.t. are shown in fig. 61.12.

Notes

(a) The dotted line in fig. 61.12(b) is the locus of the points at which pinch-off occurs.

(b) (i) The **triode region** of operation, below pinch-off, is where the drain current I_D is strongly

affected by V_{DS} as well as by V_{GS} . In this region some j.f.e.t.s are used as voltage-controlled resistors.

(ii) The *pentode* or *pinched-off region* is where the drain current is largely independent of V_{DS} . In this region j.f.e.t.s are used as *amplifiers*.

(c) The **pinch-off p.d.** is the value of V_{DS} when the value of I_D becomes effectively constant. It depends on the value of V_{GS} .

(d) The gate current I_G is negligibly small ($\sim nA$) and is independent of V_{DS} . If the junction becomes forward-biased then I_G becomes large, and the j.f.e.t. is no longer a field effect device.

(e) The input impedance is very high ($>10^9 \Omega$).

(f) The **mutual conductance**

$$g_m = \frac{\Delta I_D}{\Delta V_{GS}}$$

and the **drain or output conductance**

$$g_D = \frac{\Delta I_D}{\Delta V_{DS}} = \frac{1}{r_D}$$

where r_D = drain resistance. g_D is usually very small compared with g_m and is correspondingly less important. (E.g. at $V_{DS} = 15 V$, $g_m \approx 10^{-5} S$, $g_D \approx 10^{-5} S$.)

61.6 THE INSULATED GATE FIELD EFFECT TRANSISTOR (I.G.F.E.T.)

Figs. 61.13 and 61.14 show the construction of, and application of drain and gate p.d.s to, a p-channel i.g.f.e.t. These devices are also referred to as *metallic oxide semiconductor transistors (m.o.s.t.s)*, as the insulating layer is usually a metallic oxide.

Notes

(a) Both source and drain form p-n junctions with the channel. The channel material is n-type but the device is a **p-channel** i.g.f.e.t. as it is the

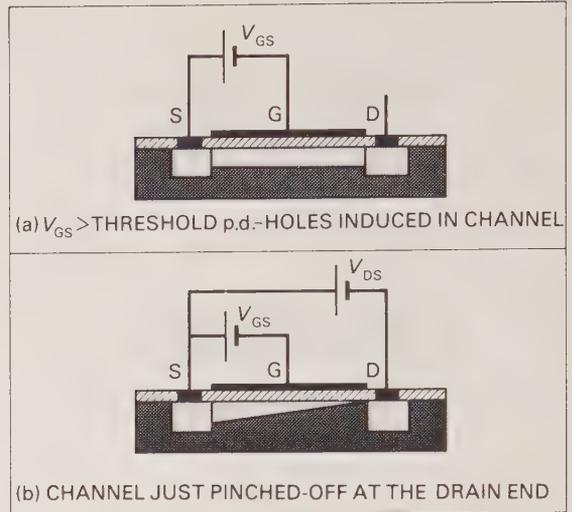


Fig. 61.14. P.d.s applied to a p-channel i.g.f.e.t.

induced *holes* that carry the current. The gate is a metal film separated from the silicon by SiO_2 or some other good insulator.

(b) Holes are induced in the channel after V_{GS} has become more negative than that necessary for the threshold p.d.

(c) If the drain is made negative relative to the source, holes flow towards it owing to the electric field. As the drain becomes more negative, the potential difference between gate and drain decreases and, when it falls below the threshold p.d., the channel is pinched off.

(d) An **n-channel** i.g.f.e.t. would have p-type substrate and *electrons* would be induced in it using a positive gate/source p.d. The circuit symbol for the n-channel type has the arrow reversed.

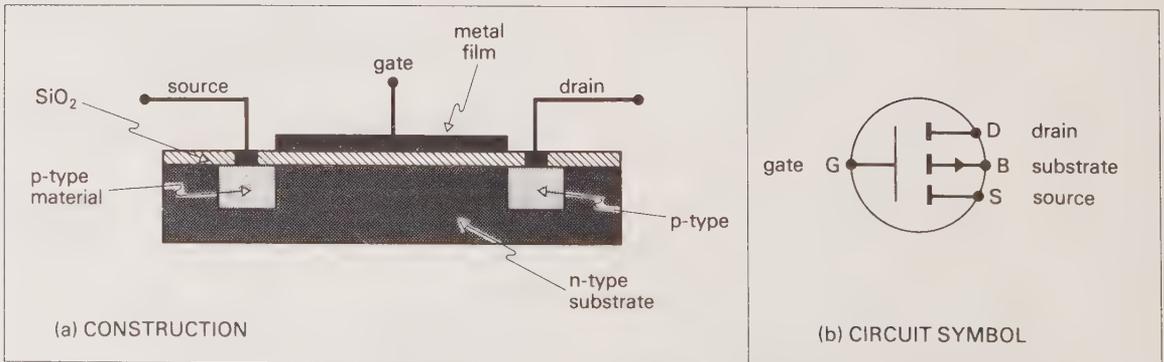


Fig. 61.13. The construction of a p-channel i.g.f.e.t. and its circuit symbol.

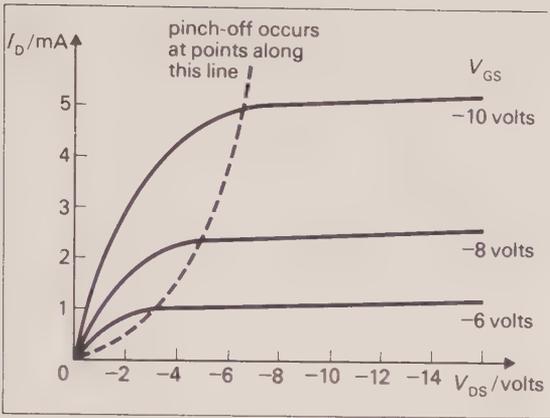


Fig. 61.15. Drain characteristics for a p-channel i.g.f.e.t.

61.7 CHARACTERISTICS OF p-CHANNEL I.G.F.E.T.S

Fig. 61.15 shows the drain characteristics for a p-channel i.g.f.e.t.

Notes

(a) The student should sketch a circuit diagram suitable for investigating these characteristics, using fig. 61.12(a) as a guide.

(b) The broken line shows the locus of the points where the channel pinches-off. This occurs when the drain/gate potential difference drops below the threshold value. At all points on this line

$$V_{DS} - V_{GS} = V_{threshold}$$

(c) I_C is very small ($\approx 10^{-14}$ A) because the gate is insulated.

(d) The i.g.f.e.t. described here is an *enhancement-mode version* as current can be enhanced only by the gate potential. (A less common device is the *depletion-mode* type in which the channel exists (because of appropriate impurity concentrations) before any gate potential is applied—the p.d. is then used to reduce the conductivity of the channel.)

61.8 THE BIPOLAR OR JUNCTION TRANSISTOR (B.J.T.)

Bipolar transistors consist of two p-n junctions very close together in a single piece of semiconducting material. Fig. 61.16 shows the construction of n-p-n and p-n-p transistors and their circuit symbols.

Notes

(a) The arrowhead in the symbol indicates the conventional direction of current flow. The n-p-n type are now more commonly used and are described here. (The action of the p-n-p type can be understood by reversing the polarity of the applied potential differences and remembering that the majority carriers have the opposite sign of charge.)

(b) The transistor is enclosed in a light-tight capsule for protection.

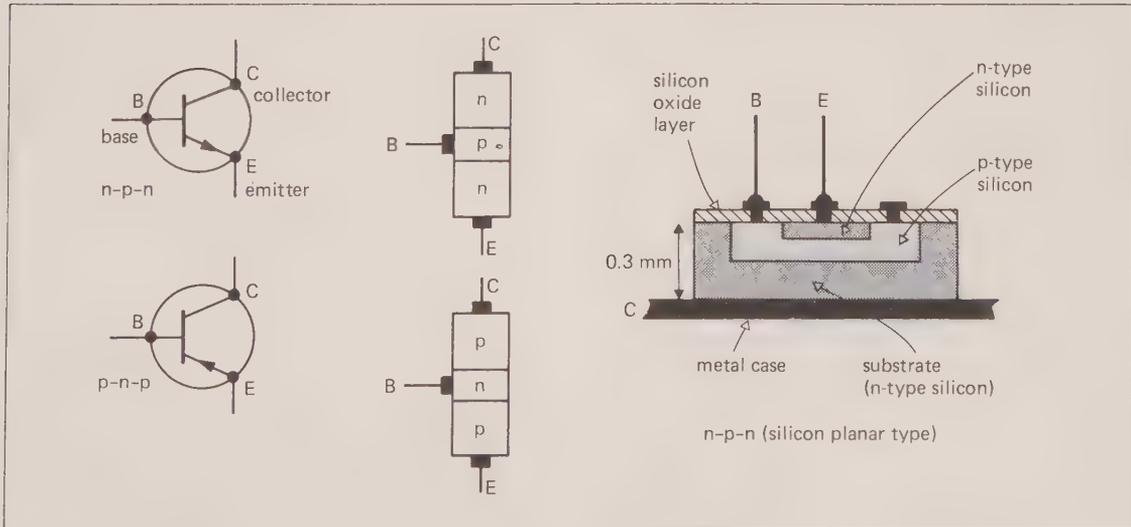


Fig. 61.16. The construction of n-p-n and p-n-p transistors and their circuit symbols.

(c) In the n-p-n transistor, the terms **emitter** and **collector** indicate what is happening to the electrons.

The action of the junction transistor

Notes

(a) Fig. 61.17 shows a circuit in which the emitter is common to base and collector circuits. This is called a **common-emitter** connection. The emitter is frequently earthed or **grounded**. Common-base and common-collector are other methods of connection which can be used.

(b) In both n-p-n and p-n-p transistors the collector-base junction must be reverse-biased: in the absence of \mathcal{E}_2 no current would flow across it. The minority-carrier number density in the base next to the collector junction is reduced.

(c) Similarly, the emitter-base junction must always be forward-biased. This raises the minority-carrier density in the base next to the emitter junction.

(d) Electrons are swept across the base-emitter junction into the very thin base region and

(i) about 99 per cent of the electrons are attracted straight through into the collector where they constitute the collector current I_C , but

(ii) only about one per cent combine with holes in the base causing a small base current I_B to flow. Thus the value of I_C may be perhaps 100 times I_B .

(e) In the p-n-p transistor, holes are swept across the p-n junction. A small proportion combine with electrons in the base to give a small base current. The electron flow in the base is maintained by supply from the external circuit.

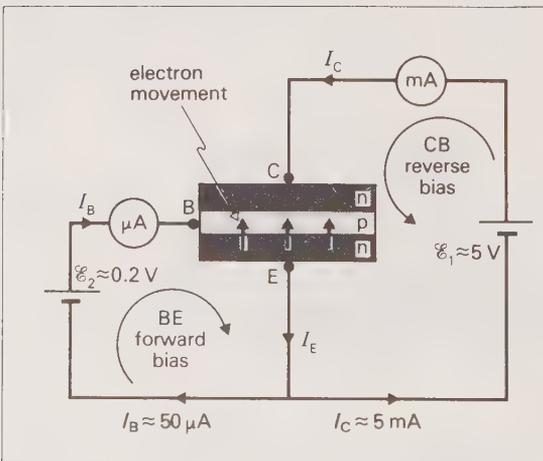


Fig. 61.17. The action of an n-p-n transistor.

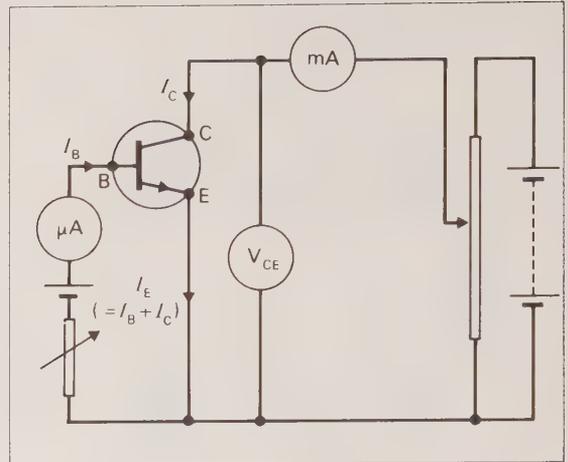


Fig. 61.18. Circuit for plotting characteristics for an n-p-n transistor.

61.9 CHARACTERISTICS OF THE JUNCTION TRANSISTOR

For the circuit of fig. 61.18, which uses the common-emitter connection, we are interested in the interdependence of

- (a) V_{CE} , the p.d. between collector and emitter
- (b) I_B , the base (input) current, and
- (c) I_C , the collector (output) current.

These relationships can be displayed by the **output** and **transfer** characteristics.

(A) The output characteristic

This shows the dependence of I_C on V_{CE} for fixed values of I_B . The results can be seen in fig. 61.19.

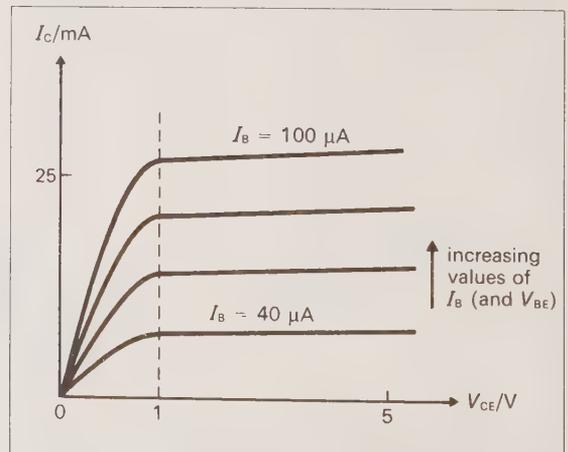


Fig. 61.19. Output characteristics for an n-p-n transistor.

The transistor is usually operated beyond the sharp change of slope. For this region we define the

$$\text{output resistance } [\Omega] = \frac{\Delta V_{CE}[\text{V}]}{\Delta I_C[\text{A}]}$$

for a fixed I_B .

It is the inverse of the slope of the characteristic. The collector-base junction is reverse-biased, which makes the output resistance of the order of 1 M Ω .

(B) The transfer characteristic

This shows the dependence of I_C on I_B for fixed values of V_{CE} . The results can be seen in fig. 61.20.

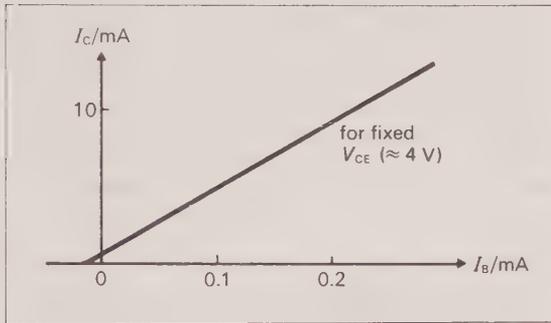


Fig. 61.20. Transfer characteristic for an n-p-n transistor.

The transfer characteristic is linear. For a fixed value of V_{CE} the quantity I_C/I_B is called the **static value of the forward current transfer ratio**, symbol h_{FE} (it used to be called the *current amplification factor*, α').

We define the **small signal forward current transfer ratio** h_{fe} by the equation

$$h_{fe} = \frac{\Delta I_C}{\Delta I_B}$$

The subscript e denotes the use of a common-emitter connection. h_{fe} is approximately equal to h_{FE} , being the slope of the characteristic. Typically it has a value ≈ 100 (no unit).

Note that the polarities and currents would be reversed in the circuit used to obtain the characteristics for a p-n-p transistor.

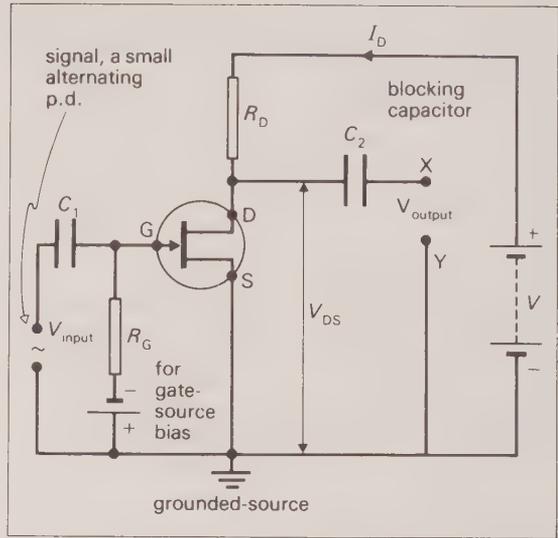


Fig. 61.21. An n-channel f.e.t. amplifier.

61.10 THE TRANSISTOR AS AN AMPLIFIER

(A) The f.e.t. amplifier

Refer to fig. 61.21.

Notes

(a) A p-channel device would have the circuit symbol arrow in the opposite direction and the polarities would be reversed.

(b) The alternating input signal changes the potential difference between gate and source, which causes an alternating current to be superimposed on the otherwise steady drain current I_D . The result is a variation in the p.d. across R_D , and so an alternating p.d. appears superimposed on the direct p.d. between drain and source. C_2 blocks the steady component so that the output p.d. V_{XY} is entirely alternating. The f.e.t. is a **voltage-controlled device**.

(c) C_1 blocks any direct component in the input signal, and the gate resistor R_G ensures that the correct value of V_{GS} is applied to the f.e.t. The arrangement is known as **resistor-capacitor coupling**.

(d) The source of the applied p.d. V_{app} supplies the power necessary for amplification.

$$V_{app} = (V_R)_D + V_{XY} = \text{constant}$$

Since V_{XY} decreases when $(V_R)_D$ (and hence V_{input}) increases, the input signal and the output are *in antiphase*.

(e) The **amplification of p.d. or gain** equals $\frac{\Delta V_{out}}{\Delta V_{in}}$. This can be shown to be approximately equal to $g_m R_D$, where g_m is the mutual conductance (p. 481).

(B) The b.j.t. amplifier

The circuit is basically the same as that for the f.e.t. amplifier (fig. 61.21), apart from the transistor symbol. Figs. 61.18 and 61.20 illustrate how the small current I_B in the base circuit controls the relatively large current I_C in the collector circuit. The b.j.t. is thus a **current-controlled** device as the current is amplified directly, whereas the f.e.t. is *voltage-controlled*. The input signal causes an alternating current to be superposed on the steady base current. This results in an alternating current (h_{fe} times as great) superposed on the steady state collector current. Variations in the current through the collector resistor R_C cause variations in the p.d. across R_C , and so in the potential of X relative to Y.

(C) The practical amplifier (fig. 61.22)

The correct operating p.d.s for a b.j.t. amplifier (V_{CE} and V_{BE}) are obtained from a single supply using the circuit of fig. 61.22. These p.d.s and the value of I_C are selected from the output characteristics (fig. 61.19) where they are linear and equally spaced so as to ensure minimum distortion of the output. A graph of I_C against V_{CE} for particular

values of V and R_C can be superimposed on the characteristics. From fig. 61.22, it can be seen that

$$V = I_C R_C + V_{CE}$$

$$\therefore I_C = -\frac{1}{R_C} \cdot V_{CE} + \frac{V}{R_C}$$

This straight-line graph of negative gradient is known as a **load line** and can be regarded as the output characteristic for the transistor and collector resistor taken together. The **operating point** is selected near the middle of the load line so that the undistorted limits (*swing*) of the output are at a maximum. The base-bias resistor R_B has large resistance so that the steady bias current flowing into the base is very small. The base is thus held at a potential of the *same polarity* as that of the collector but much smaller in magnitude. (Note that in the f.e.t. the gate is held at a potential *opposite* from the polarity of the drain.) Special bias circuits are used to stabilize amplifying circuits thereby preventing damage by overheating (**thermal runaway**) that would result from large increases in I_C . For greater amplification, several stages are connected *in cascade* (the output from one stage providing the input of the next.)

61.11 THE TRANSISTOR AS AN OSCILLATOR

The tuned circuit $L_1 C_1$ in fig. 61.23 is connected to the collector and changes in current through L_1 are fed back to the base-emitter circuit by L_2 . The battery makes the emitter negative relative to the

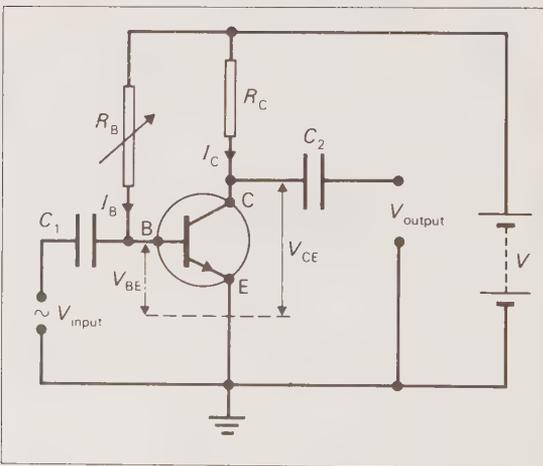


Fig. 61.22. A simple single-stage practical amplifying circuit.

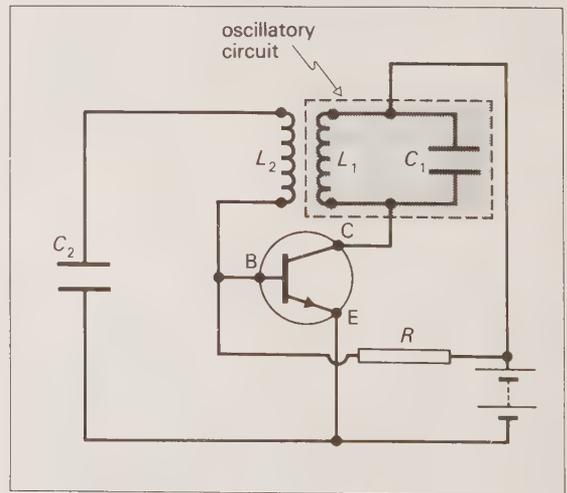


Fig. 61.23. Circuit for generating a.c. oscillations.

base and a high resistance R is used to provide the positive bias for the collector relative to the base. C_2 prevents the base and emitter from being short-circuited, so that a steady bias is maintained between them.

The circuit acts as an amplifier supplying its own input. If a small portion of the amplified p.d. is fed back in the correct phase, enough energy will be replaced in the circuit L_1C_1 to overcome the resistive and dielectric energy losses. This is an example of **positive feedback**.

Note that the coil/capacitor circuit oscillates, and the transistor merely provides the necessary energy from the battery at the right time. The *oscillation frequency* is given by $f = 1/2\pi\sqrt{L_1C_1}$.

61.12 THE TRANSISTOR AS A SWITCH

The output–input characteristic of fig. 61.24 (which refers to fig. 61.25) shows that when V_{input} is below V_1 , no current flows through the transistor so that V_{output} is +6 V. When V_{input} is above V_2 , the collector current is large enough for most of the input p.d. to appear across the collector resistor; V_{output} now remains near to 0 V as V_{input} is increased. The circuit of fig. 61.25 shows how a transistor can be used as a **switch**. If L and M are connected, the potential of X will be close to that of Y, so that V_{output} is near to 0 V. When N and M are connected, the potential of X will be the same as that of W so that V_{output} is +6 V. I.e. when

$$V_{input} = +6 \text{ V}, \quad V_{output} \approx 0 \text{ V}$$

and when

$$V_{input} = 0 \text{ V}, \quad V_{output} = +6 \text{ V}$$

The value of V_{input} determines whether the *level* of V_{output} is *high* (+6 V) or *low* (≈ 0 V). The transistor is said to be *cut off* (all the battery p.d. falls across it) when V_{input} is low, and *saturated* or *bottomed* (no battery p.d. across it) when V_{input} is high. The transistor as a *switch* operates between cut-off and saturation, giving only *two output levels*, thereby showing *non-linear* behaviour. Note that a transistor as an *amplifier* behaves *linearly* between its operating limits (fig. 61.49(a)). V_{input} is normally provided by another circuit, and switching operations can be executed extremely rapidly. A short switching time involves low power consumption and therefore reduces the chance of overheating. n–p–n transistors are preferable because their majority carriers (electrons) travel more quickly than the majority carriers (positive holes) in p–n–p types.

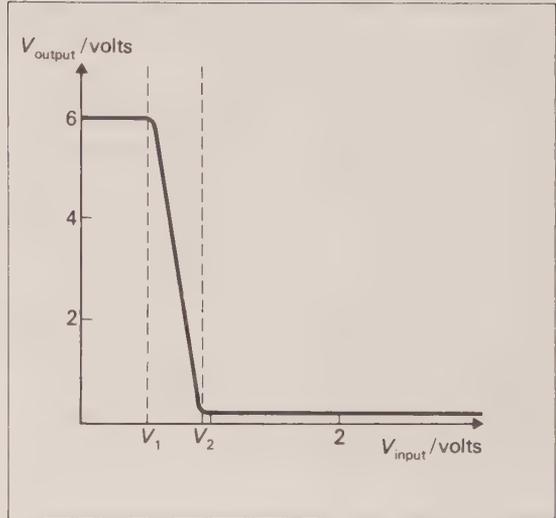


Fig. 61.24. The output–input characteristic for a single-stage transistor amplifier.

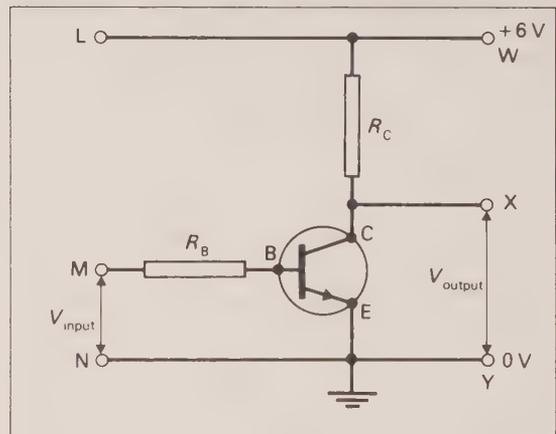


Fig. 61.25. Circuit showing a transistor used as a switch.

61.13 LOGIC GATES

Logic gates are digital circuits that are more complicated versions of the simple transistor switch and they find widespread application in electronic devices such as calculators, computers and telecommunications. The inputs and outputs of these gates involve either a *high* p.d. or a *low* p.d. *Potential difference* is usually preferable to the term *voltage* but the latter appears so often in electronics literature that it will be used, albeit relatively sparingly, in the topics that follow on from this

section. In *digital electronics* these high and low p.d.s represent the **B**inary digi**T**s (**bits**) 1 and 0. A great deal of information can be represented and transferred by a two-state signal, and digital outputs are generated from digital inputs. In *positive logic* circuits, a *high* p.d. represents *logic level 1*, and a *low* p.d. represents *logic level 0*. (*Negative logic* has the high p.d. representing logic level 0, and the low p.d. representing logic level 1. The circuits are the complements of positive logic circuits.) The different types of logic gate are switching circuits whose outputs are determined by the combination of their input signals. They are decision-making circuits using *combinational logic*. A **truth table** is used to list the outputs of each type of gate for all possible combinations of the inputs. The *true* state corresponds to a *high* (logic level 1) and the *false* state to a *low* (logic level 0). In the practical study of logic gates, they can often be treated as *black boxes* because it is more important to know how the output depends upon the input than to know all the circuit details. The *function* and *characteristics* of the black boxes are of paramount importance. In the following summary of

the basic logic gates, the American symbols are used and the less common British Standard symbols are listed at the end of the chapter (fig. 61.58).

(A) OR gate

See fig. 61.26. *The output is high if either input A OR input B OR both are high.*

(B) AND gate

See fig. 61.27. *The output is high if input A AND input B are both high.*

(C) NOT gate or inverter

See fig. 61.28. *The output is NOT high if the input is high, i.e. the input signal is always inverted.* The simple transistor switch behaves as a NOT gate.

(D) NOR gate

See fig. 61.29. *The output is high if neither input A NOR input B is high.*

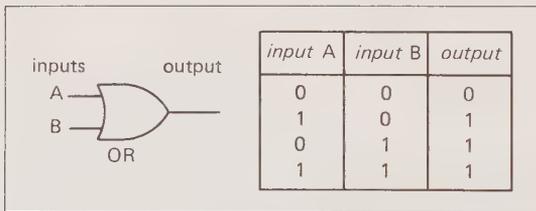


Fig. 61.26. The OR gate and truth table.

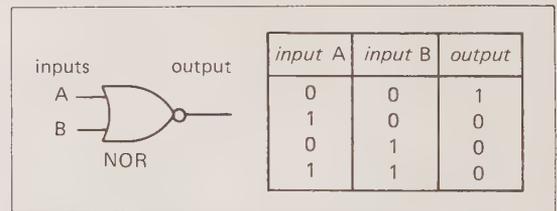


Fig. 61.29. The NOR gate and truth table.

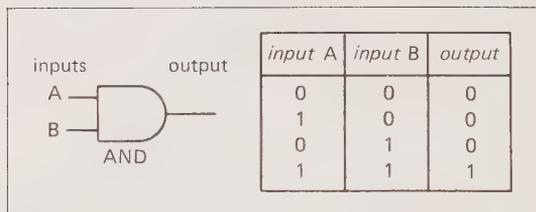


Fig. 61.27. The AND gate and truth table.

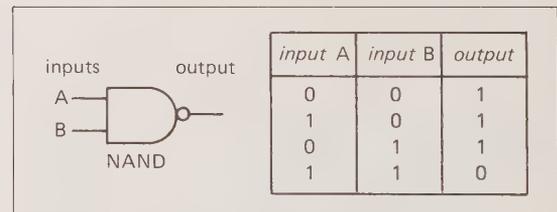


Fig. 61.30. The NAND gate and truth table.

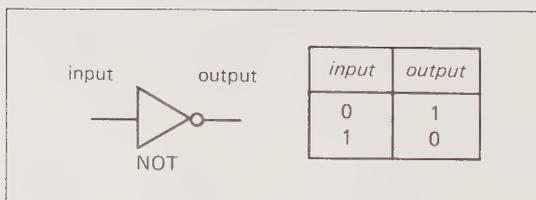


Fig. 61.28. The NOT gate and truth table.

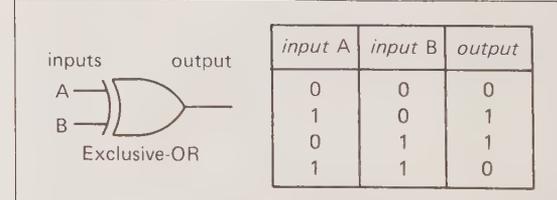


Fig. 61.31. The Exclusive-OR gate and truth table.

(E) NAND gate

See fig. 61.30. *The output is NOT high if input A AND input B are both high.*

(F) Exclusive-OR gate

See fig. 61.31. *The output is high if either input A OR input B (but not both) is high. The inputs have to be different. An Exclusive-NOR or PARITY gate is the inverse of an Exclusive-OR gate, i.e. the output is high when the inputs are the same (both high or both low).*

(G) Non-inverting gate

See fig. 61.32. *The output is the same as the input. This gate can be used to isolate one part of a system from another.*

The above truth tables can be confirmed by using suitable transistor and indicator units, such as those designed for use in school courses. If an indicator lamp lights, the output is 1; if it does not, the output is 0. Note that if the two inputs of NOR or NAND gates are joined together, they become NOT gates or inverters. NOR and NAND

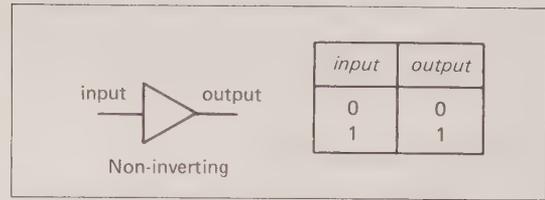


Fig. 61.32. *The Non-inverting gate and truth table.*

gates are particularly important in that combinations of each type can perform the functions of other gates. Fig. 61.33 shows how they can be used to function as NOT, AND and OR gates. The inputs shown as 1 are connected to a permanent logic 1, and those shown as 0 are connected to a permanent logic 0.

61.14 CONTROL SYSTEMS

One use of logic gates is in the *control* of electronic systems. A **transducer** or **sensor** converts changes in a non-electrical physical quantity, such as pressure, temperature or light intensity, into corresponding changes in electrical p.d. These inputs

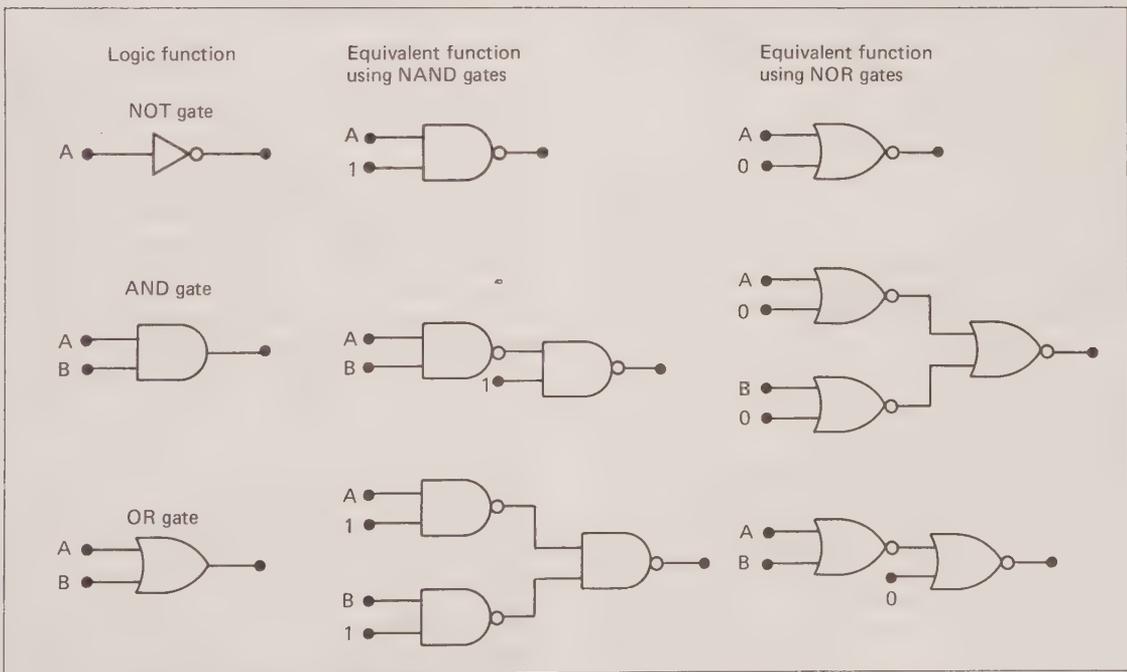


Fig. 61.33. *Combinations of NAND and NOR gates.*

are processed by the system, and the resulting output is used by another transducer to perform some required function such as switching on central heating.

(A) Light-dependent switch

When the *light intensity* decreases, the resistance of the *light-dependent resistor (LDR)* increases so that the p.d. across it becomes greater. This means that the potential of point A in fig. 61.34 must decrease. When it is sufficiently close to 0 V, the input to the inverter becomes a *low* so that the lamp is switched on. If the LDR is replaced by a thermistor, the lamp would come on when the *temperature* fell below a certain level. This is because the resistance of the thermistor increases as the temperature drops thereby decreasing the potential of point A. For more demanding loads, such as a motor or heater, a relay would be necessary to activate a separate circuit because the output current from the gate would be too small.

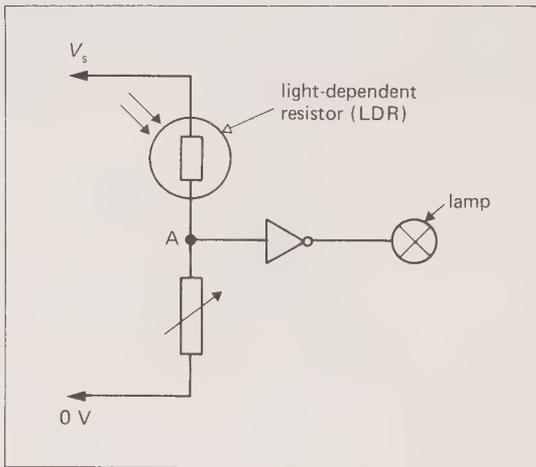


Fig. 61.34. A circuit to turn on a lamp in the dark.

(B) Car-seat heater

The control of an electronic system can be summarized using a block diagram rather than detailed circuits. Even the most complex system can be divided into a number of basic building blocks, each of which has a *function* such as amplification or switching. An understanding of these functions is usually more important than a knowledge of individual circuits. A simple example is given in fig. 61.35. This shows an AND gate which would switch on a car-seat heater when a passenger sat

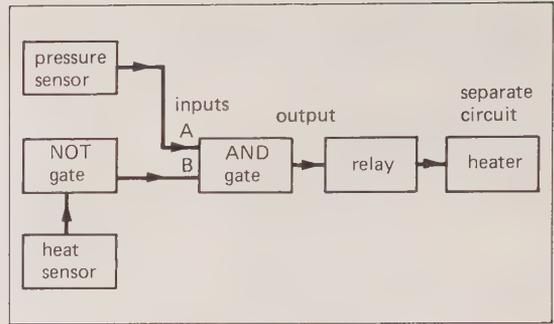


Fig. 61.35. Block diagram for car-seat heater.

on the seat which was cold. A *pressure sensor* and a *temperature sensor* would therefore be required. One form of pressure sensor is a **resilistor** (resilient resistor), which consists of a piece of foam containing carbon particles. When the resilistor is compressed, its resistance decreases. *High pressure* and *high temperature* would be *high*s (1), and a high output would be required to operate the relay. The heater would switch off if *either* the passenger left the seat *or* the temperature rose to a specified level.

61.15 BINARY ADDITION

The *decimal system* has *ten* as the *base* and therefore uses the *ten digits* 0 to 9. When a count goes beyond 9, a 1 is placed in a second column to the left of the units column to represent the tens, and so on. The values of successive columns, from the right, are $10^0, 10^1, 10^2, 10^3, 10^4$, etc. The *binary system* has *two* as the *base* and therefore only *two digits* (0 and 1). These *bits* are represented by *low* and *high* p.d. levels. Numbers are written in columns which represent, from the right, $2^0, 2^1, 2^2, 2^3, 2^4$, etc. The **least significant bit** (l.s.b.) is on the far right and the **most significant bit** (m.s.b.) is on the far left. For example, binary number 111011 can be written as follows:

m.s.b. ←	1	1	1	0	1	1	l.s.b.
	1×2^5	$+ 1 \times 2^4$	$+ 1 \times 2^3$	$+ 0 \times 2^2$	$+ 1 \times 2^1$	$+ 1 \times 2^0$	
	32	+ 16	+ 8	+ 0	+ 2	+ 1	= 59

(A) Half adder

This system adds two binary digits together to give a SUM digit and a CARRY digit. Binary addition of two bits gives:

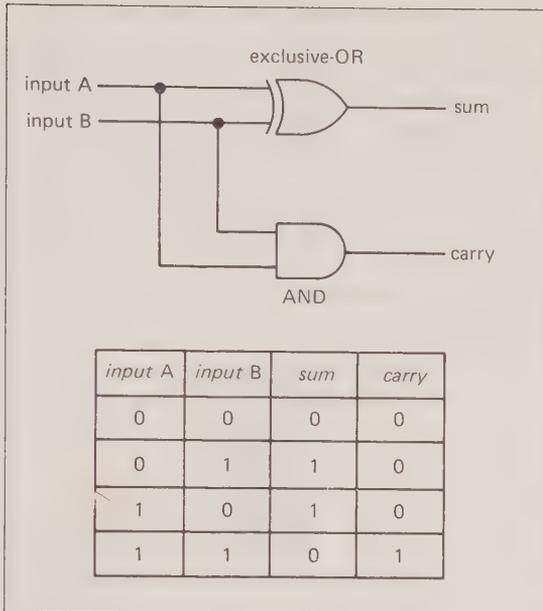


Fig. 61.36. Half-adder circuit and truth table.

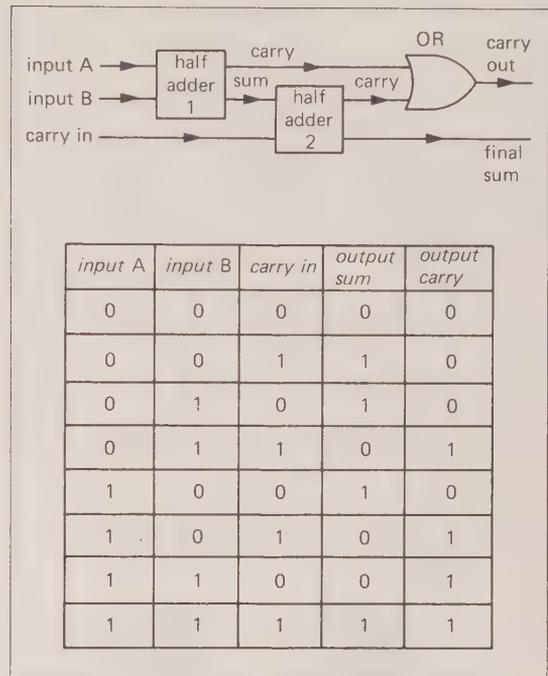


Fig. 61.37. Full adder and truth table.

bit A	bit B	SUM	CARRY
0	+ 0	= 0	0
0	+ 1	= 1	0
1	+ 0	= 1	0
1	+ 1	= 0	1

The final carry would go to the next column on the left. The circuit (fig. 61.36) needs two inputs (one for each bit being added) and two outputs (one for the *sum* and one for the *carry*). The output of the exclusive-OR gate gives the sum, and the output of the AND gate gives the carry. Binary addition occurs if both bits are applied to the inputs of both gates at the same time.

(B) Full adder

A full adder is required for adding two binary bits as well as the *carry* resulting from the addition in the least significant bit column. The circuit (fig. 61.37) needs three inputs (two for the bits and one for the carry). This can be done using a second half adder to add the incoming carry to the sum of the input bits. An OR gate combines the carry from half adder 1 with the carry from half adder 2. If two numbers, each of *n* bits, are to be added, the circuit requires *2n* inputs to *n* full adders and

(*n* + 1) outputs, e.g. the addition of two three-bit numbers requires three full adders. Binary *subtraction* can be carried out by combining an inverter and an adder. *Multiplication* and *division* involve repeated addition and subtraction.

61.16 PULSES AND MULTIVIBRATORS

Pulse production

When the switch in fig. 61.38 (overleaf) goes from 0 V to *V_s*, the value of *V_R* rapidly becomes high so that it approaches *V_s*. The output of the NOT gate therefore goes low so that the indicator lamp goes off. *V_R* then falls exponentially as the capacitor becomes charged. At *V_t*, the transition p.d. for the gate, *V_{out}* rapidly becomes high again and the lamp relights. If the inverter is replaced by a non-inverting gate, a positive-going pulse (fig. 61.39, overleaf) is obtained so that the lamp *comes on* for a short time. The length of these pulses is determined by the values of *C* and *R*. The action of the gate, in each case, is to convert a slowly changing output into a sharp pulse.

Multivibrators

A multivibrator is an oscillator consisting of two inverters coupled by capacitors and/or resistors so

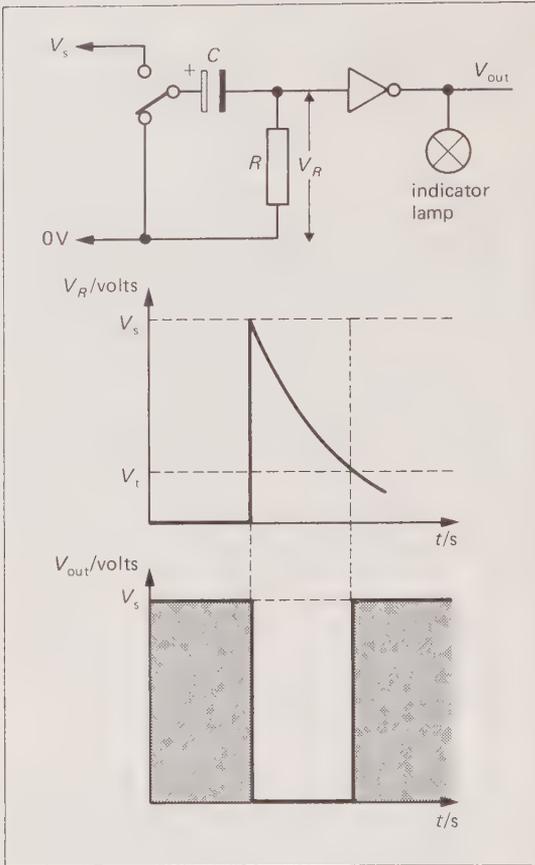


Fig. 61.38. Production of a negative-going pulse.

that the output of one provides the input for the other. The two outputs are *complementary*, i.e. when one is *high*, the other is *low*. The rapid switch-over between output states in each transistor can produce a *square-wave* output. Waves like this are a combination of a fundamental frequency and a large number of its odd harmonics—hence the term *multivibrator*. Different coupling of the transistors in the inverters produces multivibrators with different actions.

(A) The astable multivibrator

The two inverters are coupled by *capacitors*, so that the output of each stage sends a pulse to the other stage which returns it to the first. This *free-running* multivibrator has *no stable states* and switches continuously from one state to the other. These oscillations can be demonstrated using two inverters and two indicator lamps as shown in the circuit of fig. 61.40. The frequency of the flashing lamps is inversely proportional to the sum of the

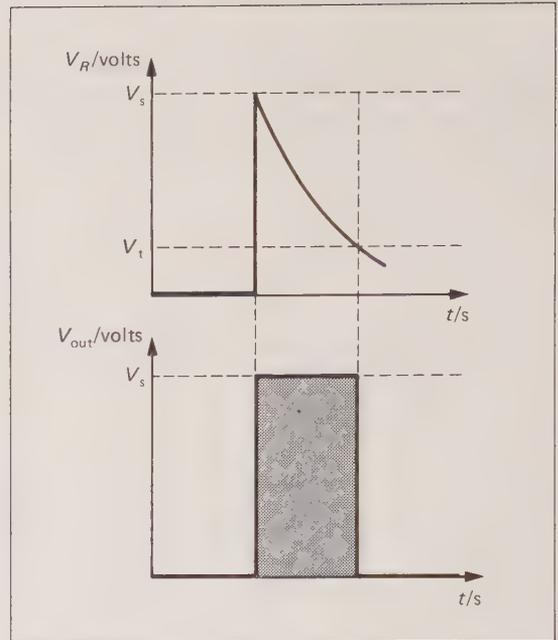


Fig. 61.39. A positive-going pulse.

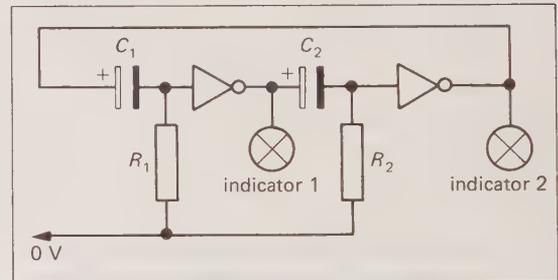


Fig. 61.40. A demonstration astable circuit.

time constants of the two circuits and therefore depends on the values of the resistors and capacitors. When the circuit is first connected, the oscillations may not start until a capacitor is shorted briefly. The continuous waveforms produced at the two outputs are *complementary* trains of almost rectangular pulses. These can be displayed on a c.r.o. (or heard in a loudspeaker if *audiofrequencies* are produced). If $C_1 = C_2$ and $R_1 = R_2$, *square-wave* pulses or oscillations are generated. The train of regular pulses can be used for many forms of timing, e.g. digital watches and computers, and for generating musical notes.

(B) The monostable multivibrator

One stage is coupled by a capacitor and the other by a resistor as shown in fig. 61.41. A monostable circuit has *one stable state* and one unstable state. The stable state for the circuit shown has a low output from B because of B's high input. When switch S is closed briefly, the high output from A and therefore the high input to B are forced to become low. B's output is therefore high and the indicator lamp comes on. The capacitor C is now being charged, and the lamp goes off when B's input is high enough to return its output to the stable low state. The length of the output pulse depends upon the values of C and R₁, and not upon the length of the trigger pulse that switched the circuit from its normal stable state to its unstable state. These circuits are important in industrial and domestic control of timings, e.g.

(a) a number of monostables can be cascaded to perform the sequence of operations involved in spot-welding,

(b) a monostable can be used to control the lighting of a communal area (when the switch is pressed, the light goes on for a predetermined time before going off again).

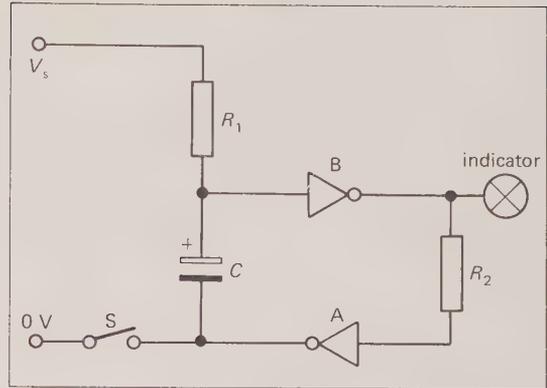


Fig. 61.41. A demonstration monostable circuit.

(C) The bistable multivibrator

The two stages are coupled by resistors and the bistable or **flip-flop** circuit has *two stable states*. This can be demonstrated using two inverters (containing the coupling resistors) and two indicators as shown in the circuit of fig. 61.42. If the input to A is high, indicator lamp 1 is off and the input to B is low. Indicator lamp 2 is on as the output of B is now high. If B's output is fed back to input A, the original input is reinforced and this represents one stable state. If the input to A is made low, the input to B becomes high. This means that indicator lamp 1 is on and indicator lamp 2 is off, illustrating the other stable state. In practice, A or B are triggered directly. The bistable circuit therefore remains in either stable state until a suitable trigger pulse makes it switch (or flip) into the other stable state. A trigger pulse applied to the other input makes the circuit switch back (or flop) into the initial stable state. This basic bistable is called an SR bistable or flip-flop, where the two states (and inputs) are labelled *set* and *reset*. The two outputs are labelled Q and \bar{Q} (not Q). Fig. 61.43 shows the SR symbol and truth table. The bistable (flip-flop) circuit can therefore be used as a single-bit memory, with one state representing bit 1 (Q = 1) and the other state representing bit 0 (Q = 0). The storage of each bit requires a separate bistable so that computer memories use large numbers of these circuits.

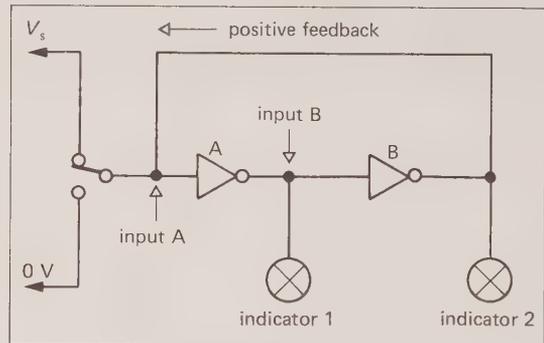


Fig. 61.42. A demonstration bistable circuit.

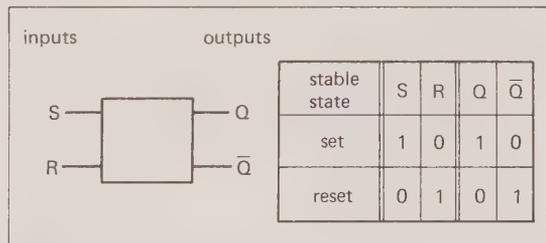


Fig. 61.43. SR flip-flop and truth table.

61.17 BINARY COUNTING

(A) Triggered bistables (T-type)

The SR bistable can be adapted so that successive pulses applied to a third input, called the *trigger* (T) or *clock* (CK), are *guided* to the other two inputs alternately. This results in a change of state for each input pulse so that *one* output pulse requires *two* input pulses. The T-type bistable or flip-flop therefore divides the input frequency by two, and this action forms the basis of *binary counting*. A square-wave pulse has a rapidly rising *positive-going* edge and a rapidly falling *negative-going* edge. Some bistables change state when the trigger pulse rises from logic level 0 to 1 (*positive-edge triggering*), and others when the trigger pulse falls from level 1 to 0 (*negative-edge triggering*).

(B) Clocked bistables

A clocked bistable has pulses of a fixed or *clocked* frequency applied to the third input (CK). The pulses can be provided by an astable multivibrator and should be high-quality square waves to ensure reliable triggering. The action of a clocked bistable can be displayed on a c.r.o. screen using the arrangement of fig. 61.44. The inverter placed

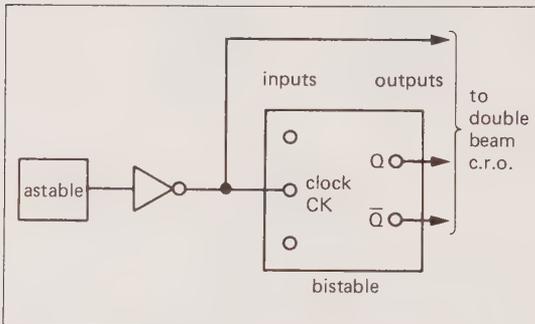


Fig. 61.44. Demonstration of a clocked bistable.

between the astable and bistable improves the squareness of the pulses. The clock input and the complementary Q and \bar{Q} output waveforms are shown in fig. 61.45. Any two of these can be displayed simultaneously by using the two Y-channels of a double-beam c.r.o. This example shows the outputs for *positive-edge triggering*, and it is clear that the output frequency is half the clock pulse frequency.

(C) The binary counter

A binary counter consists of a number of bistables connected in series (*cascade*) so that the Q output of one becomes the clock (CK) input to the next.

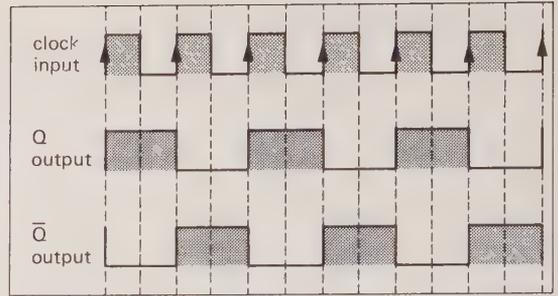


Fig. 61.45. Input and output waveforms for a clocked bistable.

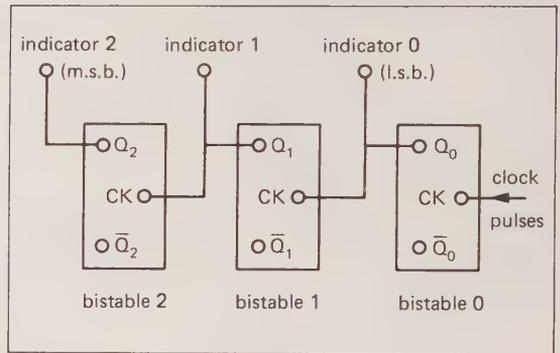


Fig. 61.46. Three-stage binary counter.

The *high* and *low* states of each Q output correspond to the bits 1 and 0. Fig. 61.46 shows a three-stage counter which can count up to seven (binary 111) input pulses before resetting. The instantaneous count is given by the states of the Q outputs. In the timing diagram of fig. 61.47, the bistables change state whenever their CK inputs

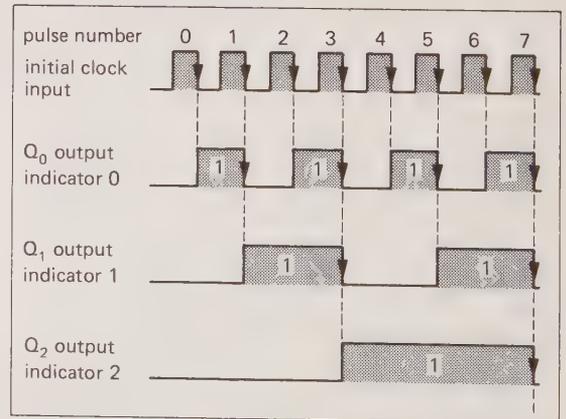


Fig. 61.47. Timing diagram for a three-stage binary counter.

move from 1 to 0, i.e. *negative-edge triggering*. Each bistable produces *one* output pulse for every *two* input pulses. The indicators should be arranged with Q_0 (l.s.b.) on the far right, then Q_1 on its left and then Q_2 (m.s.b.), so that the correct binary count is displayed. For example, clock pulse number 6 gives a binary count of 110. For counting downwards, the \bar{Q} outputs are used although the indicators are still connected to the Q outputs. Binary counters are vital components in many electronics systems such as digital watches and computers.

61.18 THE OPERATIONAL AMPLIFIER

(A) Introduction

The operational amplifier (*op amp*) was designed originally to solve complex mathematical equations electronically by performing *operations* such as addition, subtraction and integration. These mathematical operations form the basis of *analogue computing* (p. 501) in which physical systems are represented by differential equations. They were made from *discrete* components at first but are now produced in *integrated circuit* (IC) form. One of the main present-day uses is for high-gain a.c. and d.c. voltage amplification. The symbol, inputs and output are shown in fig. 61.48(a).

Notes

(a) Fig. 61.48(b) shows that the power supply is *dual rail* to allow for both positive and negative swings. It is typically in the range $\pm 3\text{ V} \rightarrow \pm 18\text{ V}$ and is often left off circuit diagrams, although it is essential for the operation of the amplifier because it supplies the power for the amplified signal. Fig. 61.48(a) shows that the centre point of the power supply is *common* to the input and output circuits.

(b) The **open-loop voltage gain** A_o (i.e. gain without feedback) is extremely high ($\sim 10^5$), although it decreases as the frequency of an a.c. input increases.

(c) In an *ideal op amp*, the *input resistance* is very high ($\sim 10^6 - 10^{12}\ \Omega$) so that negligible current is drawn from the input signal source. This means that the op amp does not alter the input p.d. The input potentials V_1 and V_2 are effectively the same because the difference between them is relatively very small. These assumptions are reasonable approximations and they simplify the theoretical derivations for amplifier gain.

(d) The *output resistance* is very low ($\sim 10^2\ \Omega$). This means that the output V_o can be transferred efficiently to a load greater than a few $\text{k}\Omega$. *Maximum voltage transfer* requires the output resistance to be very much lower than the load resistance

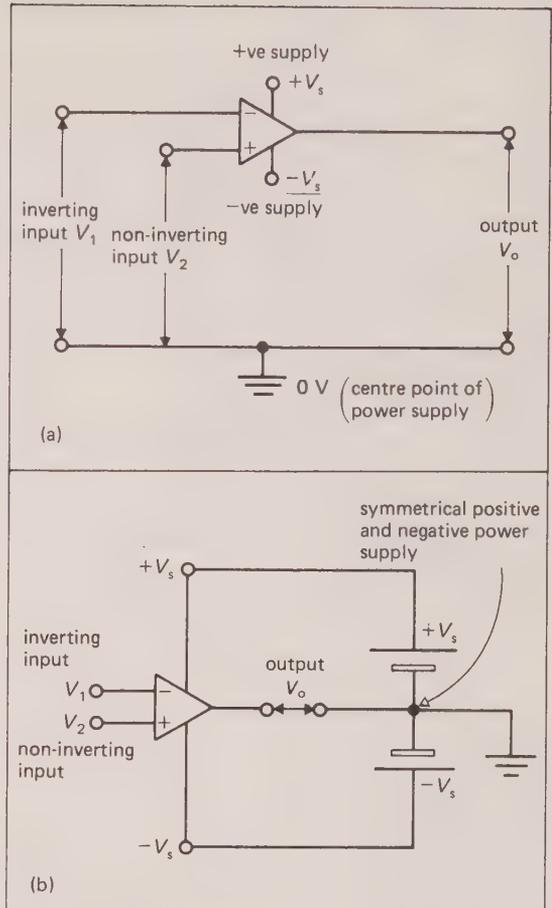


Fig. 61.48. Symbol, inputs and output for the op amp.

(this is different from *maximum power transfer* where the two resistances should be equal (p. 386)).

(e) If V_2 is positive relative to the other input, V_o is also positive; if V_2 is negative, V_o is also negative. V_2 and V_o are therefore *in phase*. If V_1 is positive relative to the other input, V_o is negative, and vice versa. V_1 and V_o are therefore *in antiphase*. The op amp is actually a *differential voltage amplifier* because it amplifies the potential difference between the inputs V_1 and V_2 . The output p.d. is given by the expression $V_o = A_o(V_2 - V_1)$. When there is no difference between V_1 and V_2 , the output V_o should ideally be at *earth* potential.

(f) Fig. 61.49(b) (overleaf) shows that the amplified output is distorted if the peak values would be greater than $\pm V_s$. This means that the amplitude of the input signal is limited. The op amp is said to be *saturated* if the inputs are outside the *linear range* shown in fig. 61.49(a). Op amps saturate at about 1–2 V below $\pm V_s$.

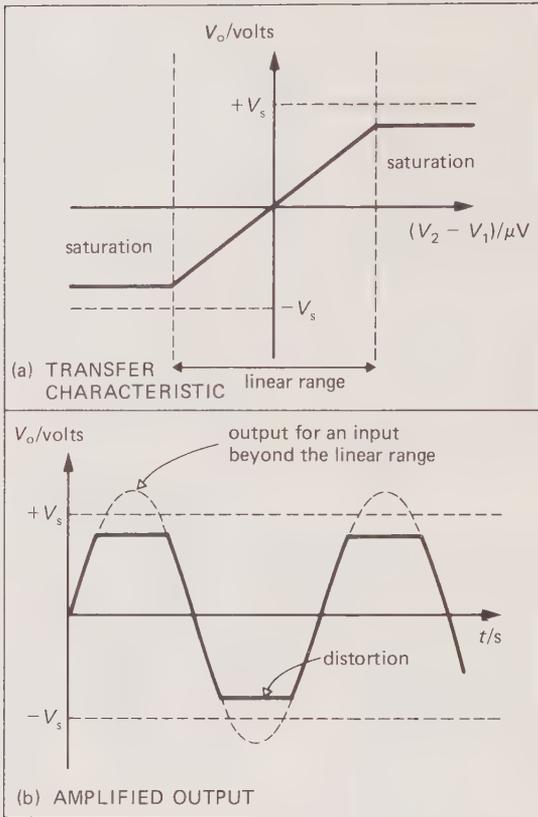


Fig. 61.49. Op amp characteristic and output.

(B) Voltage amplifier

The open-loop voltage gain A_o is usually too high to be useful, is sensitive to temperature changes and decreases greatly as the frequency increases. **Negative feedback** is when a small proportion of the output is fed back to the inverting input, thereby reducing the gain which is now called the **closed-loop gain A**. This much smaller gain is predictable and constant, and leads to less distortion as well as greater stability. Also, the gain remains constant over a wider range of input frequencies (*bandwidth*), i.e. greater negative feedback leads to improved **frequency response**. If greater gain is required, several op amp stages can be coupled together. Clearly, the amount of negative feedback in an op amp is an important *design* consideration.

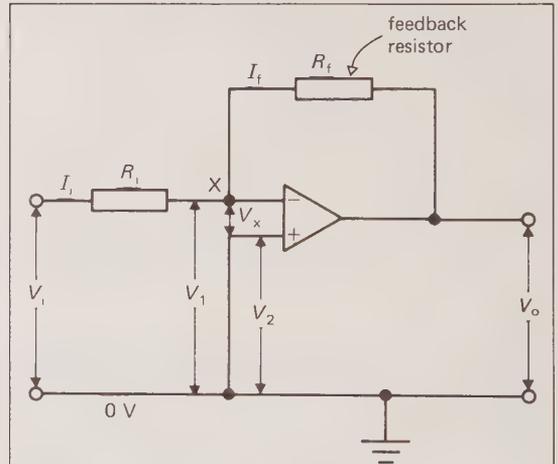
Inverting amplifier

In fig. 61.50, V_i is the input voltage to be amplified and will be in antiphase with the output V_o . R_f feeds back a certain proportion of the output voltage to the inverting input. Assuming that the op amp draws negligible current and that the input potentials V_1 and V_2 are effectively the same (because the difference between them is relatively very small),

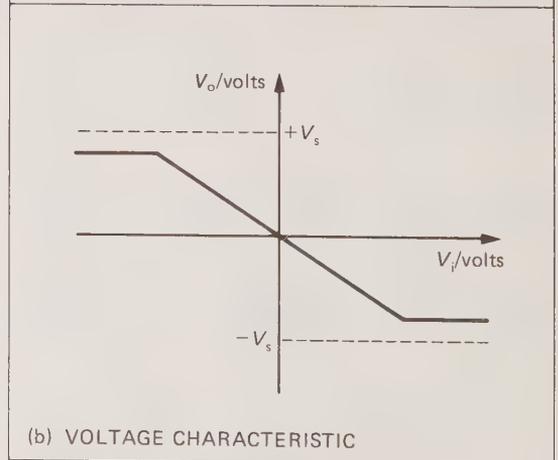
$$V_x = V_1 - V_2 = 0 \text{ V}$$

$$\text{As } V_2 = 0 \text{ V, } V_1 = 0 \text{ V,}$$

(X is called a **virtual earth point**)



(a) CIRCUIT SHOWING FEEDBACK



(b) VOLTAGE CHARACTERISTIC

Fig. 61.50. The inverting amplifier.

so that

$$I_i = \frac{(V_i - 0)}{R_i}$$

and

$$I_f = \frac{(0 - V_o)}{R_f}$$

but as $I_i = I_f$, since the op amp draws no current,

$$\frac{V_i}{R_i} = -\frac{V_o}{R_f}$$

therefore the closed-loop gain is given by

$$A = \frac{V_o}{V_i} = -\frac{R_f}{R_i}$$

Notes

(a) V_o is positive when V_i is negative, and vice versa.

(b) A depends only upon R_f and R_i , and not upon the characteristics of the op amp.

(c) If $R_f = R_i$, $A = -1$ so that the circuit behaves as a NOT gate or inverter. It is still an analogue device and therefore has a *range* of possible output values.

(d) The gradient of the voltage characteristic is equal to A , the closed-loop gain, which will be negative because V_i is in antiphase with V_o .

Non-inverting amplifier

In fig. 61.51, V_i is applied to the non-inverting input so that it is in phase with the output V_o . The fraction of V_o appearing across R_i is fed back to the inverting input to provide negative feedback. This fraction, known as the *feedback factor* or *fraction* β is given by

$$\beta = \frac{R_i}{R_i + R_f}$$

The effective output voltage $V_o = A_o(V_i - \beta V_o)$.

$$\therefore \frac{V_o}{V_i} = A = \frac{A_o}{1 + \beta A_o}$$

If A_o is large so that $\beta A_o \gg 1$, $A = 1/\beta$

$$\therefore A = \frac{R_i + R_f}{R_i} = 1 + \frac{R_f}{R_i}$$

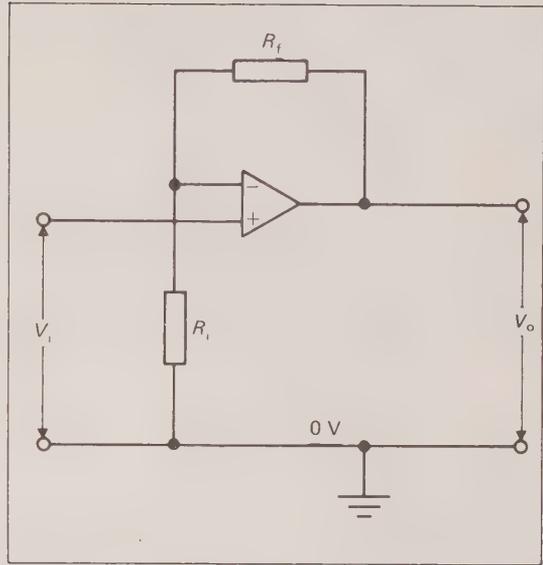


Fig. 61.51. The non-inverting amplifier.

Notes

(a) A , which is positive, depends only upon R_f and R_i , and not upon the characteristics of the op amp.

(b) If the output is connected directly to the inverting input so that there is now no potential divider, R_f becomes zero and R_i is infinite. All the output is fed back and so $\beta = 1$. If A_o is large, $A = 1/\beta$ and so A approaches 1 (*unity gain*). This is a **voltage-follower** circuit because V_o follows V_i . As it has high input impedance and low output impedance, it is used as a **buffer circuit** to *match* a high-impedance source to a low-impedance load. For example, a photodiode *current* is *amplified* to operate a chart recorder—the extra current comes from the op amp's power supply.

(c) The relationship between the gains in the inverting and non-inverting amplifier circuits is described by the following equation:

$$\left(\begin{array}{c} \text{inverting} \\ \text{circuit gain} \end{array} \right) = (\beta - 1) \times \left(\begin{array}{c} \text{non-inverting} \\ \text{circuit gain} \end{array} \right)$$

(C) Voltage comparator

From the relationship $V_o = A_o(V_2 - V_1)$, it is clear that the potential difference between the inputs is amplified to produce the output. A small change in $(V_2 - V_1)$ causes V_o to approach $+V_s$ or $-V_s$ depending on whether V_2 is greater or less than V_1 . The op amp is a **comparator**, or **differential amplifier**, in that it compares V_1 and V_2 and produces an output that results from this comparison.

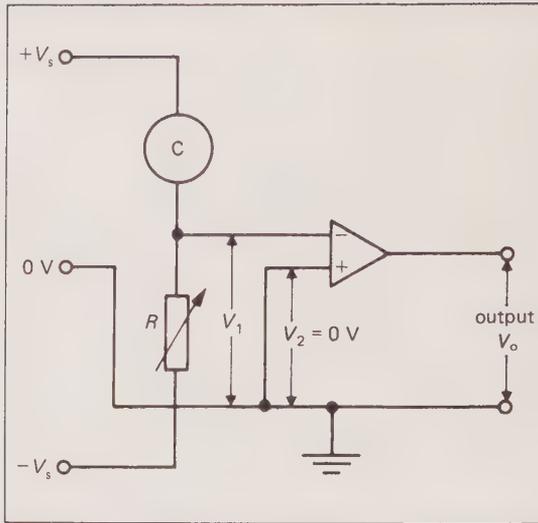


Fig. 61.52. Switching circuit using a comparator.

Fig. 61.52 shows how the comparator effectively acts as a *digital* device. Component C could be a photodiode or thermistor and it forms a voltage divider with resistor R across the dual-rail supply. The value of V_1 would be altered by variations in the external conditions, e.g. change in temperature in the case of a thermistor. The op amp compares V_1 with its other input V_2 , which is at 0 V. If V_1 becomes less than V_2 , the output V_o switches from near $-V_s$ to near $+V_s$. This output could be used to switch on a warning light. Note that a diode is required in the output if the lamp is to switch on only when $V_2 > V_1$ (rather than when $V_1 > V_2$ as well).

(D) Summing amplifier

Assuming that the op amp draws no current, the currents in fig. 61.53 are related as follows:

$$I_f = I_a + I_b + I_c$$

If X is a virtual earth point so that $V_x = 0$ V,

$$-\frac{V_o}{R_f} = \frac{V_a}{R_a} + \frac{V_b}{R_b} + \frac{V_c}{R_c}$$

$$\therefore V_o = -\left[\left(V_a \times \frac{R_f}{R_a} \right) + \left(V_b \times \frac{R_f}{R_b} \right) + \left(V_c \times \frac{R_f}{R_c} \right) \right]$$

This shows that the circuit multiplies the input voltages by different factors and adds them. The

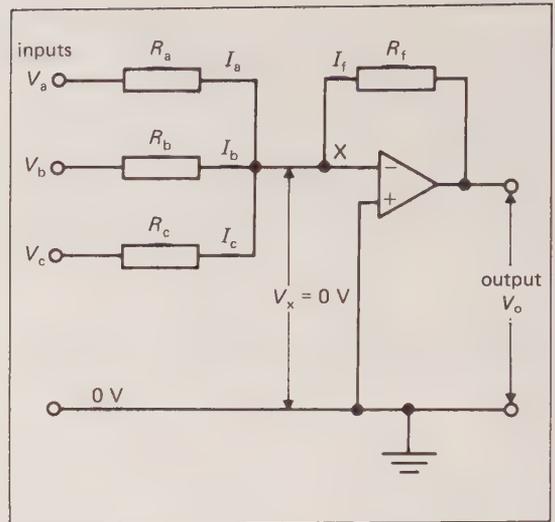


Fig. 61.53. The op amp as a summing amplifier.

gains of this **summing amplifier** depend upon the resistance values. If all the resistances are equal, $V_o = -(V_a + V_b + V_c)$. This summing circuit can be used as a *digital-to-analogue* converter. The resistors are arranged to form a binary sequence ($R, 2R, 4R, 8R \dots$) and are connected to a suitable voltage supply via switches. If R_f is given an appropriate value, closing the switches (the switch positions correspond to 0 and 1) in binary sequence will give an output voltage that increases by equal amounts. This means that the output is far more continuous than the binary input.

(E) Subtractor

All the resistors in the circuit of fig. 61.54 (opposite) have the same value. The output of the first op amp is $-V_1$. The second op amp sums its two inputs and inverts the result. The output is given by

$$V_o = -(V_2 - V_1) = V_1 - V_2$$

V_o is therefore the second input *subtracted* from the first input.

(F) Integrator

Assuming that the op amp takes zero current, I in resistor R is the charging current for capacitor C in the circuit of fig. 61.55(a). This *capacitor* provides the negative feedback. (Cf. the inverting amplifier where a *resistor* was used.) If X is a virtual earth point,

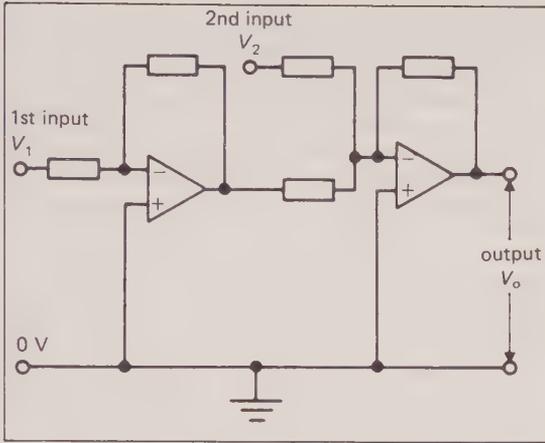


Fig. 61.54. The op amp subtractor.

$$I = \frac{V_i}{R} = \frac{dQ}{dt} \quad \left[\begin{array}{l} Q = C(0 - V_o) \\ \frac{dQ}{dt} = -C \frac{dV_o}{dt} \end{array} \right]$$

$$= -C \frac{dV_o}{dt}$$

$$\therefore \frac{dV_o}{dt} = -\left(\frac{1}{CR}\right)V_i$$

If V_i is steady, integrating gives

$$V_o = -\left(\frac{1}{CR}\right)V_i \cdot t + \text{constant}$$

The constant is zero if the capacitor is uncharged initially,

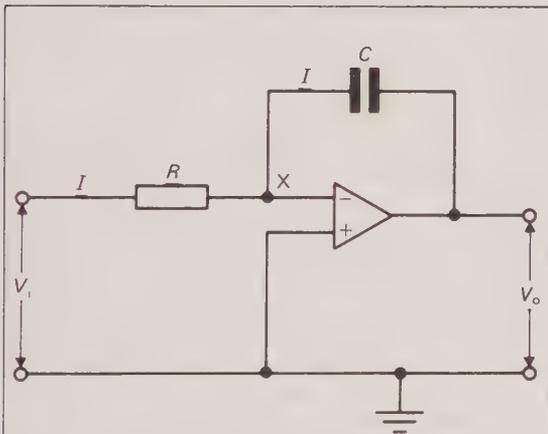
$$\therefore V_o = -\left(\frac{1}{CR}\right)V_i \cdot t$$

This shows that V_o rises steadily with time t until the op amp is saturated. V_i is therefore *integrated* over time t to give V_o , a **ramp voltage** waveform (fig. 61.55(b)) of gradient $-(1/CR)V_i$. Smaller values of C and R would give a steeper ramp. Note that V_o and the gradient are positive if V_i is negative. The *time-base* of a c.r.o. uses a ramp voltage to move the spot across the screen at a steady rate.

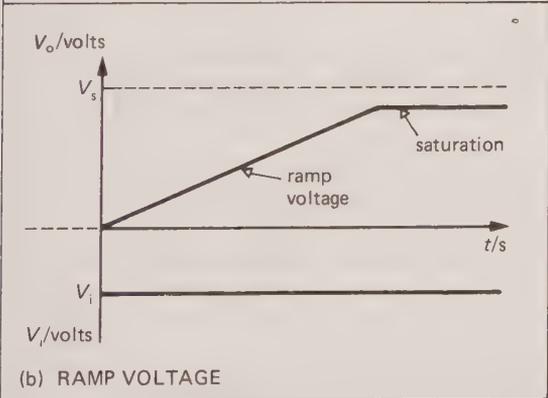
(G) Oscillator

As with the transistor oscillator (p. 486), the op amp oscillator requires controlled **positive feedback** from output to input which is *in phase* with the input and which ensures that *just enough* energy is provided from the power supply to overcome resistive and dielectric energy losses. The oscillator provides its own input and changes the d.c. from the power supply into undamped a.c. electrical oscillations which are sinusoidal. If there is too much positive feedback, the circuit becomes unstable and the amplitude of oscillation increases. **Acoustic feedback** in a public address system can lead to a piercing squeal. This is because electrical oscillations of large amplitude build up when sound is fed back from the loudspeakers to the microphone. The RC network in fig. 61.56 (overleaf) acts as an a.c. voltage divider because of the phase differences of the p.d.s across the components. It can be shown that when $R_2 = 2R_1$, the voltage output from the network is *in phase* with its input at *one* frequency f where

$$f = \frac{1}{2\pi RC}$$



(a) INTEGRATOR CIRCUIT



(b) RAMP VOLTAGE

Fig. 61.55. The op amp integrator.

The network applies positive feedback to the *non-inverting* input of the op amp. R_1 and R_2 are used

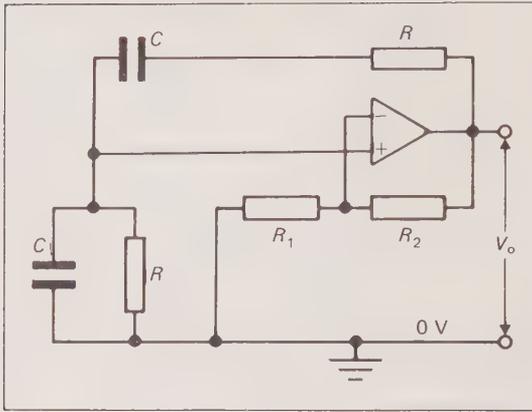
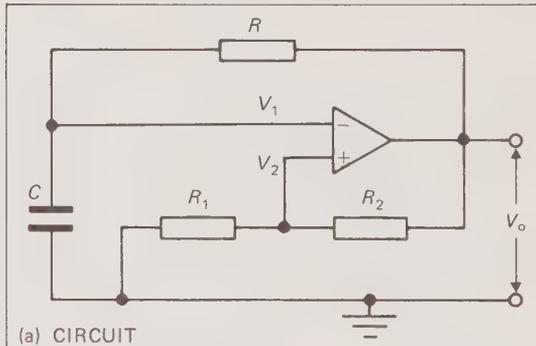
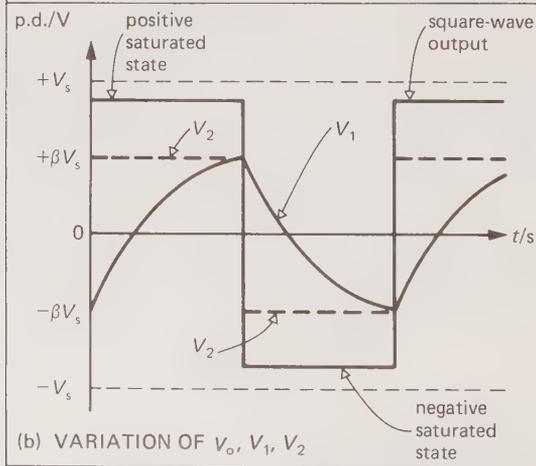


Fig. 61.56. The op amp oscillator circuit.

to supply negative feedback to the *inverting input*. The frequency of the output V_o can be varied if R and C are variable, but note that the two R values stay the same and so do the two C values. A higher R results in smaller currents so that the



(a) CIRCUIT



(b) VARIATION OF V_o , V_1 , V_2

Fig. 61.57. The op amp as an astable multivibrator.

capacitors gain or lose charge more slowly. A higher C results in the p.d. across the capacitor rising and falling more slowly for a given current.

The circuit of fig. 61.57 shows how an op amp can operate as an **astable multivibrator**. It makes use of the saturated *comparator* action and positive feedback. When V_o is positive, a fraction $+\beta V_o (= +\beta V_s)$ is fed back as V_2 , the non-inverting input, where $\beta = R_1/(R_1 + R_2)$. V_o is also fed back via R to the inverting input so that V_1 rises exponentially as C is charged. When $V_1 > V_2$, the op amp switches into negative saturation so that V_o approaches $-V_s$, and the positive feedback makes $V_2 = -\beta V_o = -\beta V_s$. C now charges up in the opposite direction so that V_1 falls exponentially. When $V_2 > V_1$, the op amp switches back to positive saturation with V_o approaching $+V_s$. The circuit therefore switches between *high* and *low* outputs, producing *square waves*, at a frequency that depends upon the time constant CR .

The op amp oscillators are used as *audiofrequency* signal generators, and the RC circuits can be adapted to produce various waveforms. Transistor oscillators, using LC circuits, generate *radio frequencies* and have been used in radio and television transmitters and receivers.

61.19 ANALOGUE AND DIGITAL ELECTRONICS

Summary

In electronics, simple *building blocks* can be assembled into *systems*. Components such as transistors lead to units such as amplifiers and oscillators. Large numbers of electronic units (built mostly as *integrated circuits*) are then combined to form complex electronics systems such as computers, televisions, telephone exchanges and compact disc players. Many different variables (such as temperature, light intensity, humidity, flow rate) can be monitored electrically so that alarms and automatic control systems can be activated as required. *Control engineering* finds application in many walks of life such as road and rail transport, flight navigation, medicine, energy distribution. In *analogue* or *linear circuits*, the output signal varies *continuously* as a linear function of the input signal. *Digital circuits* respond to *two discrete levels* of input voltage and produce two discrete levels of output voltage. The electrical signals are in the form of pulses with *high* or *low* voltage levels.

Integrated circuits (ICs)

These are of fundamental importance in building electronic systems. Each IC is a circuit of many (sometimes $\sim 10^5$) interconnected electronic cir-

cuit elements such as transistors, diodes, resistors and capacitors on a single *chip* of silicon about 5 mm square. Their manufacture involves the use of a wafer of exceptionally pure silicon on which a circuit pattern is formed. The appropriate p- and n-regions are then formed by doping, and the components are connected with aluminium strips. After testing, a wafer is cut into separate chips, each of which is mounted in a protective capsule with projecting connectors. *Analogue* or *linear* ICs, e.g. op amps, deal with continuously changing input and output signals. *Digital* ICs contain switching circuits dealing with high or low input and output signals. TTL (transistor–transistor logic) ICs use *junction* transistors whereas CMOS (complementary metal oxide semiconductor) ICs use *field-effect* transistors. CMOS circuits have lower power consumption and wider supply voltage range, and their outputs can drive a greater number of other circuits. TTL circuits provide faster switching speeds, have greater output current drive capability and are less susceptible to damage by static charges.

The advantages of ICs over conventional circuits of *discrete* components include small size, reliability, robustness, low power consumption and cheapness (volume production overcomes high design and development costs). The main disadvantages are that current and voltage sizes are restricted, and certain components such as transformers cannot be produced on ICs. They therefore deal with *information processing*, and use discrete components at their output stages. An IC is regarded as a building block and it is an understanding of the function and characteristics of the *system*, rather than the individual components, that is required.

Analogue and digital data

Although most input transducers, such as microphones, produce continuously varying analogue signals, digital **data** (information) has two main advantages over analogue data.

(a) Long-distance transmission of signals involves amplification because of attenuation and electrical interference (*noise*). With analogue signals the noise is amplified as well as the signal, leading to distortion of the *waveform*. With digital signals, only the presence or absence of a *pulse* is required.

(b) Digital systems are easier to design and build so that digital signals are more appropriate for modern telecommunications and data processing.

A process such as **pulse code modulation** (PCM) is required to change an analogue signal into a digital one. The analogue waveform is sampled at regular intervals and converted into a series of pulses by a *coder*. A *decoder* is required at the output of the system to interpret the transmitted train of pulses.

Analogue computers

In an electronic analogue computer, the electrical voltages produced are *analogous* to the physical variables being computed, e.g. the decaying oscillations of a loaded spring can be represented by a decaying sinusoidal voltage. The behaviour of physical systems can therefore be simulated by the analogue computer, and the advantage of this approach is that it is relatively easy to set up suitable circuits. For example, an electronic model can be planned for a system such as a new design of car. The effects of factors such as wind speed and loading can be investigated by using the computer to solve complex differential equations. This will be far cheaper, quicker and safer than carrying out experiments on a newly-built, untested car. Op amps carry out the main operations on the variables involved.

Digital computers

Digital computing has, to a great extent, taken over from analogue computing because

- (a) digital circuits are more robust, and are now cheaper and more compact, and
- (b) analogue inputs can now be changed reliably into digital signals, and then back again into analogue outputs.

A digital computer consists of a number of functional units:

- (a) the *input unit* where the computer reads the incoming numbers and instructions, and passes the information to the appropriate part of the computer,
- (b) the *memory unit* where information and instructions are stored,
- (c) the *central processing unit* (CPU) which consists of an *arithmetic and logic unit* (ALU) to perform the calculations and operations, *shift registers* for temporary storage, and a *control unit* consisting of a timer and program counter, and
- (d) the *output unit* which is where the results emerge.

A *microprocessor* (MPU) is a single chip which behaves as a very small CPU, dealing with the arithmetic, logic and control functions of a digital computer. A complete microcomputer would also need a memory and an input/output unit. Microprocessors are used in devices such as calculators and personal computers, and are used to control events such as traffic light changes at a road junction which has variable traffic density.

61.20 BRITISH STANDARD SYMBOLS

Some of the British Standard logic gate symbols are shown in fig. 61.58. These are in less common use than their American equivalents but note that, in both systems, the presence of a circle indicates an inverting gate.

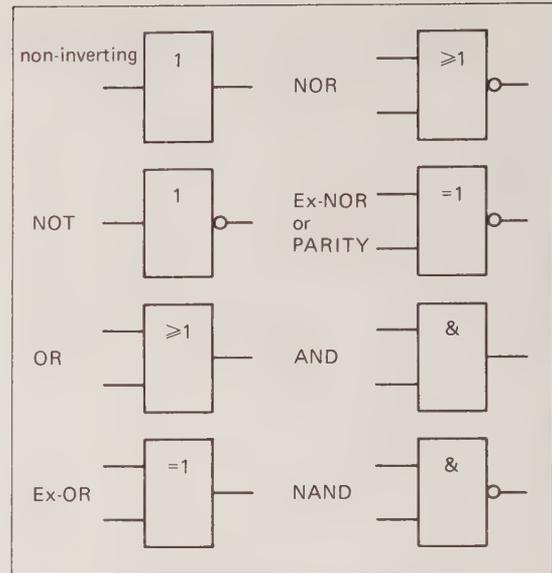


Fig. 61.58. The British Standard logic gate symbols.

62 THE PHOTOELECTRIC EFFECT

62.1	DESCRIPTION OF THE PHOTOELECTRIC EFFECT	503
62.2	EINSTEIN'S PHOTOELECTRIC EQUATION	504

62.3	MILLIKAN'S EXPERIMENT TO VERIFY THE EQUATION	505
62.4	PHOTOCELLS	506

62.1 DESCRIPTION OF THE PHOTOELECTRIC EFFECT

The first observations were made by *Hertz* in 1887.

The effect consists of the emission of electrons from the surface of a metal when electromagnetic waves of high enough frequency fall on the surface.

In fig. 62.1,

(a) if the electroscope is *negatively* charged, the leaves collapse steadily, but

(b) if it is *positively* charged, the leaves show no change of deflection.

In 1899 *Thomson* and *Lenard* showed that the liberated negative ions (**photoelectrons**) have the same (e/m_e) as any other electrons.

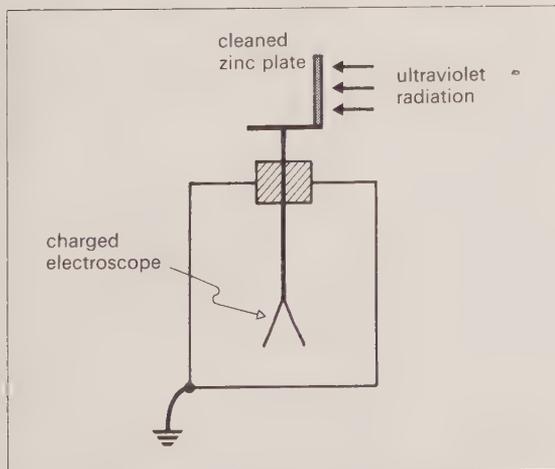


Fig. 62.1. To demonstrate the photoelectric effect for zinc.

Fig. 62.2 shows the apparatus used to study the effect quantitatively. The galvanometer is a very sensitive ammeter (such as the d.c. electrometer described on p. 364).

The results of such experiments are sometimes called the **Laws of Photoelectric Emission**.

I When the incident light is monochromatic, the number of photoelectrons emitted per second is proportional to the light intensity I' . Such an emission occurs effectively *instantaneously*.

II The k.e. of the emitted electrons varies from 0 to a maximum value. This *definite maximum* and the electron energy distribution both depend only on the *frequency* of the light, and not on its intensity.

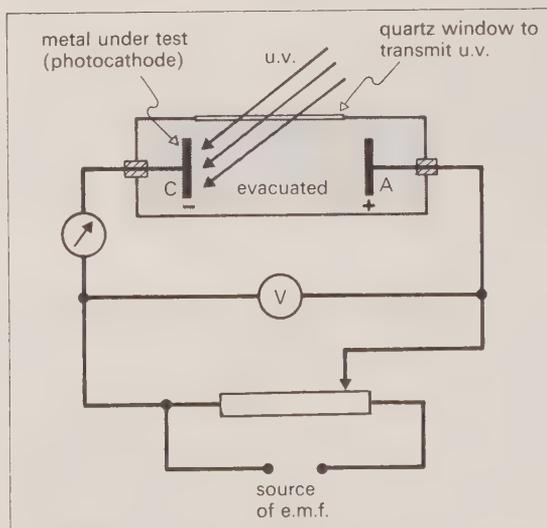


Fig. 62.2. Investigation of the photoelectric effect.

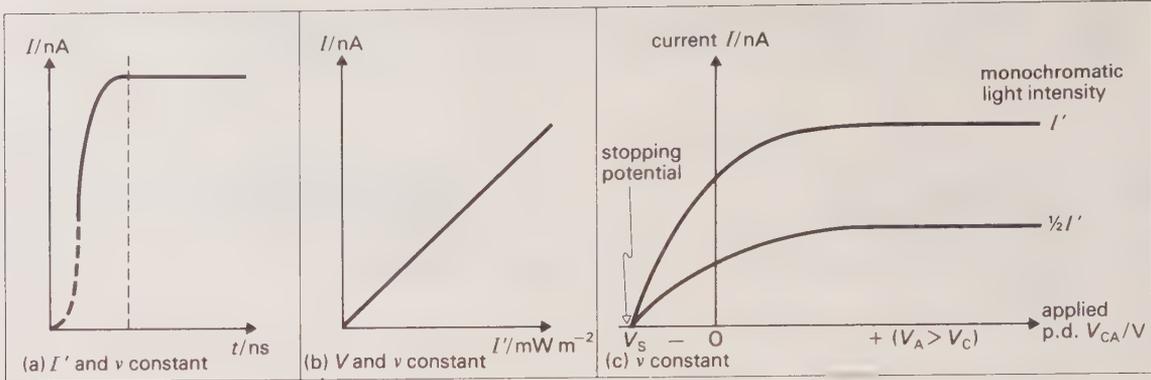


Fig. 62.3. Results of an investigation of the photoelectric effect.

III Electrons are not emitted when the light has a frequency lower than a certain *threshold* value ν_0 . The value of ν_0 varies from metal to metal.

Some of these results are demonstrated by the graphs of fig. 62.3.

The curved part of fig. 62.3(c) indicates that the photoelectrons are emitted with variable speeds.

Difficulties of a wave theory

If we consider light to be *wholly wave-like*, we would expect a uniform distribution of energy over the whole area on which the light is incident. We then cannot explain why

(a) the effect is only observed when the light frequency reaches a value ν_0 , regardless of its intensity,

(b) the electron emission is instantaneous even if very faint light is used (on the assumption that wave energy is uniformly distributed amongst the free electrons we would predict time delays of the order of 10^3 s),

(c) the maximum speed of the emitted electrons is independent of the intensity of the incident radiation.

62.2 EINSTEIN'S PHOTOELECTRIC EQUATION

Photons and the quantum theory

(a) *Planck* (1900), in an attempt to explain the distribution of energy in the black-body spectrum (p. 248), suggested that when radiation was emitted or absorbed, the emitting or absorbing oscillator always showed a *discrete sudden change of energy* ΔE . ΔE is related to the radiation frequency ν by

$$\Delta E = h\nu$$

$$\Delta E [J] = h [J s] \cdot \nu \left[\frac{1}{s} \right]$$

h is called the **Planck constant**, and has the value 6.6×10^{-34} J s. It has dimensions of $[ML^2T^{-1}]$.

For visible light the **quantum** of energy, or **photon**, carries energy

$$\begin{aligned} \Delta E &= h\nu \\ &= (6.6 \times 10^{-34} \text{ J s}) \times (10^{15} \text{ s}^{-1}) \\ &\approx 10^{-19} \text{ J} \end{aligned}$$

The exact value of ΔE depends on the frequency (colour) of the light, but is so small that for many purposes we regard energy emission and absorption as continuous processes.

Experiments on black-body radiation enabled *Planck* to calculate a value for h .

(b) *Einstein* (1905) extended *Planck's* original idea by suggesting that electromagnetic waves were not only *generated* discontinuously, but that they could exhibit particle behaviour while being *absorbed*.

Explanation of the photoelectric effect

Suppose such a photon collides directly with, and imparts *all* its energy to, an electron in a metal *surface*. Using the law of conservation of energy

$$\begin{aligned} \left(\begin{array}{c} \text{energy of} \\ \text{photon} \end{array} \right) &= \left(\begin{array}{c} \text{energy needed to} \\ \text{remove electron} \\ \text{from metal} \end{array} \right) + \left(\begin{array}{c} \text{k.e. imparted} \\ \text{to ejected} \\ \text{electron} \end{array} \right) \\ h\nu &= W + \frac{1}{2} m_e v_m^2 \end{aligned}$$

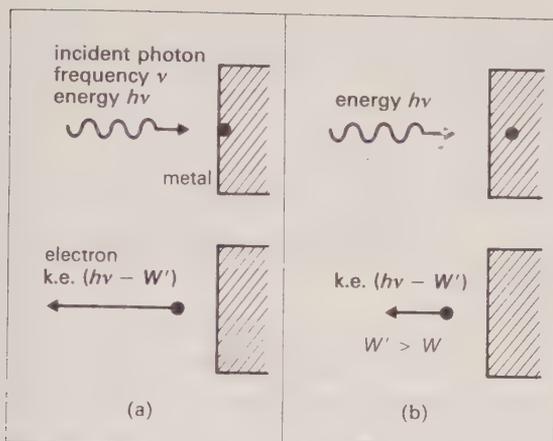


Fig. 62.4. Ejected electrons have variable speeds: (a) surface electron, (b) electron in metal. (Note that in more advanced study a radically different model is used.)

v_m is the maximum speed of ejection. Using one simple model, we imagine those electrons that are in the surface to have the least values of W , the *work function energy* (p. 473). This is illustrated in fig. 62.4.

For the surface electrons of a given metal, W always has the same value. (Note that in more advanced work this simple model is replaced by one using more sophisticated concepts.)

We explain the difficulties encountered by the wave theory (above) as follows:

$$\begin{aligned} (a) \quad \frac{1}{2} m_e v_m^2 &= h\nu - W \\ &= h\nu - h\nu_0 \quad (W = h\nu_0) \end{aligned}$$

When $\nu = \nu_0$, no individual photon imparts enough energy to an individual electron to cause ejection. (An increase in light intensity does not cause an increase in photon energy.)

(b) Electron emission is the result of a direct collision between an electron and photon, so there is no time delay before emission starts. (A photon cannot collide with more than *one* electron. This is because it has zero rest mass, and so according to relativity theory it ceases to exist once it no longer moves with the speed of light.)

(c) An increase in light intensity results in an increase in the number of photons arriving per second, but not in the energy of individual photons.

$$\frac{1}{2} m_e v_m^2 = h\nu - W$$

is called **Einstein's photoelectric equation**.

► These ideas lead to the concept of the **wave-particle duality of light**—under some circumstances it behaves as a wave motion, and under others like a particle (or photon) (p. 287).

Matter shows analogous behaviour (p. 288).

62.3 MILLIKAN'S EXPERIMENT TO VERIFY THE EQUATION

If a metal has a positive potential, more energy must be given to an electron to enable it to escape. If the metal potential is V , this additional energy is eV . This situation is illustrated in fig. 62.5.

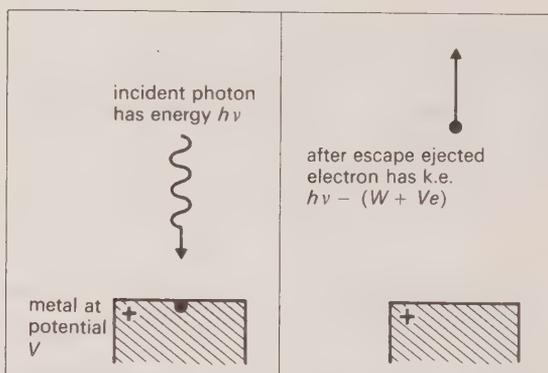


Fig. 62.5. The effect of raising the metal to a positive potential.

The **stopping potential** V_s is the potential to which the metal must be raised so that when electrons are ejected from the surface by suitable radiation, they are all just attracted back again.

Since in general $\frac{1}{2} m_e v_m^2 = h\nu - (W + Ve)$

$$v_m = 0 \quad \text{when} \quad eV_s = h\nu - W$$

When $V_s = 0$, $h\nu_0 = W$

where ν_0 is the **threshold frequency**.

When photoelectric emission has been occurring, the loss of electrons will itself cause the potential of an insulated metal to become positive: emission will cease when the stopping potential V_s has been reached.

Using apparatus similar to that of fig. 62.2 (but with the terminals reversed) *Millikan* measured the potential difference V_{CA} such that the current was just reduced to zero for different values of ν .

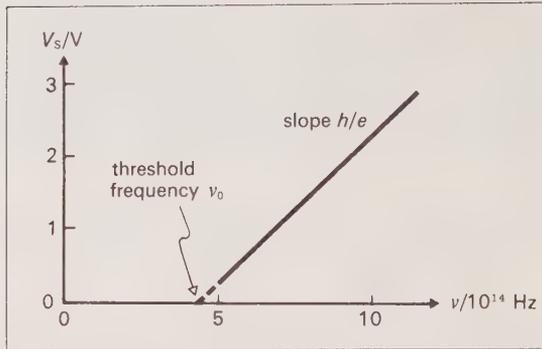


Fig. 62.6. The result of Millikan's experiment when a freshly-cut sodium surface is used.

The stopping potential

$$V_s = V_C - V_A \quad (V_C > V_A)$$

Gain of electric p.e. (QV) equals loss of electron k.e.

$$eV_s = h\nu - h\nu_0$$

$$V_s = \left(\frac{h}{e}\right) \cdot \nu - \left(\frac{h}{e}\right) \cdot \nu_0$$

If we plot a graph of V_s against ν (fig. 62.6) we find a straight line whose

- slope = h/e , and
- intercept = ν_0 (the value of ν when $V_s = 0$).

Einstein's photoelectric equation was confirmed by

- the straight-line graph, and
- the fact that the slope h/e , and the intercept $\nu_0 (=W/h)$ on the ν -axis, both enabled h to be calculated. (Millikan had already measured e .)

The determined value of h agreed with that found from Planck's radiation formula (p. 248).

62.4 PHOTOCELLS

(A) Photoelectric emission cell

In fig. 62.7 the material of the cathode can be chosen to respond to the frequency range over which the cell will operate. The response can be made proportional to the light intensity.

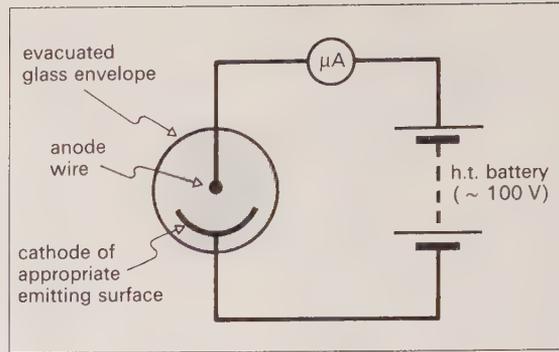


Fig. 62.7. A photoemission cell in a circuit.

Uses

The cell can be used in any situation where a beam of light which was falling on a cell is broken, e.g.

- to count vehicles or items on a factory belt,
- as a burglar alarm, or
- to open doors.

(B) Photoconductive cell

An internal photoelectric effect may free charge-carriers in a material that is otherwise an insulator, and thereby increase its conductivity by as much as 10^4 when it is illuminated. A current will flow if the material is in a circuit with a source of e.m.f.

Uses

- Detection and measurement of infra-red radiation ($\lambda \sim 10^{-6}$ m).
- Relays for switching on artificial lighting, such as street lighting.

(C) Photovoltaic cell

This may consist of a sandwich of copper, copper oxide and a thin layer of translucent gold. An e.m.f. capable of giving a current of 1 mA can be generated when the gold is illuminated.

Uses

The cell needs no source of e.m.f., and is frequently used as an exposure meter in photography.

63 X-RAYS AND THE ATOM

63.1	PRODUCTION OF X-RAYS	507
63.2	PROPERTIES OF X-RAYS	507
63.3	POSITIVE RAYS	508
63.4	DEFINITIONS	509
63.5	MASS-ENERGY EQUIVALENCE	509
63.6	PHOTON-ELECTRON INTERACTIONS	510

63.7	THE NUCLEUS	511
▶ 63.8	THE BAINBRIDGE MASS SPECTROGRAPH	513
63.9	MODELS OF THE ATOM	513
63.10	THE ORBITAL ELECTRONS	514
▶ 63.11	THE BOHR ATOM	515

63.1 PRODUCTION OF X-RAYS

X-rays were discovered by Röntgen in 1895: they were produced by electrons colliding with the walls or electrodes of a discharge tube.

The Coolidge tube

In fig. 63.1 the electrons emitted from the heated filament are attracted to and focussed on the cooled anode. (Less than $\frac{1}{2}\%$ of the electron k.e. goes to X-rays. The remainder is converted to the internal energy of the anode.) Their rapid deceleration results in the emission of X-rays, an electromagnetic radiation.

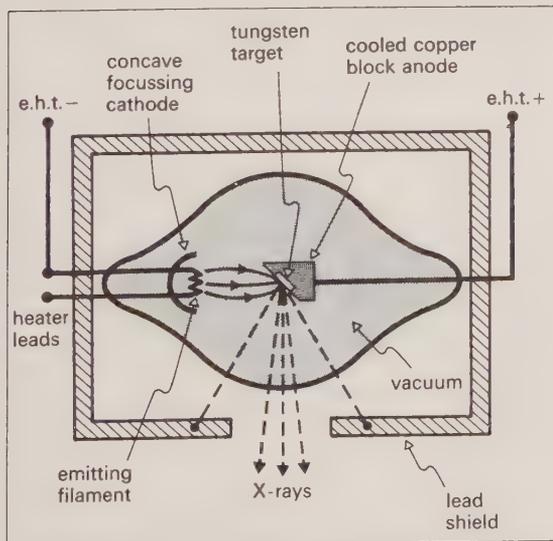


Fig. 63.1. A hard vacuum X-ray tube.

(a) The *intensity* of the X-ray beam is controlled by the filament current, which determines the number of electrons hitting the anode per second.

(b) The *penetrating power (quality)* of the beam is controlled by the p.d. between the anode and cathode. This may be 100 kV or higher.

63.2 PROPERTIES OF X-RAYS

General

(a) They *travel in straight lines*, and are not deflected by electric or magnetic fields. (They are not charged particles.)

(b) They *produce ionization* of a gas when they pass through it. An electroscope charged either positively or negatively is rapidly discharged if X-rays make the surrounding air conducting.

(c) They *expose photographic plates* in the same way as visible light waves.

(d) They cause *fluorescence*.

(e) They cause *photoelectric emission*.

(f) They *penetrate matter*. Three reasons for absorption are

(i) **Compton effect** (in which the quantum is scattered by an electron), \propto number of electrons present,

(ii) photoelectric effect (for quantum energies up to $\sim 1.6 \times 10^{-14}$ J), roughly $\propto Z^4$,

(iii) pair production (p. 510) (for quantum energies $> 1.6 \times 10^{-13}$ J), roughly $\propto Z^2$.

The absorption is thus highest for elements of high *atomic number* (dense materials). This is the basis of radiography (see overleaf).

The wave nature of X-rays

Because X-rays are electromagnetic waves, they possess the usual properties of such waves (p. 286). The typical wavelength is $\sim 10^{-10}$ m, which makes the wave properties difficult to demonstrate.

(a) *Von Laue* (1912) demonstrated that X-rays give a typical *diffraction pattern* when made to pass through a crystal. The pattern is recorded on a photographic plate (p. 311).

(b) The *Braggs* (1913) developed a simpler method of analysing the 'reflected' diffraction pattern from a crystal. Considering each ion in a crystal lattice to act as a secondary source, they developed an equation (the **Bragg equation**, p. 312) which enabled the wavelength λ of the X-rays to be calculated in terms of the separation d of the ion planes.

X-ray spectra

Fig. 63.2 shows a typical X-ray spectrum.

(a) *The continuous spectrum* shows a well-defined minimum wavelength (maximum frequency). This corresponds to an electron losing all its energy in a single collision with a target atom. The longer wavelengths (smaller energies) correspond to a more gradual loss of energy, which happens when the electron experiences several deflections and collisions and so is slowed down more gradually. All or some of the k.e. of the electron is converted into the energy of the photon(s). This radiation is called **bremstrahlung** (braking radiation). All targets show this continuous spectrum.

(b) The k.e. of a bombarding electron = Ve where V is the accelerating p.d.

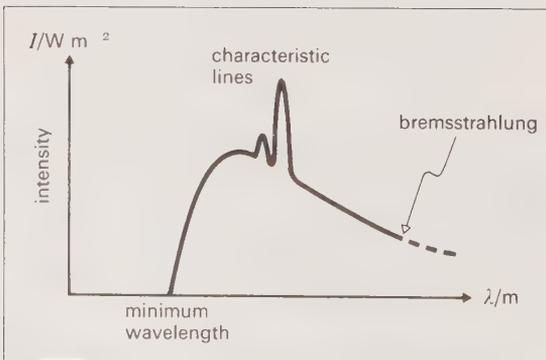


Fig. 63.2. Analysis of the radiation from an X-ray tube.

$$\therefore Ve = hv_{\max} = h \frac{c}{\lambda_{\min}}$$

where v_{\max} is the frequency of the most energetic photon (possessing all the initial k.e. of the colliding electron).

(c) *The line spectrum* is characteristic of the element used for the target in the X-ray tube. It corresponds to the quantum of radiation emitted when an electron changes energy levels very close to a nucleus. (See p. 507 and p. 515.)

Application of X-rays

(a) *Medicine*

(i) *Radiography*—dense materials such as bones absorb the radiation and this produces shadows on the film which can be analysed. This technique requires no lenses or focussing.

(ii) *Radiotherapy*—destruction of cancer cells.

(b) *Industrial radiography*—location of internal imperfections (e.g. in a casting).

(c) *X-ray crystallography*—analysis of structure of organic molecules. This can be extended to analyse fine powders containing minute crystals. X-ray spectrometer measurements (p. 313) lead to the most accurate determination of N_A .

(d) A single crystal can be used as a *monochromator* for X-rays—it can diffract an intense beam of *given* λ into a particular direction.

63.3 POSITIVE RAYS

Cathode rays (p. 468) are the streams of negative particles (electrons) that seem to come from the cathode of the discharge tube. **Positive (canal)**

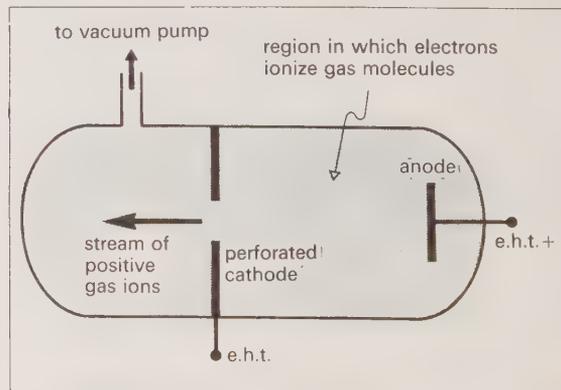


Fig. 63.3. Positive rays in the discharge tube. The pressure in the tube is ~ 0.4 kPa.

rays are the streams of positive ions that pass through a perforated cathode—see fig. 63.3.

Properties of positive rays

(a) They cause fluorescence and affect a photographic plate.

(b) They are related to the gas in the tube. (Cathode rays are *not*.)

(c) They are deflected in magnetic and electric fields, but much less than cathode rays. (The positive ions are more massive.)

(d) Their velocities are variable, since they are accelerated only from the points at which they are formed. (These vary in position between the anode and the cathode.)

Thomson's experiment

Thomson applied parallel fields of B and E which resulted in parabolic traces on a suitably placed photographic plate (p. 470). He deduced that

(a) the value of the specific charge $q = e/m_p$ for the hydrogen ion (a proton) is about 1840 times smaller than e/m_e for the electron, so

(b) if the hydrogen ion and electron carry charges of the same magnitude, the mass of the hydrogen ion is about 1840 times that of the electron.

He was also able to measure the masses of other positive ions by using different gases in the tube. This led to the discovery of **isotopes** (below).

63.4 DEFINITIONS

Models of the atom are discussed on p. 513. Consider a model in which the atom is composed of

(a) a nucleus, which contains neutrons and protons (**nucleons**) only, and

(b) an encircling cloud of electrons.

Then we can define the following terms.

A **nuclide** is a particular **species** of atom (one whose nucleus contains a specified number of protons, and a specified number of neutrons).

The **atomic number** (or *proton number*) Z of an *element* is the number of protons contained in the nucleus. This is equal to the number of electrons in a neutral atom of the element, and so determines the *chemical nature* of an element.

The **mass number** A of an *atom* or nuclide is the number of nucleons in the nucleus. It is by definition an integer.

If two different nuclides have

(a) the same atomic number Z but different mass numbers A , they are called **isotopes**,

(b) the same mass number A but different atomic numbers Z , they are called **isobars**.

(*Isotope* implies *same place* in the periodic table. *Isobar* implies *same weight*, meaning same total mass.) Isotopes are chemically identical (p. 529).

A **unified atomic mass unit** (u) is defined such that the mass of the carbon-12 atom is 12 u . By experiment

$$m_u = 1 u \approx 1.66 \times 10^{-27} \text{ kg}$$

(*N.B.* $m_u = 1 u$ is not really a *unit*, but we find it convenient to express atomic masses relative to this *physical quantity*.)

The **relative atomic mass** is the ratio

$$\frac{\text{mass of an atom}}{\text{mass of a } {}^{12}_6\text{C atom}} \times 12$$

The *mass* of an atom is usually quoted in unified atomic mass units, and is then numerically equal to its relative atomic mass. (It is often incorrectly called its *atomic weight*.) Atoms have masses which are very close to an integral number times m_u . If we assume that the proton and neutron each have a mass very close to 1 u , this is consistent with our simple model of the atom.

In practice this model is oversimplified. As can be seen below, the masses of nucleons are nearly one per cent greater than 1 u , and since some nuclei have more than 200 nucleons, the nearly integral values for atomic masses can be explained only in terms of binding energy (p. 510). The rest masses and rest energies of electrons and nucleons are:

$$\begin{aligned} \text{electron } m_e &= 9.10953 \times 10^{-31} \text{ kg} \equiv 81.8724 \text{ fJ} \\ \text{proton } m_p &= 1.67265 \times 10^{-27} \text{ kg} \equiv 0.150330 \text{ nJ} \\ \text{neutron } m_n &= 1.67495 \times 10^{-27} \text{ kg} \equiv 0.150537 \text{ nJ} \end{aligned}$$

Representation by symbol

A given nuclide is represented by



in which

- (a) the superscript A is the mass number,
- (b) the subscript Z is the atomic number, and
- (c) X is the **chemical symbol**.

X and Z present the same information. Some examples are given on p. 526.

63.5 MASS-ENERGY EQUIVALENCE

The masses of atomic particles are small. For a given energy, these particles may be accelerated to such high speeds that relativistic mechanics must

be used. The mass of a particle m moving at speed v relative to an observer would be measured by him to be

$$m = \frac{m_0}{\sqrt{[1 - (v/c)^2]}}$$

where m_0 is its rest mass. Its relativistic momentum is therefore

$$p = mv = \frac{m_0 v}{\sqrt{[1 - (v/c)^2]}}$$

The theory of relativistic mechanics gives the kinetic energy of a particle to be

$$E_k = (m - m_0)c^2$$

(Note that this is *not* equal to the classical value $\frac{1}{2}mv^2$.) If we write the total energy as E , then

$$\begin{aligned} mc^2 &= E_k + m_0c^2 \\ &= E \end{aligned}$$

The *Einstein* relationships are

$E = mc^2$	total energy
$E_0 = m_0c^2$	rest energy

The laws of energy and mass conservation are thus combined in the *law of conservation of mass-energy*.

The relationship between total energy and momentum is

$$E^2 = E_0^2 + (pc)^2$$

Some implications

These relativistic equations show that

(i) when $v \ll c$, so that *Newtonian* mechanics is a good approximation,

$$m \approx m_0, \quad p \approx m_0v, \quad E_k \approx \frac{1}{2}m_0v^2, \quad E_k \ll E_0$$

(ii) when v approaches c , so that relativistic mechanics is necessary,

$$m \gg m_0, \quad E \gg E_0, \quad p \approx \frac{E}{c}, \quad E_k \approx E$$

and

(iii) for a particle of zero rest mass,

$$m_0 = 0, \quad E = pc, \quad E_k = E, \quad v = c$$

Unbound and bound systems

(a) In an **unbound** system, the rest mass of the composite system is greater than the sum of the

rest masses of the separated particles by an amount equal to the k.e. of the amalgamating particles at combination.

(b) In a **bound** system, the rest mass of the composite system is less than the sum of the rest masses of the separated particles by an amount called the **binding energy** E_b . If a system of rest mass M_0 is split into two particles of rest mass m'_0 and m''_0 by adding energy equal to E_b , then

$$E_b = (m'_0 + m''_0)c^2 - M_0c^2$$

A measurable mass difference is obtained only when one is dealing with nuclear forces. The total mechanical energy E_m of a system of particles that have mutual attraction, is taken by convention to be zero when the particles are at rest and infinitely separated. Thus when particles are bound, E_m becomes negative; that is energy would have to be added to the system to separate the particles again completely, and thus to increase the energy to zero.

63.6 PHOTON-ELECTRON INTERACTIONS

In all such interactions the laws of conservation of charge, mass-energy and relativistic momentum can be applied, and the particle-like nature of electromagnetic radiation is emphasized. The photons have energy $h\nu$, momentum h/λ and effective mass $h/c\lambda$. These interactions usually involve high-energy photons and electrons.

(a) *The photoelectric effect* (p. 503). A photon is annihilated on colliding with a bound electron. Most of the photon's energy is transferred to the electron which is ejected, whereas most of the photon's momentum is transferred to the object to which the electron was bound. (This effect cannot, therefore, take place with a free electron.)

(b) *The Compton effect*. A photon collides with a free or lightly-bound electron, giving the electron kinetic energy and causing it to recoil. A second (scattered) photon of lower energy and therefore greater wavelength is created.

(c) *Pair production*. A photon passes near a massive nucleus and its energy is converted into matter. This cannot happen spontaneously in free space where it is not possible to satisfy simultaneously the conservation laws of mass-energy, momentum and electric charge. The photon energy is converted into

(i) the rest mass of the electron-positron pair, and

(ii) the k.e. of the particles so formed.

The equation for this can be written

$$h\nu = 2m_0c^2 + [E_k^+ + E_k^-]$$

The minimum energy of the photon for pair production is 1.64×10^{-13} J, and it can therefore be achieved only by γ -photons or X-ray photons.

(d) **X-ray production** (p. 507). An electron loses kinetic energy through collisions and deflections near massive particles. Some of the energy is converted into the energy of one or more photons in the production of bremsstrahlung. (Most of the k.e. is converted into the internal energy of the target.)

(e) **Pair annihilation**. A positron loses its kinetic energy by successive ionization, comes to rest and combines with a **negatron** (negative electron). Their total mass is converted into two oppositely directed photons (**annihilation radiation**), and the process is thus the reverse of pair production. As $h\nu_{\min} = m_0c^2$, the total energy available is 1.64×10^{-13} J and, to conserve momentum, each quantum has energy 8.2×10^{-14} J. They move off in opposite directions.

63.7 THE NUCLEUS

The radius of an atom is $\sim 10^{-10}$ m. The radius of the nucleus of an atom $\sim 10^{-15}$ m.

The electron cloud shields the nucleus, so that its behaviour is generally independent of the environment (temperature, pressure, chemical combination, etc.). Information about the nucleus comes from

- (a) *ejected particles*, such as those emitted during radioactivity,
- (b) *injected particles*, as when the nucleus is bombarded by artificially accelerated ions, and
- (c) *spectra* of all kinds.

Moseley's experiments on X-ray spectra (1913)

Using a Bragg X-ray spectrometer, Moseley measured the frequency ν of corresponding characteristic line spectra for different elements. He plotted the graph of $\sqrt{\nu}$ against Z (which we now call the *atomic number*) as shown in fig. 63.4.

The graph is described by

$$\sqrt{\nu} = (\text{constant})(Z - b)$$

where b is a constant.

Deductions

(a) Each nucleus has a *characteristic number* Z in addition to its mass number A , which increases by 1 from one element to the next.

(b) Gaps in Moseley's graph were explained by elements not then discovered.

(c) The number Z measured by this method follows the numbering of elements in the periodic table. Anomalies arising from an atomic *mass* order disappear.

(d) Comparison with Bohr's theory of the hydrogen atom shows that Z is numerically equal to the positive charge on the nucleus.

Moseley's experiment is a *direct way of measuring atomic number*.

Rutherford's experiment

Evidence for the existence of the nucleus comes from scattering experiments carried out for Rutherford by Geiger and Marsden. (See fig. 63.5, overleaf.)

Detailed measurements of the proportion of α -particles that suffer a deviation θ

- (a) confirm that each atom has a nucleus in which its mass is concentrated, and
- (b) enable the size and charge of the nucleus to be measured.

The experiment also provides confirmatory evidence that *Coulomb's Law* (which is used for the calculation) holds for atomic dimensions.

An *upper limit* for the size of the nucleus can be estimated by equating the initial k.e. of an α -particle with the p.e. of the α -particle/gold nucleus system at closest approach, in the case of a head-on collision. This assumes that the gold nucleus acquires very little k.e. from the collision.

$$\text{k.e. of } \alpha\text{-particle} = (2e)V$$

where V is the accelerating p.d.

$$\text{p.e. of } \alpha\text{-particle/gold nucleus system} = \frac{(2e)(79e)}{4\pi\epsilon_0 r}$$

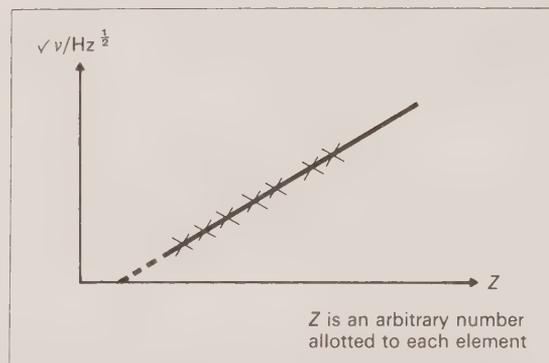


Fig. 63.4. The graph which gives the clue to atomic number.

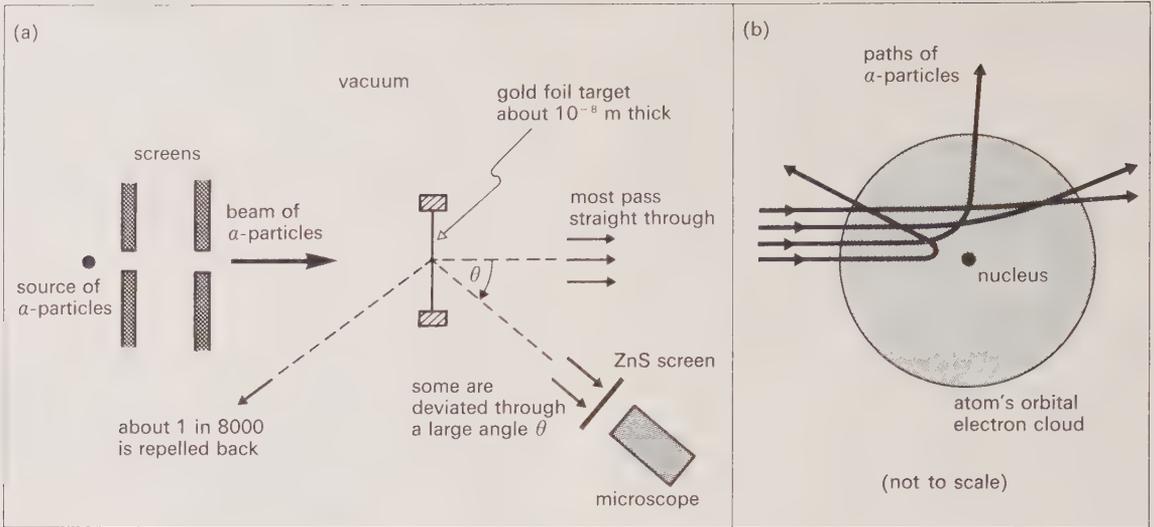


Fig. 63.5. The schemes of Geiger and Marsden's experiment: (a) the general idea, and (b) what happens on a microscopic scale.

where r is the distance of closest approach. r can be evaluated when the two energies are equated because it is the only unknown quantity.

Nuclear stability

The nuclear mass M_n is always less than the sum of the rest masses of the constituent nucleons because the p.e. associated with the (strong) interactions between nucleons is so great. If a nucleus has Z protons and N neutrons ($N = A - Z$), then

$$M_n < Zm_p + Nm_n$$

As it is the masses of neutral atoms M_a that are measured, then

$$M_a < Zm_p + Nm_n + Zm_e$$

or $M_a < ZM_H + Nm_n$ *

When Z protons and N neutrons combine to form a nucleus, they give up energy which is revealed as a mass reduction—this is the binding energy E_b of the nucleus (p. 510).

$$M_a + \frac{E_b}{c^2} = ZM_H + Nm_n$$

The greater the energy that must be supplied to unbind the system (i.e. to split it into its constituent particles), the more stable is that system.

The binding energy per nucleon E_b/A is a more useful measure of stability. It corresponds to the

atomic ionization potential as it is the energy needed to remove an average nucleon from the nucleus.

$$\frac{E_b}{A} = \left(\frac{ZM_H + Nm_n - M_a}{A} \right) c^2$$

Fig. 63.6 shows that, after the initial irregularities for nuclei with few nucleons (note the stability

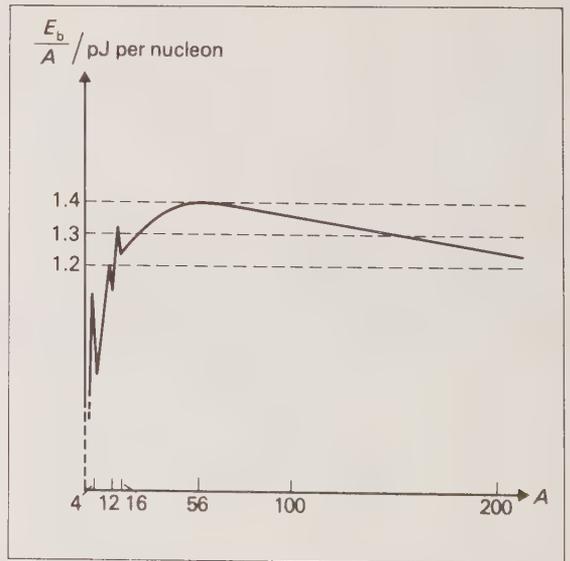


Fig. 63.6. Average binding energy per nucleon for naturally occurring nuclides.

* This step is justified because the strong (nuclear) interaction between protons and neutrons is about 10^6 times as strong as the electromagnetic interaction between protons and electrons.

peaks for ${}^4_2\text{He}$, ${}^{12}_6\text{C}$ and ${}^{16}_8\text{O}$), the curve reaches a maximum at about ${}^{56}_{26}\text{Fe}$ and then drops gradually towards the massive nuclei. When $A > 20$, the total binding energy is roughly proportional to A . ($E_b/A \approx 1.3$ pJ/nucleon, on average.)

Nuclear forces tend to give constant E_b/A and to make the value of $Z = N$. Nevertheless some deviation is caused because

(a) electrostatic repulsion between protons in heavy nuclei tends to reduce E_b/A and to make $Z < N$, and

(b) surface energy in light nuclei tends to reduce E_b/A .

All nuclear changes happen in an attempt to achieve maximum stability by increasing E_b/A . They therefore tend to minimize the mass of the system, since $A = \text{constant}$.

► 63.8 THE BAINBRIDGE MASS SPECTROGRAPH

(A) The velocity selector (fig. 63.7(a))

Suppose the ions carry a charge Q , and have mass m . Only those ions which experience zero net force will emerge from the last slit. These ions all have the *same speed* v such that

$$EQ = B'Qv$$

$$v = \frac{E}{B'}$$

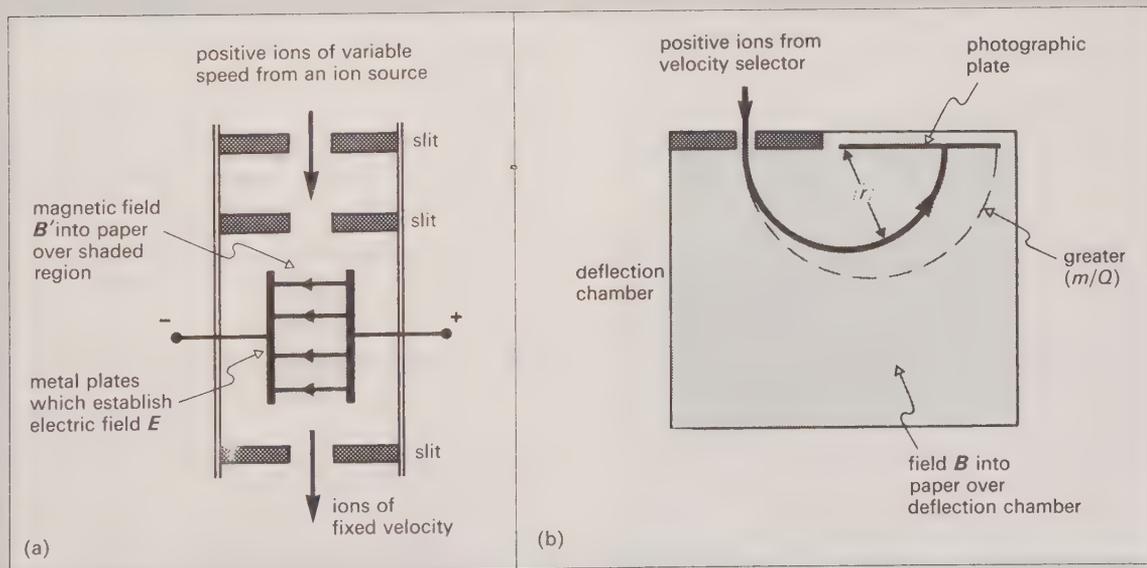


Fig. 63.7. Essentials of the Bainbridge mass spectrograph: (a) the velocity selector, and (b) the magnetic deflection through π rad.

(B) The magnetic deflection (fig. 63.7(b))

For the semi-circular path of radius r

$$F = ma$$

gives

$$BQv = \frac{mv^2}{r}$$

so

$$r = \frac{m}{Q} \cdot \frac{v}{B}$$

Since all the positive ions move at the same speed v in the same magnetic field B

$$r \propto \frac{m}{Q}$$

For singly charged ions

$$r \propto m$$

The ions form dark lines on the plate. Their positions enable m to be calculated once the spectrum has been calibrated by admitting gases which produce ions of known mass.

When ions of isotopes are present several lines will be produced in slightly different positions owing to the different masses of the nuclei. From the degree of exposure of the photographic plate the relative abundance of isotopes can be deduced.

63.9 MODELS OF THE ATOM

For a discussion of the term *model*, see p. 4.

(A) Thomson's plum pudding model (1904)

Thomson's model suggested that the atom's positive charge was spread throughout its whole volume, and that the electrons vibrated about fixed centres embedded in the sphere. It gave a qualitative explanation of dielectric polarization (p. 366).

(B) Rutherford's nuclear atom (1911)

Rutherford's model resulted from Geiger and Marsden's experiment: it confined the positive charge to a central core (**nucleus**) of radius $\sim 10^{-15}$ m, around which the electrons were supposed to move in a spherical volume of radius 10^{-10} m. The electrons in orbit account for the atom's volume.

Orbiting electrons are accelerated electric charges, and by classical theory should lose energy by radiating electromagnetic waves: they should spiral into the nucleus.

(C) Bohr's atom (1913)

This is discussed more fully opposite. Bohr applied some of the quantum theory ideas to the Rutherford atom, and this enabled him to make quantitative predictions for hydrogen and singly ionized helium.

His model correlates well with

- (a) the Franck and Hertz experiment (right),
- (b) Moseley's work on X-ray spectra (p. 511),

and in addition can be used to predict

- (1) some spectroscopic results with an accuracy of 1 part in 10^6 , and
- (2) the correct value for the most probable location of the electron orbit in the hydrogen atom.

The Bohr atom has now been replaced by the ideas that come from wave mechanics.

► (D) The wave mechanics atom

The wave mechanics model is too difficult for detailed study in this book. It has the advantage of not making arbitrary assumptions, but can only be described adequately using mathematics. The procedure is (1) to set up a *wave equation* for a particular situation, (2) to solve that equation where possible, and (3) to deduce from the solution the *probability* of finding an electron at a particular distance from the nucleus.

The idea of orbits is replaced by that of an **electron cloud** of particular density, but the energy levels of the Bohr atom remain (p. 516).

63.10 THE ORBITAL ELECTRONS**(A) Barkla's experiment (1911)**

Barkla measured the number of electrons in an atom by observations on the scattering of X-rays. He concluded that

$$(\text{number of electrons}) \sim \frac{1}{2}(\text{mass number})$$

This raised the question of an atomic property Z as distinct from A .

(B) Bohr's theory (1913)

Bohr's theory gave a quantitative prediction of the energy of an orbital electron in the hydrogen atom.

If an electron is to change from an allowed energy level E_2 to a lower level E_1 (fig. 63.8), then the energy difference appears as an electromagnetic wave of frequency ν , where

$$(E_2 - E_1) = h\nu \quad (\text{p. 504})$$

The theory leads to a detailed account of the spectral lines of hydrogen, and the ionization energy.

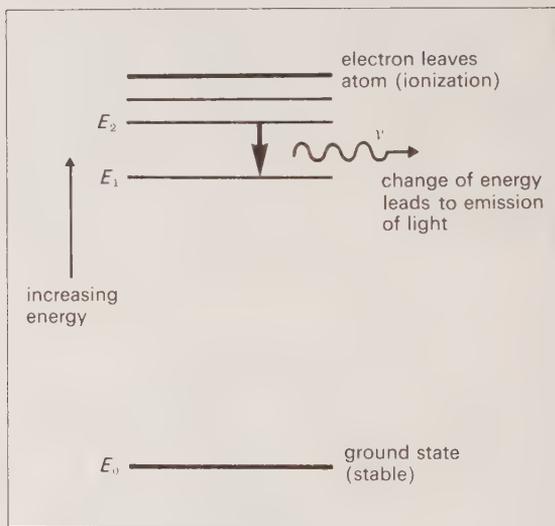


Fig. 63.8. Energy level diagram for hydrogen.

This is indirect evidence for the existence of energy levels in the orbital electrons.

(C) The Franck and Hertz experiment (1916)

The experiment provided further evidence for discrete energy levels within the atom. (See fig. 63.9.)

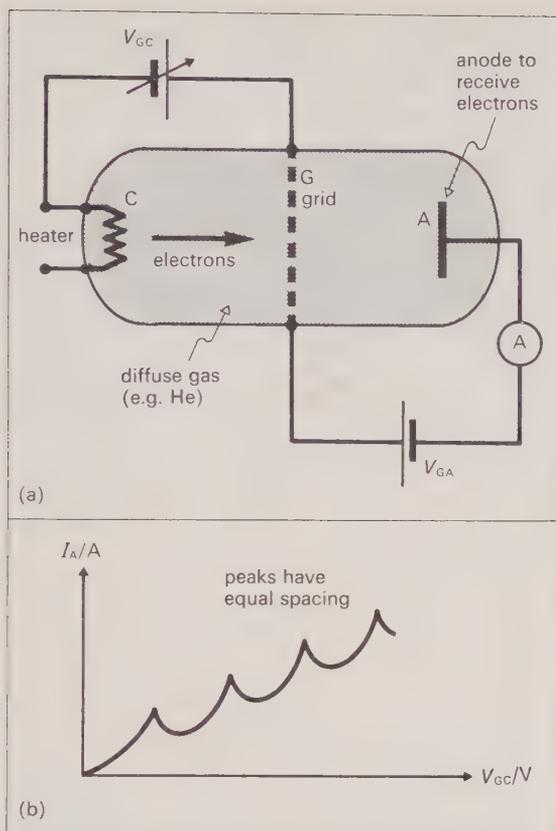


Fig. 63.9. The Franck and Hertz experiment: (a) the apparatus, and (b) the results.

Procedure

(1) Electrons emitted by the cathode are attracted to and pass through the grid. They just reach the anode because $V_{GC} > V_{GA}$.

(2) Increase V_{GC} and plot a graph as shown in fig. 63.9(b).

Interpretation

Electrons colliding with He atoms do not lose energy unless the He atoms are excited (unless the orbital electrons are raised to the next energy level). When they do, those electrons do not reach the anode A, and the current drops. Further increase in V_{GC} leads to an increased current, until another peak is reached, corresponding to a second excitation. Each peak represents an *inelastic collision* with energy exchange between the free electron and the atom.

► (D) The wave mechanics energy levels

Solution of the *Schrödinger* wave equation predicts the number of electrons that can fit into each *shell* of an atom, and describes, using *quantum numbers*, the allowed energies that each electron may have. However, this equation and the *Heisenberg* uncertainty principle indicate that it is impossible to talk about the precise position or orbit of an electron. *Orbitals* are used to describe the regions round the nucleus in which there is the greatest probability of finding a given electron. They can be regarded as a three-dimensional electronic contour. Each orbital accommodates a maximum of two electrons which must have opposed spins. Wave-mechanical calculations show that different shapes and sizes of orbitals correspond to different energy levels. For example, the first two orbitals are spherically symmetrical about the nucleus but have different radii. The next orbitals consist of three dumb-bell shapes of the same size and corresponding energy level, arranged mutually at right angles with the nucleus at their centres. Electrons are predicted to be within these various orbital shapes but their positions cannot be known exactly. If an atom is raised to an *excited state*, an electron will move to an orbital corresponding to a higher energy level.

There is exact agreement between the energy level predictions of wave mechanics and empirical methods such as those based on chemical behaviour and the periodic table.

► 63.11 THE BOHR ATOM

The spectrum of atomic hydrogen

Hydrogen *atoms* in a discharge tube emit characteristic spectral lines whose frequencies ν fit the general formula

$$\nu = cR_H \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right)$$

c is the speed of light, R_H is called the **Rydberg constant** (for hydrogen), and n_1 and n_2 are integers, such that $n_1 < n_2$.

By experiment $R_H = 1.09678 \times 10^7 \text{ m}^{-1}$.

Spectral series

(a) When $n_1 = 1$, $n_2 \geq 2$, the frequencies correspond to the **Lyman series**, which is in the *ultra-violet*.

(b) When $n_1 = 2$, $n_2 \geq 3$, the frequencies correspond to the **Balmer series**, which is in the *visible region*.

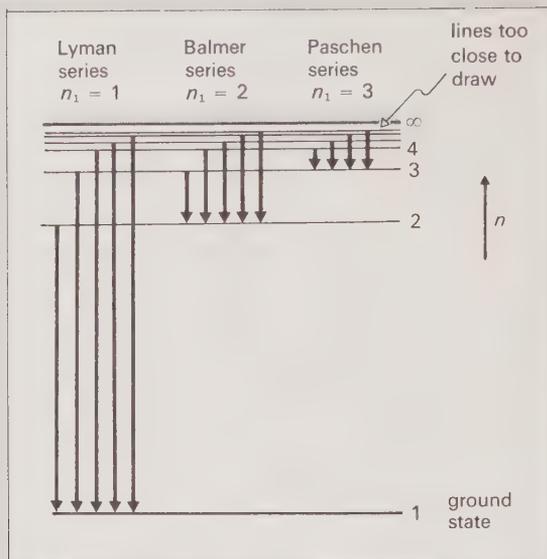


Fig. 63.10. The three main spectral series of atomic hydrogen.

(c) When $n_1 = 3$, $n_2 \geq 4$, the frequencies correspond to the **Paschen series**, which is in the *infra-red*.

These spectral series are illustrated in fig. 63.10 and they provide convincing evidence for the existence of discrete energy levels in atoms.

Classical physics could not explain the general equation above.

Bohr's analysis of the hydrogen atom

Consider an electron of mass m_e and charge $(-e)$ in circular motion of radius r around a proton of much greater mass and charge $(+e)$.

Bohr postulated

(1) That when the electron had an angular momentum given by

$$m_e v r = \frac{nh}{2\pi} \quad (1)$$

then the orbit was stable: although the electron was accelerated, it did not radiate electromagnetic waves.

(2) That when the electron changed from an allowed orbit in which the system had energy E_2 to a different allowed orbit of lower energy E_1 , then it radiated electromagnetic waves of frequency ν such that

$$E_2 - E_1 = h\nu$$

where h is the *Planck* constant.

For the circular orbit, the centripetal force is provided by the *Coulomb* force

$$\left(\frac{1}{4\pi\epsilon_0}\right) \cdot \frac{e^2}{r^2} = \frac{m_e v^2}{r} \quad (2)$$

Eliminating the speed v from (1) and (2)

$$r = \frac{\epsilon_0 n^2 h^2}{\pi m_e e^2} \quad (3)$$

The total energy E of the system is given by

$$E = E_p + E_k$$

where k.e.

$$E_k = \frac{1}{2} m_e v^2 = \left(\frac{1}{4\pi\epsilon_0}\right) \cdot \frac{e^2}{2r} \quad (\text{equ. 2})$$

and p.e.

$$E_p = -\left(\frac{1}{4\pi\epsilon_0}\right) \cdot \frac{e^2}{r}$$

So total energy

$$E = -\left(\frac{1}{4\pi\epsilon_0}\right) \cdot \frac{e^2}{2r} = -\frac{m_e e^4}{8\epsilon_0^2 n^2 h^2}$$

using the value of r from equation (3).

This equation shows

(a) that the system always has *negative* energy —this is because the electron is *bound* to the proton, and

(b) that the energy of an orbit varies as $-1/n^2$, i.e. becomes greater for larger values of n . Since $r \propto n^2$, this means for larger values of r .

The frequency ν radiated by a transition from

$$n = n_2 \text{ to } n = n_1 \quad (n_2 > n_1)$$

is given by

$$h\nu = E_2 - E_1 = \frac{m_e e^4}{8\epsilon_0^2 h^2} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)$$

which can be written

$$\nu = cR_H \left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right) \quad (4)$$

where

$$R_H = \frac{m_e e^4}{8\epsilon_0^2 c h^3}$$

Notes

(a) The values obtained for ν from this expression are in excellent agreement with those found empirically.

(b) Using values of ϵ_0 , m_e , e and h found by other methods, R_H can be found, and agrees with the values found spectroscopically.

(c) By putting $n = 1$ in equation (3), we can calculate the *electron orbital radius* (the *Bohr radius*) in the ground state a_0 to be 0.53×10^{-10} m. Methods based on gas kinetic theory give an atomic diameter $\sim 10^{-10}$ m, which is in agreement.

(d) By putting $n_1 = 1$ and $n_2 = \infty$ in equation (4) we can calculate the energy E required to ionize atomic hydrogen to be about 2.2×10^{-18} J, which agrees with experiment. E is called the **ionization energy**.

Using $E = VQ = V_1e$, we can show that the *ionization potential*

$$V_1 = 13.6 \text{ V}$$

(e) All energy states above the ground state are called *excited states*. The difference between the energy of an excited state and the energy of the ground state is called the **excitation energy**.

The **wave mechanics atom**, which is mentioned on p. 515, has replaced the *Bohr atom* for the following reasons:

(i) The *Bohr theory* cannot account for the spectra of atoms or ions having more than one electron.

(ii) The postulate that the angular momentum takes values $nh/2\pi$ turns out to be an oversimplification.

(iii) High-resolution spectroscopy indicates that each spectral line predicted by the theory consists in fact of many closely-spaced lines.

64 RADIOACTIVITY

64.1	GENERAL DESCRIPTION OF RADIOACTIVITY	518
64.2	METHODS OF DETECTION	519
64.3	MEASUREMENTS WITH A D.C. AMPLIFIER	520
64.4	IDENTIFICATION OF THE RADIATIONS BY SELECTIVE ABSORPTION	521
64.5	THE EFFECT ON THE RADIATIONS OF <i>E</i> AND <i>B</i>	522

64.6	SUMMARY OF THE PROPERTIES OF THE RADIATIONS	523
64.7	RADIATION HAZARDS	523
64.8	RADIOACTIVE DECAY	525
64.9	NUCLEAR REACTIONS	526
64.10	THE LAW OF RADIOACTIVE DECAY	526
64.11	ENERGY CHANGES IN RADIOACTIVE DECAY AND NUCLEAR REACTIONS	528
64.12	SOME USES OF RADIOACTIVITY	529

64.1 GENERAL DESCRIPTION OF RADIOACTIVITY

History

In 1896 *Becquerel* noted that radiation from uranium affected photographic plates.

In 1898 *Curie* noted that thorium and radium showed the same effect.

In 1899 *Rutherford* distinguished

- (1) a low penetration radiation (α -particles),
- (2) a radiation of higher penetration (β -particles).

In 1900 *Villard* discovered a third type of radiation of even greater penetration (γ -rays).

Radioactivity

Radioactivity is the spontaneous disintegration of the nucleus of an atom, from which may be emitted some or all of the following:

- (1) α -particles,
- (2) β -particles,
- (3) γ -rays,
- (4) other particles or types of electromagnetic radiation (less important at this level).

The process represents an attempt by an unstable nucleus to become more stable, and is not affected by

- (a) chemical combination, or
- (b) any change in physical environment other than nuclear bombardment.

We suspect that it is the *nucleus which is concerned* with the effect, but not the orbital electrons.

A source of radiation continuously emits *energy* in large quantities as

- (a) k.e. of ejected particles, and/or
- (b) γ -radiation.

Natural sources of radiation include cosmic rays, rocks (especially granite), luminous paint and television screens. These contribute to *background* radiation which must be taken into account in the measurement of activities.

Radioactive series

Naturally occurring radioactive nuclides can be classified into three groups, in each of which a series of successive disintegrations produces eventually a different lead isotope. For example

- (a) ${}_{92}^{238}\text{U}$ produces ${}_{82}^{206}\text{Pb}$ (uranium series)
- (b) ${}_{90}^{232}\text{Th}$ produces ${}_{82}^{208}\text{Pb}$ (thorium series)
- (c) ${}_{92}^{235}\text{U}$ produces ${}_{82}^{207}\text{Pb}$ (actinium series)

At any instant some of all the intermediate elements produced by disintegration will be present. For laboratory work it is often convenient to choose a radioactive element which can be isolated (e.g.

chemically), and which emits *only one type of radiation*. For example

- (a) $^{241}_{95}\text{Am}$ for α -particles only
- (b) $^{90}_{38}\text{Sr}$ for β^- -particles only
(The β -particles also produce bremsstrahlung (p. 508).)
- (c) $^{60}_{27}\text{Co}$ for γ -rays only.
(β -particles are also emitted but these have low energy and are relatively easily absorbed.)

Activity of sources

The **activity** A of a source is the number of disintegrations it undergoes per unit time. The SI unit of activity is called the **becquerel** (Bq). Thus $1 \text{ Bq} = 1 \text{ s}^{-1}$.

Laboratory sources have activities of about 10^4 Bq, whereas some medical sources might have an activity of about 10^{14} Bq.

64.2 METHODS OF DETECTION

Most detectors indicate the arrival of *energy*, which may produce

- (a) *ionization*, as in
 - (i) the ionization chamber,
 - (ii) the Geiger–Müller (G.M.) tube,
 - (iii) the cloud chamber, and similar chambers,
 - (iv) the spark counter,
- (b) *exposure* of a photographic emulsion,
- (c) *fluorescence* in the phosphor of a scintillation counter, and
- (d) *mobile charge-carriers* in a semiconductor.

Some of these detectors are discussed in this book but details are not included.

(A) Ionization methods

Ionization is the removal of one or more of the extra-nuclear electrons from an atom, thus creating an **ion-pair**, consisting of the *removed electron*, negatively charged, and the *massive remainder of the atom*, positively charged. All radiations can produce ionization of the atoms of solids, liquids and gases.

α -particles have the greatest ionizing power of the three types of radiation, because

- (a) they have twice the charge of a proton which enables them to attract electrons strongly, and
- (b) they travel comparatively slowly and consequently remain a relatively long time in the vicinity of each atom near their path.

Ionization needs energy, and each ion-pair formed results in the ionizing agent losing that amount of kinetic energy. α - and β -particles decrease in speed in the process, and if γ -quanta are not wholly absorbed, their frequency is considerably decreased.

The distances travelled by α - and β -particles of given energy in a given medium are fairly constant, as these particles are slowed down in a large number (10^3 – 10^5) of very small steps, *in each of which* an ion-pair is formed.

X- or γ -rays are absorbed exponentially as a quantum usually loses all its energy (as in pair production) or much of it (as in the *Compton effect*) in a *single event*.

The X- and γ -rays cause ionization by ejecting electrons (secondary radiation).

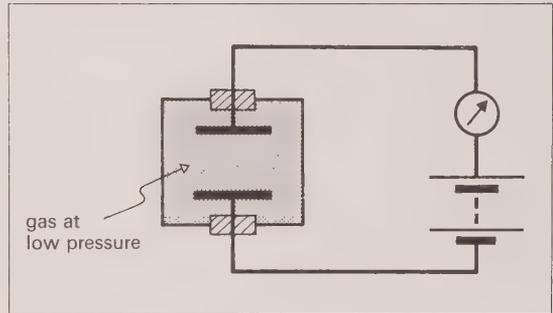


Fig. 64.1. Principle of the ionization chamber.

Ionization chamber

In fig. 64.1, radiation which enters the chamber ionizes gas molecules by collision, and the applied p.d. causes an electrical pulse to pass through the circuit. The pulse can be used to activate

- (a) an amplifier and loudspeaker,
- (b) a d.c. amplifier (p. 364),
- (c) an electronic counting device called a **scaler**, or
- (d) a **ratemeter**, which records the average rate of pulse arrival.

The G.M. tube

This is a very sensitive type of ionization chamber and is shown in fig. 64.2 (overleaf).

γ -rays easily penetrate the walls of the tube but less penetrating radiation enters through the mica window. For detection of α -particles and slow β -particles the window has to be extremely thin.

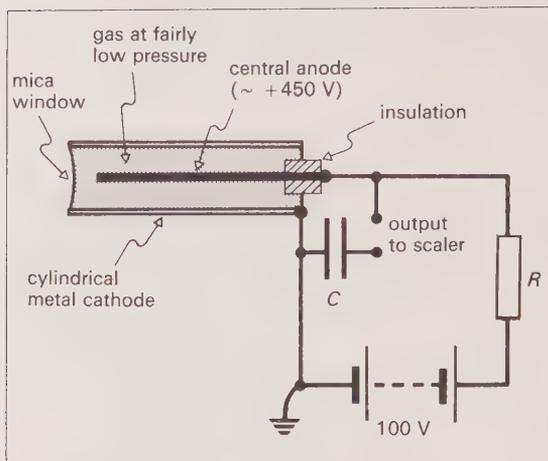


Fig. 64.2. A typical G.M. tube.

The cloud and similar chambers

A charged particle travelling at speed leaves behind a trail of ions. If these occur in a region in which there is vapour about to condense, or liquid about to boil, then the ions will encourage the formation of droplets and vapour bubbles respectively. Either will leave a trail to show the path of the particle, which may be photographed if the chamber is brightly lit from the side. These principles are used in several devices:

(a) *The Wilson cloud chamber*, in which a sudden adiabatic expansion of the right amount cools water vapour so as to produce condensation on ions of either positive or negative charge.

(b) *The diffusion cloud chamber*, in which supersaturated alcohol vapour condenses. The pattern seen in the chamber is characteristic of the nature of the radiation which causes the ionization.

(c) *The bubble chamber*, in which the charged particle leaves a trail of bubbles in liquid hydrogen. This has the advantage that any photograph will not be distorted by expansion processes.

The spark counter

This uses the ionization produced by particles to trigger a spark discharge in a region where air is close to breakdown in a strong electric field.

(B) Photographic emulsion methods

A primary ion colliding with an emulsion gives a line of silver grains on processing. This enables a well-defined permanent record to be made, which can be made three-dimensional by stacking the plates. The emulsion used for detecting radiation

has a high density of silver halide which results in the particle tracks being short.

(C) Scintillation methods

The impact of radiation on a suitable material (which can be matched to the radiation) causes the emission of a minute flash of light—a scintillation. In the *spinharscope* α -particles cause zinc sulphide (a **phosphor**) to scintillate. A single crystal of sodium iodide is used for counting γ -rays.

A **scintillation counter** consists of an appropriate phosphor combined with a photomultiplier tube, which enables the weak flashes produced by β -particles and γ -rays to be detected.

Scintillation detectors have several advantages:

(a) The energies of the particles detected can be measured.

(b) They can deal with very high counting rates with pulse durations as short as 1 ns (10^{-9} s).

(c) Their efficiency in counting γ -rays is nearly 100%.

(D) The solid state detector

If electrons and holes are suddenly created at the junction of a junction diode (p. 476) by ionizing radiation, an applied p.d. will cause a very short pulse to pass through the circuit. Amplification and counting by scaler or ratemeter now enables α -, β -, γ - and X-radiation, as well as protons and neutrons, to be detected.

64.3 MEASUREMENTS WITH A D.C. AMPLIFIER

The d.c. amplifier (46.9) can be used to detect and measure very small ionization currents by using the resistance ranges and connecting a milliammeter at the output (fig. 64.3). The working potential difference for an ionization chamber can be established by using an α -source and the d.c. amplifier to plot a graph of ionization current against applied p.d. V.

(a) The d.c. amplifier can be used to investigate the *range of radiation* by selecting a source which is known to emit only one kind of radiation and placing it at different distances d from the ionization chamber. A graph can then be plotted of the corresponding ionization currents I against the distances d (fig. 64.4), from which the range of that particular radiation in air can be deduced. The experiment can be repeated using layers of different materials.

(b) The d.c. amplifier can also be used for *determination of half-life* (of, for example, thoron gas pumped into the (closed) ionization chamber),

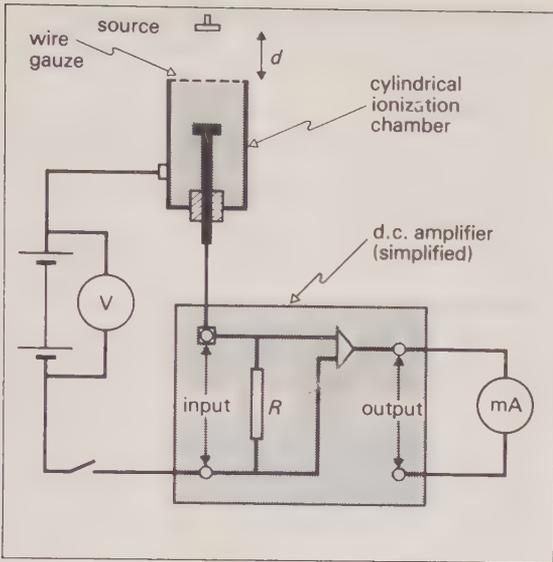


Fig. 64.3. A d.c. amplifier being used to measure the current in an ionization chamber.

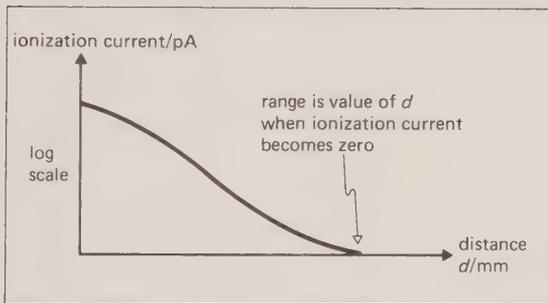


Fig. 64.4. Measurement of range of radiation.

since the ionization currents are directly related to the activities (p. 519). The half-life (p. 527) can be deduced from a graph of $I:t$ or $\log I:t$ (fig. 64.7, p. 527).

64.4. IDENTIFICATION OF THE RADIATIONS BY SELECTIVE ABSORPTION

Effect of different radiations

(a) α -particles penetrate a few tens of mm of air or a thin foil of Al. For a particular disintegrating nuclide all the particles have about the same speed — there is often a line spectrum of nearly equal energies.

They have a small range, because they may produce about 10^4 ion-pairs per mm of path in air at atmospheric pressure.

(b) β -particles are emitted with very variable speeds even for a particular disintegrating nuclide. Low-energy β -particles have a smaller penetration than α -particles, but typically they have about ten times the range. The ionization they produce per unit length of path is about one hundredth that of α -particles.

(c) γ -rays have variable energies which fall into several distinct monoenergetic groups from any particular emitter. The most energetic radiation may penetrate up to several tens of mm of lead, whereas the least energetic can pass through only very thin foils.

These measurements are usually done with a G.M. tube since γ -rays typically produce 10^{-4} times as many ion-pairs per unit length as do α -particles.

When passing through a given material, a monoenergetic (or homogeneous) beam of γ -rays (or X-rays) is absorbed by the processes of the photoelectric effect, the Compton effect and pair production. If I_0 is the incident intensity (rate of energy arrival per unit area) and I is the transmitted intensity after passing through a thickness x ,

$$I = I_0 e^{-\mu x}$$

where μ is the linear absorption coefficient. μ depends upon both the absorber (high values for high Z and high-density materials) and the γ -ray beam (high values for low photon energies). The half-value thickness reduces the intensity of a narrow monoenergetic beam to half its original value ($I = I_0/2$) and is given by

$$x_{1/2} = \frac{\ln 2}{\mu} = \frac{0.693}{\mu}$$

A graph of $\ln I:x$ is a straight line of gradient $-\mu$, from which $x_{1/2}$ can be determined. $x_{1/2}$ is often used in describing the quality (penetrating power) of the radiation.

Air acts as an almost perfectly transparent medium to γ -rays, and the intensity of γ -rays emanating from a point source varies inversely as the square of the distance from the source. The inverse square law does not hold for α - and β -particles in air because it absorbs them. The law would hold for beams of these particles when travelling through a vacuum, provided they emanated from a point source.

Effect of different absorbing materials

Ability to absorb depends on

- the material density, and
- the material thickness.

The atomic number of the material has little effect on the absorption of α - and β -particles, but for γ -rays in high and low energy ranges there is very much higher absorption in materials of high atomic number.

It is convenient to quote data in terms of *mass per unit area*.

Suppose an α -particle could penetrate 4×10^{-2} m of air of density 1.3 kg m^{-3} . Then the mass per unit area to stop these particles is $5.2 \times 10^{-2} \text{ kg m}^{-2}$. If we used aluminium of density 2700 kg m^{-3} , we would require a thickness of 2×10^{-5} m.

Distinguishing two simultaneous radiations

We can plot a graph of ionization currents against different thicknesses of absorbing material. A graph like fig. 64.5 which has an abrupt change of slope shows the range of the less penetrating radiation.

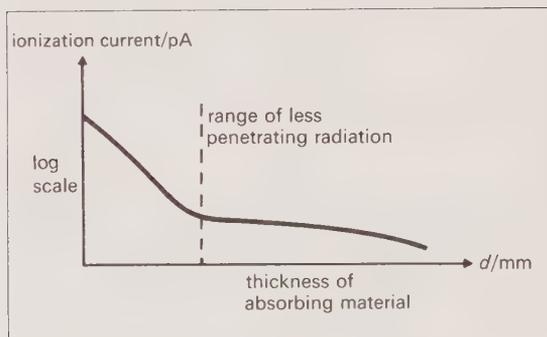


Fig. 64.5. The form of an absorption curve for two simultaneous radiations.

64.5 THE EFFECT ON THE RADIATIONS OF E AND B

(A) Electric field E

- (a) α -particles are deflected as positive charge would be deflected,
- (b) β -particles are deflected as negative charge would be deflected,
- (c) γ -rays are undeflected and have no charge.

(B) Magnetic field B

- (a) α -particles show a small deflection (fig. 64.6), which suggests they may be *particles* of relatively large mass,
- (b) β -particles show a large deflection—they are *particles* of small mass. They also show considerable dispersion, which demonstrates that they are emitted with variable speeds.
- (c) γ -rays are undeflected.

(C) Crossed E and B

This enables the *specific charge* and *speed* of the particles to be measured (p. 470).



Fig. 64.6. Schematic deflection of radioactive radiations.

64.6 SUMMARY OF THE PROPERTIES OF THE RADIATIONS

(A) α -particles

Their specific charge is half that of the proton. When they acquire electrons which neutralize them, they show the spectral lines typical of helium, so we deduce that they are *helium nuclei* (Rutherford and Royds—1909). Each α -particle consists of four nucleons—two protons and two neutrons. This is a particularly stable arrangement as each energy level in the shell model of the nucleus can accommodate $2p$ and $2n$.

α -radiation is the result of the ejection at high speed of α -particles from a heavy nucleus. Many α -emitters are naturally occurring nuclides whose atomic number is between that of lead (82) and uranium (92). One very important emitter, plutonium-239, is manufactured in nuclear reactors.

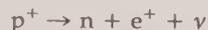
(B) β -particles

(a) *Negative β -decay.* In this decay the nucleus attains greater stability by the conversion of a neutron to a proton. The energy released is the k.e. of an ejected *electron* plus the energy of the *antineutrino*—a particle with zero rest mass and zero charge. Since the rest mass is zero it has an energy pc only by virtue of travelling at speed c .



β^- -emission is characteristic of nuclei having a large proportion of neutrons. (The ejected electrons come from the *nucleus* rather than from the electron cloud, and yet do not exist as electrons while in the nucleus.)

(b) *Positive β -decay.* Positron decay is possible if the mass of the parent atom exceeds that of the daughter atom by at least $2m_e$ (or its energy equivalent). Any extra energy appears as the k.e. of the daughter nucleus, the positron and the *neutrino*—a particle with its sense of spin opposite from that of the antineutrino.



β^+ -emission is characteristic of nuclides with low atomic number which have a high proportion of protons.

If a nucleus has insufficient energy for positron decay, it may attain the same result by **electron capture** from one of the innermost shells. The energy released is that of a neutrino and there will also be X-radiation owing to the vacancy in the electron shell. A particular species of nuclide often decays by either positive β -decay or electron capture.

Quite a few β -emitters are found in nature (e.g. potassium-40), but most are prepared in the laboratory by irradiating matter with neutrons.

All the decays mentioned above involve a change in Z and N with A remaining constant. If a graph of *mass defect* of nuclides is plotted against Z for an isobar of odd number A , the result is a parabola. The most stable position is at the lowest point of the curve. The β -decay of other nuclei will help them to reach this point. For an isobar of even number A , two parabolas are obtained.

(C) γ -rays

These (being electromagnetic waves) have properties similar to X-rays (p. 507), but originate in

Table of properties of the radiations

Property	α	β^-	γ
fluorescence on a phosphor	yes	less	very little but is observed with NaI
effect on photographic plate	yes	yes	yes
number of ion-pairs per mm in air	$\sim 10^4$	$\sim 10^2$	~ 1
absorption: stopped by up to	$\sim 10^{-2}$ mm Al \sim tens of mm air	\sim mm Al \sim m air	~ 100 mm Pb
deflection by E and B	as +ve charge	as -ve charge	nil
speeds	$\sim 10^7$ m s $^{-1}$	$\sim 10^8$ m s $^{-1}$ but variable	3×10^8 m s $^{-1}$

the nucleus and generally have a smaller wavelength.

When a nucleus is excited it can return to its ground state by emitting a γ -ray. It is often found that α - and β -emitters also emit γ -rays because nuclei are frequently produced in excited states as a result of their decay. However, many radioactive substances do not emit γ -rays as the resultant nuclei are produced in the ground state. The wavelength, λ , is measured by crystal diffraction methods.

Using

- (a) $c = v\lambda$, and
(b) $E = hv$

their energies E may be measured.

► Antiparticles

The positron is said to be the antiparticle of the neutron, and the antineutrino is the antiparticle of the neutrino. All elementary particles have such counterparts (e.g. proton–antiproton and neutron–antineutron pairs) but these are not often detected as a large amount of energy is required to generate such pairs. The annihilation process is often the most effective method of detecting antiparticles, and occurs when an antiparticle meets its corresponding particle.

64.7 RADIATION HAZARDS

Biological effects

The ionization produced by high-energy radiation can alter and upset the function of complex cell molecules such as DNA, and can produce constituents that may have adverse effects on metabolic rates. The cells of reproductive organs, thyroid gland, intestines and blood-forming tissues such as bone marrow are particularly sensitive. The long-term effects of cell damage include anaemia, leukemia, cancer, sterility and genetic mutation. More immediate effects include blistering of the skin, radiation sickness, loss of hair, emaciation and even death.

Radiation exposure and dose

The exposure to radiation, X ($C\ kg^{-1}$), is defined as follows:

$$X = \frac{Q}{m}$$

where Q is the total charge of the ions of one sign produced in air when all the electrons and positrons, liberated by X- or γ -ray photons in a volume of air of mass m , are completely stopped in air. The absorbed dose, D ($J\ kg^{-1}$), is restricted neither to X- and γ -radiation nor to air, and is defined for irradiated material as follows:

$$D = \frac{E}{m}$$

where E is the energy absorbed by material of mass m . D is measured in grays (Gy) where $1\ Gy = 1\ J\ kg^{-1}$. As direct measurement of D is difficult, it is usually calculated from X using the relationship

$$D\ (J\ kg^{-1}) = f\ (J\ C^{-1}) \times X\ (C\ kg^{-1})$$

where f depends upon the biological medium and the photon energy. dX/dt ($C\ kg^{-1}\ s^{-1}$) is the exposure rate and dD/dt ($Gy\ s^{-1}$) is the absorbed dose rate. Equal doses of different ionizing radiations provide the same amount of energy in a given absorber, but may produce different biological effects. The effective dose or dose equivalent H (Sv) is defined by

$$\left(\begin{array}{c} \text{dose} \\ \text{equivalent} \end{array} \right) = \left(\begin{array}{c} \text{total} \\ \text{absorbed} \\ \text{dose} \end{array} \right) \times \left(\begin{array}{c} \text{relative biological} \\ \text{effectiveness (r.b.e.)} \end{array} \right)$$

where the r.b.e., or quality factor Q , depends upon the nature of the radiation. H is measured in sievert (Sv) to distinguish it from D .

$$H\ (Sv) = D\ (Gy) \times Q$$

$Q = 1-2$ for X-, γ - and β -radiation; and $Q = 10$ for fast neutrons, protons and α -particles. For background radiation, the dose rate $dH/dt = 1\ mSv$ per year, whereas for localized therapy $H \sim 50000\ mSv$. The ICRP (*International Commission on Radiological Protection*) has laid down safety standards in terms of maximum permitted dose levels (MPLs). For example, the MPL for someone engaged on radiation work is $50\ mSv\ year^{-1}$ (whole body or reproductive organs) but can be as high as $750\ mSv\ year^{-1}$ (hands, forearms, feet).

Hazards and protection

Sources of background radiation include cosmic rays and naturally occurring radioisotopes, both in the body, e.g. ^{40}K , and in rocks and buildings, e.g. ^{238}U , ^{226}Ra . Nuclear testing and the increasing use of radiation in diagnostic and therapeutic medicine represent significant hazards to the general population. Astronauts are vulnerable

when they pass through the Earth's *Van Allen* belts, and would be in real danger of a high radiation dose if a solar flare occurred during a space flight. Radiation can be a hazard to a human being when the body is exposed to the energy from a source which may be either internal or external.

(a) A source should never be present *inside the body*, since at all times its radiation will damage the living cells. Therefore protective clothing should be worn to prevent the contamination of the body by radioactive substances. For example, strontium-90 is absorbed into the bones where its radiation damages the bone marrow.

(b) An *external source* is less dangerous, because it will cause damage only while it is close to the body. Even while it is there

(i) the α -radiation cannot penetrate the skin, but care must still be taken because some α -emitters create radioactive gaseous daughter products which can be absorbed into the body;

(ii) the other radiations can be screened: e.g. β -radiation by a sheet of Perspex, and γ -radiation by tens of mm of lead or by a few metres of concrete.

As a general rule materials will not become radioactive on exposure to radiation from other nuclei, but they may do so when used as a target in a particle accelerator.

Radiation workers are protected by

(a) lead shielding of radioactive materials in store,

(b) handling samples by remote control or at a safe distance,

(c) shielding from equipment producing radiation (and scattered radiation),

(d) wearing *film badges*.

The **film badge dosimeter** provides a cheap, permanent record of radiation exposure by using the degree of blackening of a photographic film or emulsion. The badge is checked regularly and provides the most common method of personnel *monitoring*. If the MPL is exceeded, a worker must be removed from radiation work for a stipulated period of time.

In the **school laboratory**, naturally occurring radioactive substances and prepared sealed sources must be handled with tongs rather than fingers. They should be held at a distance and pointing away from the body. The sources must never be tampered with and should be stored in their lead containers when not in use. Radioactive liquids should be handled above a spill tray, lined with absorbent paper. Protective clothing should

be worn and hands should be washed before and after each experiment. Radioactive material should *never* be allowed to enter the body.

Waste disposal

Nuclear power stations discharge waste gaseous radioactive materials into the atmosphere and liquids into the sea at recognized safe levels of activity. In the disposal of the bulk of radioactive waste, which is mainly liquid, the following aspects must be considered:

(a) the storage of *chemically active* materials,

(b) the containment of *radioactive* materials,

(c) the heating effect caused by the energy released.

A further factor is the length of the half-lives of radioactive decay products, e.g. $T_{1/2}$ for ^{137}Cs is 27 years so that it must be stored securely for over 500 years before the activity reaches the recommended level of $A_0 \times 10^{-6}$. The *transuranic* elements require millions of years.

Some early methods of storage, such as porous soil beds and underground salt beds, have been abandoned because of subsequent contamination problems. Low-level waste can be stored safely in steel drums but high-level waste is potentially more dangerous. Present research is investigating the possibility of incorporating the latter in suitable, stable solids such as borosilicate glass and concrete. These materials must be capable of withstanding

(a) contact with water,

(b) high temperatures,

(c) internal radiative bombardment.

Ideally these solids should then be buried deep in the Earth—perhaps where two tectonic plates are colliding so that the radioactive waste can be subducted into the mantle. However, there are many difficulties to be overcome, and the whole problem becomes increasingly urgent because the volume of waste is increasing as is the concern regarding security.

64.8 RADIOACTIVE DECAY

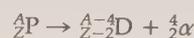
The Russell, Fajans and Soddy rules (1913)

(a) In α -decay, an α -particle is emitted, so

(i) A decreases by 4, and

(ii) Z decreases by 2.

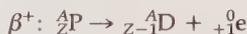
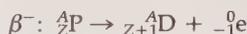
If P represents the parent nuclide and D the daughter, then



(b) In β -decay

(i) A does not change since no nucleon is emitted, and

(ii) Z increases by 1 (β^- -decay) or decreases by 1 (β^+ -decay).



(c) A γ -ray represents the emission of energy from a nucleus which is returning to its ground state (p. 516). Both A and Z remain unaffected.

(excited nucleus) \rightarrow (more stable nucleus) + γ

When the emission of a γ -ray accompanies the emission of an α -particle or a β -particle, rule (a) or (b) will apply.

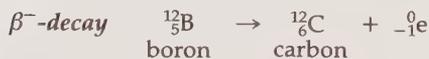
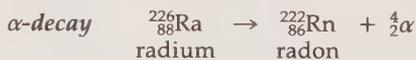
Examples of radioactive decay

We represent the changes by equations that show what happens to *the nucleus*. They must be distinguished from equations that represent chemical reactions (which are concerned with the bonding of an *atom's outermost electrons*). In all such equations, the sums of

(a) the mass numbers (representing the *number of nucleons*), and

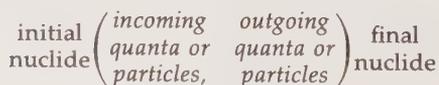
(b) the atomic numbers (representing *electric charge*)

must be the same on each side of the equation if it is correct. (The number of nucleons and charge are both *conserved*.)

**64.9 NUCLEAR REACTIONS**

A nuclear reaction is an induced change involving a spontaneous decay of a nucleus from a highly excited state. It requires the bombardment of a nucleus by high-energy particles (such as those from a particle accelerator) or by γ -radiation which results in their capture by, and the subsequent decay of, a compound nucleus.

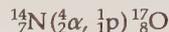
The symbolic way of expressing a nuclear reaction is:



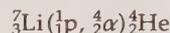
Reactions are normally classified in terms of the quantities enclosed by the brackets. The equations which follow are examples of what can take place.

(A) A (d, p) reaction

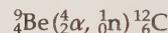
A stable nucleus absorbs a deuterium nucleus which leads to the ejection of a proton and the formation of a new radioactive nucleus.

(B) An (α , p) reaction

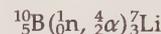
This reaction was carried out by *Rutherford* (1919), and was the first artificial transmutation. The presence of the oxygen isotope was shown spectroscopically.

(C) A (p, α) reaction

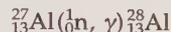
This reaction was carried out by *Cockcroft and Walton* (1932), and was the first transmutation achieved using artificially accelerated ions (the protons).

(D) An (α , n) reaction

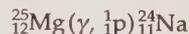
This was carried out by *Chadwick* (1932) and led to the discovery of the **neutron**. This is an example of a *neutron source*.

(E) An (n, α) reaction

This is one method of detecting neutrons.

(F) An (n, γ) reaction

The unstable isotope of the target nucleus is formed by neutron capture.

(G) A (γ , p) reaction

This reaction is known as *photodisintegration*.

64.10 THE LAW OF RADIOACTIVE DECAY

Individual nuclei disintegrate independently and so the decay is a *random process*. This is well illustrated by the spark counter. The number of

nuclei present in a small sample is so great that we can treat the process statistically, and determine the rate at which a parent element disintegrates into its daughter element.

Nevertheless, strictly only the probability of disintegration in a particular time interval can be stated.

The radioactive decay constant λ

Suppose at any instant there are N active nuclei present in a sample. We assume

$$\left(\begin{array}{c} \text{rate of} \\ \text{disintegration} \\ \text{or activity} \end{array} \right) \propto \left(\begin{array}{c} \text{number of} \\ \text{nuclei available} \\ \text{to disintegrate} \end{array} \right)$$

$$A \propto N \quad \text{or} \quad -\frac{dN}{dt} \propto N$$

We define the radioactive decay constant λ by the equation

$$A = -\frac{dN}{dt} = \lambda N$$

$$A[\text{Bq}] = -\frac{dN}{dt} \left[\frac{1}{\text{s}} \right] = \lambda \left[\frac{1}{\text{s}} \right] \cdot N$$

Suppose at $t = 0$ we have N_0 active nuclei, and after a time t we have N active nuclei. Then

$$\int_{N_0}^N \frac{dN}{N} = -\lambda \int_0^t dt$$

$$\ln \left(\frac{N}{N_0} \right) = -\lambda t$$

$$N = N_0 e^{-\lambda t}$$

The number of parent nuclei present decays exponentially.

The half-life period $T_{1/2}$

Suppose $T_{1/2}$ is the time taken for the number of active nuclei to be reduced to $N_0/2$. Then

$$\frac{N}{N_0} = e^{-\lambda T_{1/2}} = \frac{1}{2}$$

$$e^{-\lambda T_{1/2}} = \frac{1}{2}$$

$$\therefore T_{1/2} = \frac{\ln 2}{\lambda} = \frac{0.693}{\lambda}$$

The value of $T_{1/2}$ depends on the emitting nuclide, being 10^9 years for uranium-238 and 10^{-6} s for some products of radium decay.

Fig. 64.7 illustrates the concepts of $T_{1/2}$ and exponential decay. From $N = N_0 e^{-\lambda t}$, it is clear that

$$\ln N = -\lambda t + \ln N_0$$

The second graph is a semilogarithmic plot and the decay constant can be determined from the gradient ($=-\lambda$). The intercept is $\ln N_0$. Similar graphs are obtained if N is replaced by a quantity such as activity A or ionization current I .

The decay of radioactive parent atoms and the growth in the number of daughter atoms (assumed stable) are illustrated graphically in fig. 64.8 (overleaf) as functions of time. Assuming that the daughter atoms are stable, the number of daughter atoms present at any time equals the number of parent atoms that have decayed, namely $N_0 - N$.

From above,

$$N_0 - N = N_0 - N_0 e^{-\lambda t}$$

$$= N_0 (1 - e^{-\lambda t})$$

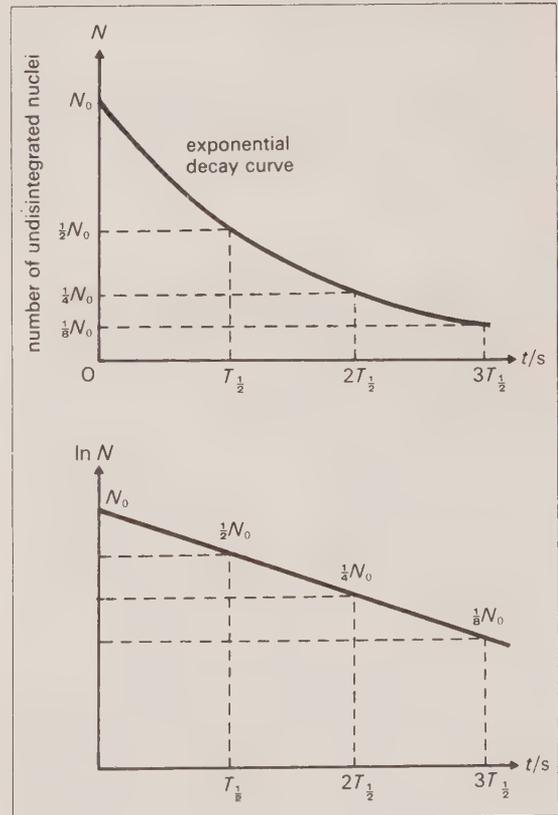


Fig. 64.7. Exponential decay and half-life.

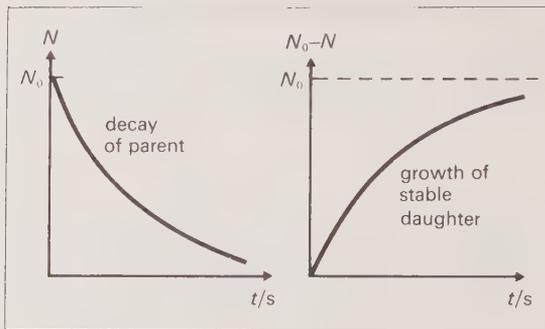


Fig. 64.8. Decay of parent atoms and growth in number of daughter atoms.

Radioactive equilibrium

If a sample of the first member of a radioactive series has been left for some time, all members of the series will be present together. The *relative* amounts of the various nuclides will be constant at radioactive equilibrium, which occurs when each nuclide is decaying at the rate at which it is formed.

$$\text{Therefore} \quad A_1 = A_2 = A_3$$

$$\text{which means that} \quad \lambda_1 N_1 = \lambda_2 N_2 = \lambda_3 N_3$$

$$\text{and so} \quad \frac{N_1}{(T_{\frac{1}{2}})_1} = \frac{N_2}{(T_{\frac{1}{2}})_2} = \frac{N_3}{(T_{\frac{1}{2}})_3}$$

64.11 ENERGY CHANGES IN RADIOACTIVE DECAY AND NUCLEAR REACTIONS

The *disintegration* or *reaction energy* (Q value) is the energy released in a radioactive decay or a nuclear reaction per disintegrating nucleus.

$$Q = \left[\left(\begin{array}{c} \text{total rest mass} \\ \text{going into} \\ \text{reaction} \end{array} \right) - \left(\begin{array}{c} \text{total rest mass} \\ \text{coming out of} \\ \text{reaction} \end{array} \right) \right] c^2$$

If $Q > 0$, nuclear energy is released—an **exothermic** reaction.

If $Q < 0$, nuclear energy is absorbed—an **endothermic** reaction.

(a) In α -decay the α -particles have discrete energies which lie in the range 0.6 to 1.6 pJ.

(b) In β -decays the negatrons or positrons have a continuous energy spectrum up to the calculated maximum value, since the remainder of the energy is carried away by the antineutrino or the neutrino which is also emitted. Examples:

(i) β^- -decay of $^{12}_5\text{B}$ to $^{12}_6\text{C}$ has $Q = +2.1$ pJ per disintegration.

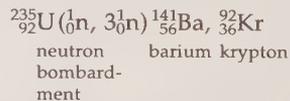
(ii) β^+ -decay of $^{12}_7\text{N}$ to $^{12}_6\text{C}$ has $Q = +2.6$ pJ per disintegration.

(c) In γ -decay the γ -photon energies cover the range 0 to about 1 pJ.

High-energy nuclear reactions involve bombarding particles of energies ≈ 10 nJ; *low-energy* reactions involve particle energies up to ≈ 3 pJ. For example, in the *Cockcroft and Walton* (p, α) reaction, $Q = +2.8$ pJ, whereas in the *Rutherford* (α, p) reaction, $Q = -0.19$ pJ.

Nuclear fission

Nuclear fission, which is the disintegration of a massive nucleus into two fragments with comparable masses, is a special type of low-energy nuclear reaction. Sufficient excitation energy for the nucleus to split may be provided spontaneously (a relatively rare event) or, more usually, by neutron bombardment and capture. The *reaction cross-section* of a nucleus measures the probability *in space* of a nuclear reaction. (The decay constant λ measures the probability *in time* of the decay of an unstable nucleus.) The value may be greater or smaller than the *geometrical* cross-section and depends upon the particular nucleus and bombarding particle, as well as the latter's energy. Oscillations of the excited nucleus cause a deformation of the spherical shape which allows the *Coulomb* repulsion forces to overcome the supposed nuclear surface tension (short-range nuclear interaction). This means that there is a greater probability that the nucleus will deform so much that it divides into two, rather than return to its spherical shape. The two nuclei produced will lie in the middle of the curve of fig. 63.6 so that the final value of E_b/A is greater than its original value. A nucleus undergoing fission does not split into the same product nuclei each time. For uranium-235 the product masses have values of A between 70 and 165. An example is given below:



The products have too many neutrons to be stable and so these are either released simultaneously or they change into protons via β^- and γ -decays. The total energy released per disintegration is very high compared with a typical exothermic nuclear reaction and can be calculated from either

- the loss of mass, or
- the change in the value of E_b/A .

A typical energy release is 32 pJ per disintegration, of which about 85 per cent is the k.e. of the fission fragments. The rest of the energy is shared by antineutrinos, γ -rays, β -particles and neutrons. The fission process provides a great source of energy as it can result in a *self-sustaining chain reaction*. The requirement for this is that for every uranium atom split there must be at least one neutron, out of the average of 2.5 produced, that will split another uranium atom. A chain reaction occurs when the *critical mass* of fuel has been assembled—this will depend upon the fuel being used as well as the shape of the assembly. An *enriched fuel* has an artificially increased proportion of material able to undergo fission, e.g. 3% of ^{235}U rather than the naturally occurring 0.7%. This proportion decreases with time of use and so the fuel needs taking out to allow for the removal of waste material. The flux inside a **nuclear reactor** may be $\approx 10^{19}$ neutrons $\text{m}^{-2} \text{s}^{-1}$. Loss of neutrons is minimized by

- (a) having a large nuclear reactor,
- (b) slowing down the neutrons using a **moderator** (such as beryllium or graphite)—the aim being to increase the fission *cross-section* but to avoid neutron capture, and
- (c) arranging the uranium fuel rods appropriately.

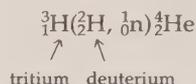
The rate at which fission reactions occur is regulated by control rods which can easily absorb neutrons. When each fission produces *less than one* further fission the reactor is **subcritical**. When it produces *at least one* further fission the reactor becomes **critical** and the reaction is self-sustaining. In a *nuclear power station*, a nuclear reactor provides the heat required to produce steam which is then used, as in a conventional power station, to drive a turbo-generator. In *thermal reactors*, thermal neutrons have been moderated to give them energies which match the average k.e. of the surrounding atoms. The fission energy released is extracted by means of a circulating fluid called a *coolant*. *Gas-cooled* reactors use coolant gas at high temperatures and pressures. *Pressurized-water* reactors use high-pressure water as both moderator and coolant. In a *fast-breeder* reactor, no moderation is required as fission occurs by fast neutrons. The fuel in a *breeder* reactor consists of

- (a) a *fissionable* material, e.g. plutonium-239, and
- (b) a *fertile* material, e.g. uranium-238.

The fertile material can be converted in the reactor into fissionable material. Liquid sodium metal is used as the coolant.

Nuclear fusion

Light elements can fuse together to increase their binding energy per nucleon. This mass reduction results in the release of large amounts of energy (of the order of pJ per fusion). A maintained temperature of 10^8 to 10^9 K is needed to give the nuclei enough energy to overcome the initial electrostatic repulsion and to come together (hence '**thermonuclear**' reaction). Note that there is no such repulsion in nuclear *fission* because the neutron has no electric charge and can therefore approach a nucleus even if its k.e. is very small. If the conditions are appropriate, the energy liberated in fusion is sufficient to excite other nuclei so that a chain reaction results. This chain reaction becomes a nuclear explosion. The radiation of solar and stellar energy is caused by cycles of thermonuclear reactions, in which the overall effect is the formation of helium from hydrogen. For example,



64.12 SOME USES OF RADIOACTIVITY

(A) Radioisotopes

A radioactive isotope is called a **radioisotope**. Radioisotopes are usually made within a nuclear reactor where a large flux of neutrons is available.

Uses

(a) **Radiotherapy**. Cobalt-60 emits γ -rays which are used *therapeutically* for treating certain cancers (e.g. of the pelvis, cervix, larynx and pituitary gland) by a single dose or by a course lasting several days or weeks. In general, healthy tissue is more radiation-resistant than cancer cells but great care must be taken over absorbed doses. γ -rays can also be used to sterilize medical instruments and dressings.

(b) **As tracer elements**. A radioisotope will experience the same chemical treatment as its non-active isotope, but its activity enables it to be detected even in minute quantities. Tracers are used in biological research (e.g. investigating the transport of carbohydrates in plants) and agriculture (e.g. studying the effects of different fertilizers). They also have many *diagnostic* medical applications, e.g. in circulation studies, haemorrhage location, radiocardiography, assessment of thyroid function, localization of tumours. Industrial uses include the monitoring of fluid flow in, for example, polluting effluent in rivers and in leaks from underground pipes. Geologists and

geographers can study the flow of underground water and the movement of sediments in estuaries.

(c) *Miscellaneous*. The thickness of a material, such as paper and paint films, can be measured from the decrease in the intensity of the radiation that has passed through it. This can be extended to checking the level of filling of various containers. The storage life of foods can be increased by irradiation. The small size of a radioisotope source makes it more suitable than an X-ray machine for radiography of certain metal objects such as welded pipelines, ships' hulls and legs of drilling platforms. Radioisotope-powered generators for space satellites and for terrestrial use in remote places are increasingly in demand, for they last a long time and need no maintenance.

(B) Atomic energy

(a) *Nuclear fission* has been described in 64.11 as the basis for the atomic reactor within which the k.e. of the fission fragments is a source of thermal energy. When each fission reaction produces *well above the one* further fission, the reactor is said to be **supercritical**. An example of this is the *atomic bomb*.

(b) *Nuclear fusion* as described in 64.11 is the basis of the *hydrogen bomb*, which needs a fission bomb to provide the required starting temperature. A controlled thermonuclear reaction would provide an enormous source of energy (using naturally occurring deuterium for example) but, as yet, research work has been unsuccessful. The difficulties include maintaining the extraordinarily high temperatures required, and containing the very hot nuclei and electron mixture, called a

plasma. A great deal of work is being done on the design of containers (*magnetic bottles*) which use magnetic fields to prevent the plasma from touching the walls. The pressure of the unheated plasma has to be very low because of the very large pressure increase when it is heated to its working temperature. Extraction of the energy released presents a further problem.

Laser fusion research is investigating the effects of firing short, powerful ($\sim 10^{14}$ W) laser pulses at small pellets of fusion material such as deuterium or tritium.

(C) Dating methods

Geological

Uranium has a convenient half-life ($\sim 4.5 \times 10^9$ years). The age of rocks can be measured from observations on the relative amounts of ^{232}Th , ^{238}U , ^{208}Pb (from Th), ^{206}Pb (from ^{238}U) and ^{207}Pb (from ^{235}U) in the rocks. The ages of the Earth, meteorites, and rock from the Moon have been estimated by radiometric dating. Other methods, appropriate for younger rocks, use the ratios $^{40}\text{K}/^{40}\text{Ar}$ and $^{87}\text{Rb}/^{87}\text{Sr}$.

Archaeological

Radioactive ^{14}C has a convenient half-life ($\sim 5.7 \times 10^3$ years), and also has a constant ratio to ^{12}C in *living matter*. Measurement of their proportions in dead organisms enables an estimate to be made of the time that has elapsed since death. This method of dating is important for plants and animals, using well-preserved parts such as wood and bones.

XII ENERGY RESOURCES

65 ENERGY RESOURCES

65.1 CONSERVATION AND CONVERSION OF ENERGY	532
65.2 SOURCES OF ENERGY	532
65.3 CONSUMPTION OF ENERGY	533

65.4 POLLUTION	534
65.5 ALTERNATIVE SOURCES OF ENERGY	534

65.1 CONSERVATION AND CONVERSION OF ENERGY

The importance of the concept of energy has been emphasized throughout this book and a table of comparative energy values was given in section 1.7. In section 4.2, the Law of Conservation of Energy stated that

There is no change in the total energy of the Universe.

Various recognizable forms of energy are encountered in physics, namely kinetic, potential, thermal and internal, radiant, chemical, nuclear and mass energy. Forms such as electrical, magnetic and sound, not mentioned explicitly in this list, can be treated in terms of kinetic and potential energies. The different forms yield a large number of theoretically possible transfers from one form of energy to another. In practice, many are difficult to imagine (e.g. sound to nuclear), and only a few are useful (e.g. chemical to mechanical—in a rocket for example). An important feature of energy conversion is that transfers can take place *from any energy form into thermal and internal energy*. Useful forms of energy, such as electrical and chemical, are called *high-grade* whereas forms such as internal, which cannot easily be converted into anything else, are called *low-grade*. The change from high-grade to low-grade is called *degradation* of energy (26.15) and is accompanied by an increase in entropy (26.11). Although total energy is conserved, fuels are not, and so as energy becomes increasingly disordered, new sources of high-grade energy are required.

An *energy converter* is a device which takes in one form of energy and changes it into one or

more other forms. The efficiency of a conversion process is a measure of the proportion of the energy input that is produced for the required job. It is expressed as a fraction or a percentage and is defined by the following equation:

$$\text{efficiency} = \frac{\text{useful energy output}}{\text{total energy input}}$$

If, for example, a dynamo is required to light a lamp, the efficiency would be given by

$$\frac{\text{light energy from lamp}}{\text{k.e. of dynamo}}$$

The unwanted heating of the circuit and the surroundings reduces the efficiency of the energy-converting dynamo. Devices which involve chemical change and transfer of thermal energy (e.g. burning fuel in boilers) and those which involve mechanical change and electromagnetic processes (e.g. a hydroelectric turbine), can attain quite high efficiencies in the range 60–90%. However, devices called **heat engines** (26.12), which involve chemical change and transfer of thermal energy to produce mechanical work, *cannot* be as efficient, e.g. a diesel engine may be 40% efficient and a steam locomotive only 10%. As described in 26.12, the maximum efficiency of a heat engine is determined by the source and sink temperatures. In 26.14 it was emphasized that a *real* heat engine will be even less efficient due to frictional forces, turbulence and loss of thermal energy.

65.2 SOURCES OF ENERGY

Primary fuels, such as crude oil and coal, are used in the form in which they are obtained originally, and the total energy that they can provide is called

primary energy. A primary fuel may undergo a number of conversions, perhaps by refineries and power stations, until it reaches its *final-use* stage, e.g. electricity or petrol, when it provides *final-use* energy. The energy available to the final user is called *functional* energy, and the ratio of this to the final-use energy is determined by the final energy converter, e.g. car or electric motor. *Exhaustible* (or *finite* or *non-regenerative*) sources include the **fossil** fuels *coal, peat, oil and natural gas*, and the **nuclear** fission fuels (64.11). *Inexhaustible* (or *renewable* or *regenerative*) sources include those of **solar** origin such as *wind, water, biofuels*, as well as **tidal** and **geothermal** sources.

The *energy density* of a fuel such as coal is relatively high so that the converter required can be correspondingly small, e.g. a furnace. For solar energy, the energy density is much lower so that a solar converter needs to be larger to produce the same power. An energy source should be available *when* and *where* it is required and should be capable of meeting variable demand. *Storage* and *transportability* are therefore important factors. Some energy conversions are more *feasible* than others, e.g. coal for heating requires simpler apparatus than is necessary for solar or wave energy. The storage of electrical energy is a major problem which has yet to be resolved satisfactorily, for it is often produced when not required. The energy density in E - and B -fields is proportional to E^2 and B^2 respectively. High-intensity fields can therefore store a lot of energy but difficulties include the short duration of storage (e.g. capacitors) and the need for (expensive) superconducting magnets for strong B -fields. The problem is solved partially by converting the electrical energy into other forms of energy which can be stored more effectively. Some possibilities include

- (a) *potential energy*—of water in a pumped storage hydroelectric scheme, or of air when compressed,
- (b) *chemical energy*—in batteries (which tend to be heavy and expensive), and fuel cells using hydrogen and oxygen,
- (c) *thermal energy*—in solids and fluids, as latent heat, and in reversible chemical reactions.

Transport costs are an important factor for all forms of energy. The practical maximum distance of transmission of each form determines the ideal size and output of the power plant, e.g. electricity—100 km, 1000 MW; oil—10 Mm, 2 TW (a typical oilfield). Superconducting, and therefore energy-saving, cables for the transmission of electrical energy are a possibility in the relatively near future.

65.3 CONSUMPTION OF ENERGY

At present the annual world energy consumption is $\sim 10^{20}$ – 10^{21} J, and this relies heavily on the *fossil* fuels. There is a large expansion programme planned for *nuclear* power over the next few decades but objections to this are increasing the urgency for research into *alternative* sources and their exploitation. The approximate contributions to the world's consumption are oil 40%, coal 30%, gas 20%, hydroelectricity and nuclear power 10%. Dependence on fossil fuels varies between different areas of the world and a critical factor is therefore the availability of resources to these areas. Some countries have to rely heavily on imports of certain fuels. *North America* and *Western Europe* consume about 50% of the world's annual fuel production, and the inhabitants of some areas of the world effectively consume 30 times as much primary fuel as some others. Since 1925 the world demand for fuel has often increased *exponentially*, although the doubling times have differed over various time periods. However, the growth rate levelled off during the oil crisis of 1973, and the demand has been decreasing since 1979 (although it remains extremely high). It is difficult to predict the pattern of future demand for this will depend upon factors such as the world population growth, human expectations and the development of industry. A lot will also depend upon the extent of the world's stock of fossil and nuclear fuels, and the effectiveness of the development of alternative *inexhaustible* sources. In the *UK* about 30% of the primary energy is lost as internal energy in heating the surroundings, during conversion and distribution. The delivered energy, in forms such as petrol, diesel oil, electricity and coke, is used in the following broad categories: industrial production 33%, industrial and domestic space heating 41%, transport 16%, other domestic uses 10%. A further 30% of the primary energy is lost when delivered energy is converted to final-use energy, so that the overall efficiency for the whole conversion process is about 40%. Efficiencies of the complete power station system can be increased significantly by making use of the 'waste' heat by pumping it to nearby housing estates, greenhouse complexes, etc. As space heating is such an important factor, a lot of research has been aimed at reducing energy losses from buildings. A typical *UK* house loses the heat supplied in the following ways: roof 25%, windows 10%, walls 35%, draughts 15%, floors 15%. Improved thermal *insulation*, double glazing and draught excluders can reduce the overall rate of flow of heat from a building by about 60%. This leads to a corresponding saving on fuel consumption. However, it is

important that there should be sufficient *ventilation* to prevent condensation on surfaces and within structures such as roofs.

65.4 POLLUTION

The use of fossil and nuclear fuels contributes to *pollution*, which includes any undesirable effects on the environment. The Earth's temperature remains constant as long as the rate of arrival of radiant energy is equal to the rate of emission. Fuel consumption increases the Earth's energy input, and therefore increases the Earth's temperature. The combustion of fossil fuels produces CO₂ which absorbs energy radiated from the Earth, thereby producing a 'greenhouse effect'. *Thermal pollution* is an important consideration because relatively small temperature changes can produce significant changes in weather patterns and sizes of polar ice caps. Other discharged gases, such as sulphur oxides, produce *acid rain* when they combine with water in the atmosphere, and this may fall at considerable distances from the origin of the gases. If the nuclear power expansion goes ahead as planned, there is real concern regarding radioactive waste disposal (64.7), reactor accidents, sabotage and the dispersal of plutonium from breeder reactors. Nuclear power stations had a very impressive safety record until the explosion and fire at *Chernobyl* near *Kiev* in April 1986. This led to serious radioactive contamination on a dramatically large scale, and high levels of activity were recorded thousands of kilometres away.

65.5 ALTERNATIVE SOURCES OF ENERGY

(A) Hydroelectric

Hydroelectric power is already used widely, especially in mountainous countries such as *Norway* and *Scotland*. The p.e. of water stored in a high-level reservoir behind a dam is converted to k.e. as it falls to a lower level and drives turbines which are used to generate electrical energy.

hydroelectric power \propto vertical height through
which water falls
 \propto rate of flow of water

The efficiency of conversion is high and there is no pollution or residual waste. Careful design, e.g. the *Cruachan Dam* near *Loch Awe* in *Scotland*, reduces potential environmental complaints. In a *pumped storage* station, water is pumped from a low-level reservoir to a higher-level reservoir using electrical energy from continuously operated power stations such as nuclear types, or from

stations operated at various times by winds or tides, during off-peak periods. When the demand for electrical energy becomes high, the water is allowed to flow back to the low-level reservoir. Hydroelectric power contributes only about 2% of the total *UK* electricity supply but, world-wide, it remains an important energy source on a local scale.

(B) Solar

In 30.4(F) the solar constant was quoted as equaling 1.35 kW m⁻² and defined to be the solar power incident on an area of 1 m² placed normal to the radiation at the edge of the Earth's atmosphere. The power reaching the Earth's surface is far less than this (c. 200 W m⁻² in the *UK*) and depends upon factors such as the geographical location, the time of year, the time of day, the altitude and the weather. A major drawback is that peak energy demand occurs in the winter when solar energy is at its least available.

The solar radiation also results in (a) evaporation from the oceans and therefore clouds, precipitation and *hydroelectric* power, (b) convection air currents leading to *wind* power and *wave* power, (c) photosynthesis and plants which contribute to *biofuels*.

Solar heating

This is achieved directly when solar radiation enters windows and is distributed by thermal transfer throughout a building. For this to be effective, the building materials should have high specific heat capacities and the rooms should be well-insulated and double-glazed. *Solar panels* are used to heat water for domestic hot-water supplies and swimming pools. Fig. 65.1 shows a typical design. The orientation of the panel must be calculated to ensure maximum annual output. A *solar furnace*, e.g. *Odeillo* in the *French Pyrenees*, uses a large number of mirrors which follow the motion of the Sun and concentrate its radiation via a huge parabolic reflector onto a small area, thereby producing a power of 1 MW and temperatures of about 3300 K.

Solar electricity

Solar cells use the photovoltaic effect to convert solar energy directly into electrical energy. Each cell is a p-n junction with a large surface area in a suitable semiconducting material. Fig. 65.2 shows a typical design. A standard cell may produce about 1 W 0.4 V at an efficiency of about 10%. Arrangements of connected cells give greater power outputs but costs are still high and total cell

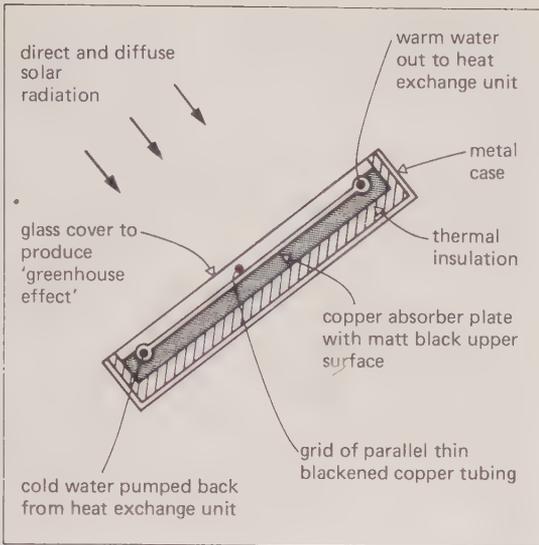


Fig. 65.1. A flat-plate solar energy collector in cross-section.

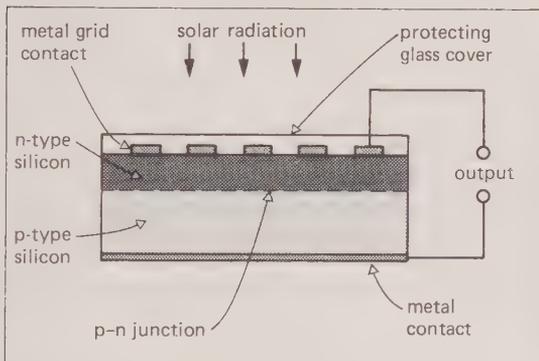


Fig. 65.2. A silicon n-on-p solar cell.

areas required are very large. Their most important use so far has been on the Moon and in space vehicles such as *Skylab*, but they are now finding uses on Earth, e.g. on navigation buoys.

(C) Wave and tide

Wave energy

Sea waves are formed when energy is transferred from wind blowing across the water surface. Storm waves in the *Atlantic*, resulting from high winds, may reach linear wavefront power levels of 1500 kW m^{-1} . In certain regions round the *British Isles*, average power levels of $40\text{--}50 \text{ kW m}^{-1}$ are predicted. An obvious advantage for *Britain* is that

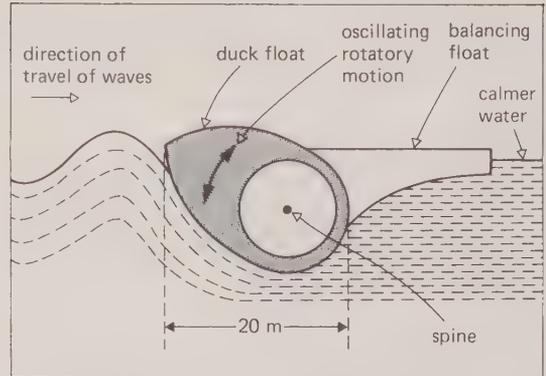


Fig. 65.3. Cross-section of a Salter duck.

the power levels would be at their greatest when the energy demand is at its highest. Numerous devices have been designed to convert wave energy into electrical energy. Fig. 65.3 shows a cross-section of a *Salter duck*, designed by *Dr S. Salter* of *Edinburgh University*, which has achieved efficiencies of over 90% in the laboratory. The device extracts energy through a rotatory motion produced by the incident waves. It is proposed that short duck sections should be mounted on a long central core called a *spine*. One solution to the problem of transport and storage of energy is to use the electrical energy to produce hydrogen by the electrolysis of sea water. At present, however, the construction and maintenance costs of the device would be unacceptably high.

Tidal energy

Tides are generated by the interaction of the gravitational pulls of the Moon and the Sun on the oceans. *Spring* tides result when the lunar and solar pulls reinforce, and *neap* tides occur when the pulls are at right angles to each other. The time and height of high water at a given place can be predicted accurately, giving tidal energy an advantage over wind, wave or solar energy. A tidal power station is sited across a river mouth where there is a large vertical range in tides. The barrage consists of (a) large sluice gates which open to allow sea water to flow in during the *flood* tide and then close at high tide, and (b) water turbines and small gates which are opened when the *ebb* tide has produced an appropriate head of water. The largest tidal power station is at *La Rance* near *St. Malo* in *France* where the average tidal range is 8.4 m and the maximum output is 240 MW. In the *UK*, the *Severn*, *Morecambe Bay* and the *Solway Firth* have been considered but capital costs are extremely high and there are many environmental

problems to be overcome. It is estimated that these three sites could provide 12% of the UK's electrical energy requirements.

(D) Wind

A proportion of the k.e. of the wind can be extracted by a wind turbine and then converted into other energy forms. It can be shown that the power P developed by a wind turbine with a horizontal shaft is given by

$$P = kAv^3$$

where k is a constant that depends upon the density of air, A is the area swept out by the turbine rotor, and v is the wind speed. The requirement for large rotor diameters (up to 100 m) and high wind speeds (usually found in the attractive settings of coastlines, moorlands and mountains) causes environmental objections. These could be resolved by siting the turbines on towers and buoys out at sea where wind speeds would be even higher. The hourly, daily and seasonal variations in wind speed and direction require effective systems of energy storage. Once installed, there would be no fuel costs and no pollution. Again capital costs are high but systems are being designed with outputs of tens of MW. These are relatively low powers but would have benefits locally and would be useful when linked with other systems. An advantage for the UK is that the prevailing wind speeds are highest in the winter when the demand for energy is greatest.

(E) Biofuels

Biofuels are renewable energy sources derived mainly from living things and therefore originating as stored solar energy in the form of starch and cellulose made by photosynthesis. *Biomass* includes

- (a) vegetable matter such as trees, crops and natural vegetation,
- (b) animal dung,
- (c) domestic and industrial refuse.

Biofuels are obtained by *processing* the different types of biomass. *Biogas* is a mixture of methane (60–70%) and carbon dioxide, and is produced when organic material decays by bacterial action under anaerobic (no oxygen) conditions. This happens naturally, e.g. in marsh mud, but the process can be accelerated using suitably insulated tanks called *digesters*. Animal dung and green plants such as *water hyacinth* are particularly effective. The gas can be burned for heating and for operating mechanical engines. *Alcohol* is produced by the fermentation of biomass such as cassava

plants (starch) and sugar cane (sugar). The alcohol can be added to petrol for use as motor fuel. An associated problem is that the waste products from fermentation have polluting effects and come in large quantities. *RDF* (refuse derived fuel) is produced from the combustible components of municipal refuse such as plastic, paper, wood, leather, dust. Loose RDF can be used where suitable incinerating installations are close to the refuse disposal plant. Pelleted RDF can be transported or stored as required. The fuel can be mixed and burned with conventional fossil fuels.

(F) Geothermal

The *geothermal gradient* (the rate at which temperature rises with depth) in the Earth's crust has an average value of 30 K km^{-1} but much steeper values occur in areas of igneous activity. The most important heat source is the decay of long-lived radioisotopes, such as U, Th, K, concentrated in the low-density siliceous rocks like granite, which are situated relatively high in the crust. Regions such as *New Zealand, Italy, Iceland, Japan, and California* have *geothermal fields* where the ground water may be at about 525 K and at a pressure of 30 atmospheres. In these regions hot springs and geysers expel hot water and steam at the surface. Boreholes result in the release of high-pressure steam which can be used to drive electricity generators and to provide industrial and domestic heating. *Reykjavik* is almost entirely heated by energy from geothermal wells. In many other parts of the world, such as *Britain*, there are no similar geothermal fields but two methods of hot water extraction are being investigated.

Geothermal aquifers

These are permeable, sedimentary rocks such as sandstone in areas where the geothermal gradient is about 30 K km^{-1} so that at depths of about 2 km the water would be hot enough ($>333 \text{ K}$) to make extraction worthwhile. Fig. 65.4 illustrates the *Hampshire* and *Paris Basin* schemes. The two shafts should be well separated so that the reinjected cool water does not get back to the extraction point too quickly otherwise the life of the system would be limited.

Hot, impermeable rocks

Some impermeable rocks, such as granite which contains long-lived radioisotopes, have high temperature gradients. The rock is fractured using small explosive charges or water at high pressure (*hydrofracturing*). Ideally the fracturing should produce a vertical, disc-shaped region of cracks.

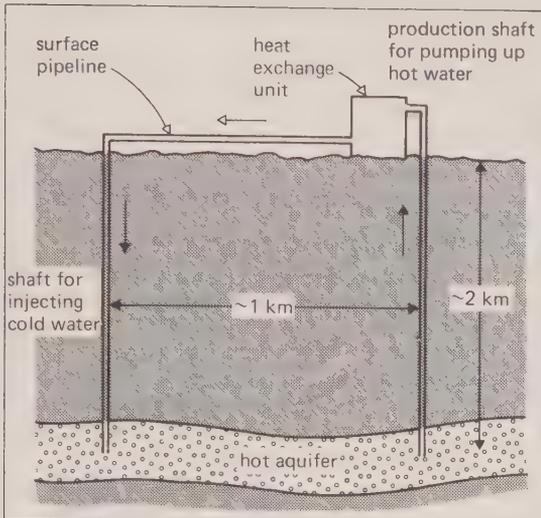


Fig. 65.4. Extraction of hot water from a geothermal aquifer.

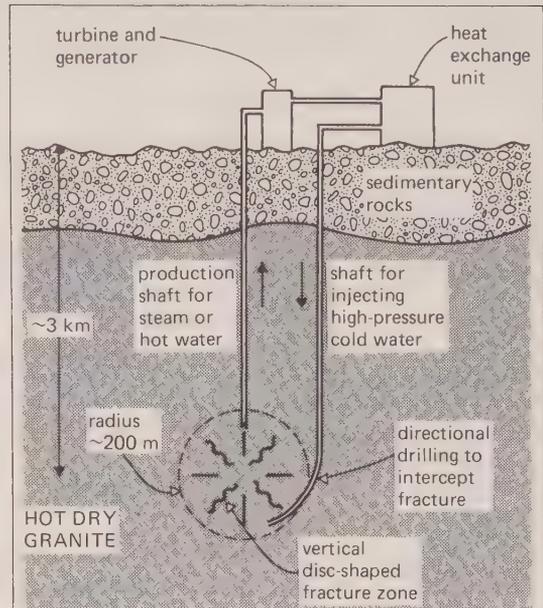


Fig. 65.5. Extraction of geothermal energy from hot, impermeable rocks.

Cold water can be injected so that it flows over the fractured surfaces and extracts heat from the rock before emerging from the top of the production shaft. Fig. 65.5 illustrates the scheme carried out at Los Alamos in New Mexico, USA; similar schemes are being used in Cornwall.

Conclusion

Some of the alternative sources of energy outlined above have considerable potential but their full

development may require another 25–50 years. Furthermore, even full implementation would make only a relatively small contribution to the total world energy requirement. It seems that, if present standards of living are to be maintained, the major energy sources will continue to be the fossil and nuclear fission fuels. A real dilemma exists: should the development of nuclear fission with all the associated risks go ahead, or is mankind willing or able to cope with a prolonged energy shortage?

APPENDIX A. SPECIAL NAMES AND SYMBOLS FOR DERIVED SI UNITS

The quantities in this table are measured in derived units to which special names have been allotted.

<i>Physical quantity</i>	<i>Symbol for quantity</i>	<i>Name of SI unit</i>	<i>Symbol for unit</i>	<i>Immediate definition of unit</i>	<i>Definition in terms of basic units</i>
frequency force work or energy	<i>f</i> or <i>v</i> <i>F</i> <i>W</i>	hertz newton joule	Hz N J	s^{-1} $kg\ m\ s^{-2}$ N m	s^{-1} $kg\ m\ s^{-2}$ $kg\ m^2\ s^{-2}$
power pressure electrical charge	<i>P</i> <i>p</i> <i>Q</i>	watt pascal coulomb	W Pa C	$J\ s^{-1}$ $N\ m^{-2}$ A s	$kg\ m^2\ s^{-3}$ $kg\ m^{-1}\ s^{-2}$ A s
electrical potential difference electrical capacitance electrical resistance	<i>V</i> <i>C</i> <i>R</i>	volt farad ohm	V F Ω	$J\ C^{-1}$ $C\ V^{-1}$ $V\ A^{-1}$	$kg\ m^2\ s^{-3}\ A^{-1}$ $A^2\ s^4\ kg^{-1}\ m^{-2}$ $kg\ m^2\ s^{-3}\ A^{-2}$
electrical conductance magnetic flux density magnetic flux	<i>G</i> <i>B</i> Φ	siemens tesla weber	S T Wb	Ω^{-1} $N\ A^{-1}\ m^{-1}$ $T\ m^2$	$kg^{-1}\ m^{-2}\ s^3\ A^2$ $kg\ s^{-2}\ A^{-1}$ $kg\ m^2\ s^{-2}\ A^{-1}$
inductance activity absorbed dose	<i>L</i> or <i>M</i> <i>A</i> <i>W/m</i>	henry becquerel gray	H Bq Gy	$Wb\ A^{-1} =$ $V\ A^{-1}\ s$ s^{-1} $J\ kg^{-1}$	$kg\ m^2\ s^{-2}\ A^{-2}$ s^{-1} $m^2\ s^{-2}$

APPENDIX B. RECOMMENDED VALUES OF SELECTED PHYSICAL CONSTANTS

These are given to only four significant figures, as at this level we seldom need to use more. Nevertheless, the reader should examine the last column carefully.

<i>Physical quantity</i>	<i>Symbol</i>	<i>Value</i>	<i>Approximate uncertainty in parts per million</i>
speed of light in a vacuum	c	$2.998 \times 10^8 \text{ m s}^{-1}$	0.004
permeability of a vacuum	μ_0	$4\pi \times 10^{-7} \text{ H m}^{-1}$ exactly	—
the permittivity constant	ϵ_0	$8.854 \times 10^{-12} \text{ F m}^{-1}$	0.01
mass of proton	m_p	$1.673 \times 10^{-27} \text{ kg}$	5
mass of neutron	m_n	$1.675 \times 10^{-27} \text{ kg}$	5
mass of electron	m_e	$9.110 \times 10^{-31} \text{ kg}$	5
charge of proton	e	$1.602 \times 10^{-19} \text{ C}$	3
Boltzmann constant	k	$1.381 \times 10^{-23} \text{ J K}^{-1}$	32
Planck constant	h	$6.626 \times 10^{-34} \text{ J s}$	5
Stefan–Boltzmann constant	σ	$5.670 \times 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$	125
Rydberg constant	R_∞	$1.097 \times 10^7 \text{ m}^{-1}$	0.08
Avogadro constant	N_A	$6.022 \times 10^{23} \text{ mol}^{-1}$	5
molar gas constant	R	$8.314 \text{ J mol}^{-1} \text{ K}^{-1}$	31
Faraday constant	F	$9.648 \times 10^4 \text{ C mol}^{-1}$	3
gravitational constant	G	$6.672 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$	615

BIBLIOGRAPHY

- (a) *Essential Principles of Physics* needs to be amplified by more detailed texts. The following are recommended as being particularly useful throughout an A-level course.

Physics: A Textbook for Advanced Level Students, by Duncan, 2nd edition (John Murray, 1987).

Physics, by Akrill, Bennet and Millar (Arnold, 1979).

Physics: Concepts and Models, by Wenham, Dorling, Snell and Taylor, 2nd edition (Addison–Wesley, 1984).

Advanced Physics Project for Independent Learning: Students' Guides. Revised edition (John Murray, 1988).

- (b) *Practical Physics*, by Squires, 3rd edition (Cambridge University Press, 1984). This is not a catalogue of experiments but emphasizes the development of a critical approach to experimental work.

- (c) The *Revised Nuffield Advanced Physics Course* makes a major contribution to school physics teaching. The *Students' Guides* (and *Teachers' Guides*) contain comprehensive lists of reference material.

- (d) The following, listed in approximate order of increasing sophistication, comprise complete texts which lead up to first- and second-year standard of British universities. They discuss fundamentals very carefully, and this makes them invaluable for bridging what is sometimes a difficult gap between school and degree physics. Of the many texts available, these reflect very much the authors' personal choice.

College Physics, by Sears, Zemansky and Young, 6th edition (Addison–Wesley, 1984).

University Physics, by Sears, Zemansky and Young, 7th edition (Addison–Wesley, 1987).

Fundamentals of Physics, by Halliday and Resnick, 3rd edition (Wiley, 1988).

Elementary Classical Physics (2 vols), 2nd edition (1973), and *Elementary Modern Physics*, 3rd edition (1980), by Weidner and Sells (Allyn and Bacon).

Physics (2 parts), by Halliday and Resnick (Wiley, 1978).

Fundamental University Physics (3 vols), by Alonso and Finn, 2nd edition (Addison–Wesley, 1979).

Berkeley Physics Course (5 vols), (McGraw-Hill, 1967 to 1984).

Feynman Lectures on Physics (3 vols), by Feynman, Leighton and Sands, (Addison–Wesley, 1963 to 1965).

- (e) The following are useful books which do not form part of a complete text.

The Structure of Matter, by Guinier (Arnold, 1984).

Gases, Liquids and Solids, by Tabor, 2nd edition (Cambridge University Press, 1979).

- Properties of Matter*, by Flowers and Mendoza (Wiley, 1970).
Newtonian Mechanics, by French (Norton, 1980).
Vibrations and Waves, by French (Van Nostrand Reinhold, 1982).
Special Relativity, by French (Van Nostrand Reinhold, 1982).
Forces and Particles, by Pippard (Macmillan, 1972).
Optics, by Hecht, 2nd edition (Addison–Wesley, 1987).

- (f) Further suggestions for more general reading can be found in the authors' *Questions on Principles of Physics* (John Murray, 1979).

To that list should be added the publications of the **Open University**, details of which may be obtained from

The Marketing Division
The Open University Educational Enterprises
12 Cofferridge Close
Stony Stratford
Milton Keynes
MK11 1BY

INDEX

Key 316 refers to a main entry 217*d* refers to a definition 184*t* refers to a table

A

- aberration**, 256, 267
chromatic 268
spherical 267
- abscissa** 26
- absolute error** 16
- absolute measurement**,
current 421
p.d. 357
resistance 432
- absolute permeability** 460*d*
- absolute permittivity** 366*d*
- absolute refractive index** 253*d*
- absolute temperature** 181
- absolute zero** 181
- absorbed dose rate** 524
- absorptance**, spectral 245*d*, 247
- absorption factor** 245*d*
- absorption of radiation** 243
- absorption spectra** 247, 271, 274
- a.c. circuits**,
capacitive 448
inductive 446
resistive 445
RLC parallel 451
RLC series 449
- acceleration**,
angular 55*d*
average 28*d*
centripetal 33, 39
circular motion 39
due to gravity 38, 172
in s.h.m. 83
instantaneous 29*d*
of charge 282
projectiles 30
radial 32, 56
tangential 56
as a vector 29
- acceptor impurities** 381
- accuracy** 15
- achromatic lens (doublet)** 268, 270
- achromatic prism** 270
- acid**, fatty, molecular diameter of 136
- actinium series** 518
- activity of radioactive source** 519
- adder**,
full 491
half 490
- addition of vectors** 18, 445
- adhesion** 154
- adiabatic change**,
entropy changes in 219
equation for 209
reversible nature of 209
sound wave as 327
work done during 210
- adiathermanous** 243
- aerial** 103
- aerodynamic lift** 77
- air cell** 255
- air column**, vibrating 94, 329
- Airy* 176
- Alcomax* 461
- alpha emission** 518, 523, 525
- alpha particles**,
artificially produced 526
bombardment by 526
charge 523
deflection in electric and magnetic fields 522
detection 519
energy 521
identification 521
ionizing effect 519
range 521
scattering of 511
speeds 523
- alternating current**,
generator of 428
measurement of 443
motors 434
- alternator** 435
- Amagat* 232
- americium** 519
- ammeter**,
a.c. 443
d.c. 394
moving-coil 418
moving-iron 418
shunt 394
thermocouple 443
- amorphous solids** 134
- ampere** 9*d*, 372, 416
absolute measurement of 421
realization of 11
- Ampère's Law*,
in a medium 465
in vacuo 415
- amplification factor** 485*d*
- amplifier**,
b.j.t. 486.
d.c. 364
differential 497
f.e.t. 485
inverting 496
non-inverting 497
operational 495
practical 486
summing 498
voltage 496
- amplitude**,
of s.h.m. 83
of a wave 103
- analogue computing** 495, 501
- analogue electronics** 500
- analogues** 454
- analyser** 318
- anchor ring**, the Rowland 459
- Anderson* (speed of light) 291
- Andrews's experiments* 231
- angle**,
critical 108, 254, 326
definition of 22
of attack 77
of contact 152*d*, 155, 158
of dip 429
of incidence 252
of minimum deviation 108, 257
of reflection 252
of refraction 252
of rotation 318
polarizing 315*d*, 318
trigonometric relationships 22
visual 275*d*
- angular acceleration**,
average 55*d*
constant 55
instantaneous 55*d*
- angular displacement** 55
- angular frequency** 85, 441
- angular impulse** 61, 420
- angular magnification** 275*d*
- angular momentum** 61*d*
conservation of 61
orbital 62
photon 288
spin 61
vector nature of 61
- angular speed**,
average 31*d*
instantaneous 31*d*
- angular velocity** 31*d*
- anharmonic oscillation** 102, 187, 325
- anisotropic** 319
- annihilation** 338, 511
- anode** 473, 508
- anticlastic surface** 156
- antimony** 380
- antineutrino** 523, 524
- antinodal lines** 112
- antinode** 111, 289, 330

- antiparticle** 524
aperture, lens or mirror 267, 268
apparent depth 254
apparent expansion 188
Arago 318
archaeological dating 530
Archimedes's Principle 72
area as a vector 71
areal moment, second 150
armature 433
arsenic 380
artificial radioactivity 526
artificial transmutation 526
astable multivibrator 492
astronomical telescope,
 reflecting 279
 refracting 278
atmospheric pressure 73
atom, 126
 Bohr 514, 515
 energy levels in 376, 508, 514, 515, 526
 electric charge of 126
 ionization energy of 517
 models of 513
 quantum mechanics and 514
 structure of 126
atomic bomb 530
atomic clock 12
atomic energy 530
atomic mass,
 measurement of 513
 number *A* 509*d*, 526
 relative 509
atomic number *Z* 509*d*, 526
atomic spectra 271, 515
atomic weight 509
atomicity of gases, 211
 and speed of sound 328
attracted disc electrometer 356
audible range 324
avalanche effect 479
average value 23, 199, 442
Avogadro constant, the 135, 196*d*, 539
Avogadro's Law (Hypothesis) 4, 200
axial vector 19
axis, principal 259, 266
- B**
back e.m.f. 431, 434, 436
background radiation 518
Bainbridge mass spectrograph 513
balance,
 ballistic 12, 52
 beam 12
 current 394, 421, 443
 inertial 12, 68
 spring 12, 37
balancing columns 188
ballistic balance 12, 52
ballistic galvanometer,
 theory of 420
 uses of 421
Balmer series 515
band spectra 273
band theory of conduction 376, 381
banking of track 40
bar magnet, electromagnetic
 moment of 408
Barkhausen effect 464
Barkla 514
barometer 73
barrier, potential 477
Barton's pendulums 93
base (transistor) 483
basic quantity 4
battery,
 e.m.f. of 388*d*, 396
 internal resistance of 389, 396
beam balance 12
beat frequency 115, 327
beat period 115
beats,
 general 114, 176
 light 289
 sound 327
Becquerel 518
becquerel 519*d*
Belham measurement of resistance 433
bending moment 149
Bernoulli's equation 75
beryllium 526
beta decay,
 negative 523
 positive 523
beta particles,
 absorption of 521
 charge of 523
 deflection in electric and magnetic fields 523
 detection of 519
 identification 468, 523
 ionizing effect 519
 pair production 510
 range 523
 specific charge 468, 470
 speed variation 522
bifilar suspension 63
binary addition 490
binary counter 494
binary counting 494
binding energy 510
binding energy per nucleon 512
binomial theorem 21
biofuels 536
biomass 536
Biot-Savart Law 410
bipolar transistor (b.j.t.) 483
biprism, Fresnel 296
birefringence 317
bistable multivibrator 493
Bitter photograph 463
black body, 244
 absorber 244
 emitter 245
 energy spectrum of 248
 radiator 245
 Sun as a 246, 248
 temperature 245
Bloch walls 463
blocking capacitor 485
blooming of lenses 299
blue of sky 118, 217, 317
Bohr's theory of the atom 515
boiling 228
boiling point,
 effect of pressure on 228
 standard 228*d*
bolometer, Langley's 243
Boltzmann constant 211, 219
bomb,
 atomic 530
 hydrogen 530
boron 526
Bouguer 176
bound charges 373
bound system 510
boundary conditions, 25
 general 111
 in light 293
boundary layer 165
Bourdon gauge 73
bow wave 123
Boyle temperature 233
Boyle's Law, 197
 derived from microscopic definition of ideal gas 199
 deviations from 232
 macroscopic definition of ideal gas 197
Boys' image 263
*Boys' method for *G** 177
Boys' radiomicrometer 243
Bradley 290
Bragg equation 312
Bragg's Law, 312
 use for measurement of N_A 135
Brahe, Tycho 168
bremsstrahlung 508
Brewster's Law 315
bridge,
 De Sauty 362
 metre 398
 rectifier 420, 443
 Wheatstone 376, 397
brittle fracture 143
brittleness 143
broadening of energy levels 376
broadening of spectral lines 122
Brownian movement 135
bubble, excess pressure in 155
bubble chamber 520
buffer circuit 493
bulk modulus 141*d*

- adiabatic 148, 327
isothermal 148
Bunsen ice calorimeter 226
- C**
- caesium** 473
calcite 317
calculus 23
calibration,
 curve for instrument 25
 of potentiometer 395
 points for thermometer 181, 228
Callendar 192
calorimetry 190
camera,
 lens 264
 pinhole 264
canal rays 508
candela 10
capacitance 358*d*
 measurement of 361, 421
 of coaxial cylinders 360
 of Earth 359
 of isolated sphere 358
 of parallel plates 359
capacitative circuits 392, 448
capacitors,
 blocking 485
 charge stored in 359
 coaxial cylinder 360
 electrolytic 360
 energy of 363
 exponential discharge of 393
 force on plates of 356
 guard-ring 241, 357, 360
 in parallel 362
 in series 263
 oscillatory circuit 453, 486
 parallel-plate 359, 360, 365, 369
 reservoir 456
 smoothing 458
 tuning 453, 486
 use for measurement of ϵ_0 361
capacity,
 heat 189*d*
 molar heat 190, 194*d*, 207*d*
 specific heat 189*d*
capillarity, 157
 use for measurement of γ 159
carbon-12, use for
 carbon dating 530
 definition of mole 9, 196
 definition of relative atomic and
 molecular masses 196, 509
carbon dioxide, isothermals for 232
carburettor 78
Carnot cycle 221
Carnot's theorem 222
cascade process 235
Cassegrain reflector 279
cathode 468, 473
cathode ray oscilloscope,
 construction 473
 uses 475
cathode rays,
 identification of 468
 in electric and magnetic fields 469
 speed of 471
caustic curve 254
Cavendish 177
cavity radiator 244
cells,
 e.m.f. 388*d*, 396
 in parallel 389
 in series 389
 internal resistance 386, 389, 396
 photoelectric 506
 photovoltaic 506
 standard 394
 Weston 394
Celsius scale 184*d*
centigrade 184
central force 169, 337
centre, optic 259
centre of curvature 156, 259, 266
centre of gravity, 69*d*
 and stability of equilibrium 69
centre of mass 68*d*
 location of 68
centre of oscillation 63
centre of suspension 63
centrifugal force 40
centripetal acceleration 33, 38
centripetal force 39, 516
Čerenkov radiation 123
Chadwick 526
chain reaction 529
chamber,
 bubble 158, 520
 cloud 158, 520
 diffusion 520
 ionization 519
change,
 adiabatic 206, 209, 327
 isothermal 206, 208
 of phase 130, 225
Chappuis 185
characteristics,
 electric 375, 456
 field effect transistor 481, 483
 junction diode 478
 junction transistor 484
charge,
 acceleration of 282, 507, 516
 conservation 338, 389, 449, 510
 density 348
 electric 336, 372*d*
 induced 346, 353, 359, 428
 measurement of 338
 on electron 337, 471
 on isolated conductor 354
 quantization of 337
 separation of 359, 477
 sign of 336, 356, 403
charge sensitivity 420*d*
charging by induction 346, 356
Charles's Law 196
chemical bond 130
chemical properties of atom 509
choke, smoothing 458
chromatic aberration 268
chromosphere 273
circle of least confusion 267
circuits,
 capacitative 392, 448
 energy transfer in 385
 inductive 437, 446
 integrated 500
 Kirchhoff's rules for 389
circular coil, field of 412
circular motion,
circular motion, in a vertical plane
 46
 non-uniform 56
 uniform 32, 39
circular orbits 174, 514
circular polarization 102
Claude process 235
Clausius-Clapeyron equation 229
Clement and Désormes 209, 212
clocks 12
clocked bistables 494
cloud chamber 158, 520
coating of lenses 300
coaxial cylindrical capacitor 360
cobalt 519
Cockcroft and Walton 526, 528
coefficient of friction 36*d*
coefficient of performance 222
coefficient of restitution 52*d*
coefficient of viscosity 141, 163*d*, 203
coercive force 461
coherence,
 general 112
 in light 283, 292
cohesion 154
coil,
 field along axis of 412
 Helmholtz 413, 471
 moving-coil galvanometer 418
 torque on 406
collector 484
collimator 270
Collins 235
collision,
 elastic 52
 electron and lattice 379
 explosive 52
 frequency of gas molecules 129
 inelastic 52, 198, 378, 515
 superelastic 52, 198
colours,
 and frequency 254
 of bloomed lenses 299

- of diffraction fringes 307
- of interference fringes 297
- of sky 118, 217, 317
- of spectrum,
 - grating 306
 - prism 269, 306
 - of thin films 299
- column of liquid**, pressure due to 72
- combination of errors** 16
- combination of lenses** 262, 268
- combination of prisms** 270
- combination of s.h.m.** 90, 475
- common-emitter circuit** 484
- common logarithm** 21
- common temperature** 184
- commutator** 433
- comparison of**
 - e.m.f.s 396
 - frequencies 90, 327, 332, 475
 - resistances 396, 397
 - speed of sound in gases 332
- components of a vector** 18
- components of Earth's magnetic field** 429
- composite materials** 147
- compound microscope** 278
- compound pendulum** 62
- compound slab** 239
- compound-wound electrical devices** 433
- compressibility** 140*d*
- compression** 323
- compression ratio** 223
- computers**,
 - analogue 501
 - digital 501
- Compton effect** 507, 510
- concave mirror** 266
- concave surface** 260
- concrete** 147
- condensation** 130, 323
- conditions for visibility of a fringe system** 292
- conductance**, 374
 - drain 482*d*
 - mutual 482*d*
 - thermal 238
- conduction**, band theory of 376, 381
- conduction, electrical**, in
 - gases 373, 515
 - liquids 373
 - metals 373, 378
 - semiconductors 373, 377, 379
- conduction electrons** 131, 236, 373, 376, 378
- conduction, thermal**,
 - measurement 241
 - mechanism 236
- conductivity**,
 - electrical 374*d*, 374*t*
 - thermal 237*d*, 238*t*
- conductors**,
 - distribution of charge on 345, 355
 - electric field close to surface of 348
- conservation laws**,
 - angular momentum 61
 - electric charge 338, 389, 449, 510
 - linear momentum 50, 510
 - mass 74
 - mass-energy 510
 - mechanical energy 46
 - total energy 43, 389, 449
- conservative field of force**,
 - electric field as 337, 427
 - gravitational field as 46, 173
- constant pressure head** 167, 192
- constant volume gas thermometer** 181, 184, 186
- contact**, angle of 152*d*, 155, 158
- contact p.d.** 391, 476
- continuous-flow**,
 - Callendar and Barnes 192
 - Henning 226
 - Scheel and Heuse 208
- continuous spectrum**,
 - light 273
 - β -particle energies, 522
 - X-rays 508
- contour fringes** 300
- control systems** 489
- convection**,
 - forced 193
 - natural 191, 193
- convention**, sign, for light 260, 266
- convention for direction of electric current** 372
- converging lens** 101, 259
- convex mirror** 266
- Coolidge X-ray tube** 507
- cooling**,
 - by evaporation 228
 - five-fourths power law 193
 - method of, for measurement of c 193
 - Newton's Law of 193
 - on reversible adiabatic expansion 209
- cooling correction** 190
- coordination number** 130
- Copernicus** 168
- copper**, conduction in 374*t*
- copper losses** 431
- cork-screw rule** 411, 427
- corona discharge** 355
- corpuscular theory of light** 107, 290
- correction**,
 - cooling 190
 - of real-gas temperature scale 182, 185
- cosine law** 22
- cosmic rays** 518
- coulomb**, the 337, 372*d*
- Coulomb's Law** 336
 - and dielectrics 367
 - and Gauss's Law 345
 - test of 337, 511
- counter**, Geiger-Müller 519
- couple**,
 - on a coil in a magnetic field 406
 - on an electric dipole 342
 - on a permanent magnetic dipole 407
 - work done by 59, 342, 408
- coupling, intermolecular** 236
- covalent binding** 130
- covolume** 234
- creep fracture** 145
- critical angle** 108, 254*d*, 326
- critical damping** 92, 420
- critical isothermal** 231, 235
- critical mass** 529
- critical pressure** 231
- critical temperature** 231, 235
- critical velocity** 163, 165
- critical volume** 232
- cross product of vectors** 19, 66, 401
- cross section**, reaction 528
- crossed Nicol prisms** 317
- crosswires** 270, 273
- crystal**,
 - energy of 129, 133
 - spacing of atoms in 135, 311
 - structure of 132
- cubic expansivity** 188*d*
- Curie** 518
- Curie temperature** 464
- current**,
 - absolute measurement of 421
 - eddy 430
 - electric, in
 - liquids 373
 - metals 373, 378
 - semiconductors 373, 379, 380
 - ionization 520
- current amplification factor** 485*d*
- current balance** 394, 421, 443
- current carriers** 373, 403
- current density**,
 - electric 374*d*, 379, 404
 - energy 202*d*
 - molecular 201*d*
 - momentum 203*d*
- current element** 411
- current sensitivity** 419*d*
- current transfer ratio** 485
- curvature of conductor**, relation
 - with surface density of charge 355
- customary temperature** 184
- cut-off frequency** 506
- cycle**, 83
 - Carnot 221
 - hysteresis 141, 461
 - on indicator diagram 205
- cyclotron** 403
- cyclotron frequency** 403
- cylindrical capacitor** 360

- D**
- Dalton's Law of partial pressures** 200
- damping**,
eddy current 430
effect on forced vibration 93
in oscillatory circuit 454
in s.h.m. 91
- data**,
analogue 501
digital 501
- dating**,
archaeological 530
geological 530
- Davison and Germer** 288
- d.c. amplifier** (electrometer) 338, 364, 520
- d.c. generator** 428, 433
- d.c. meter** 418
- d.c. motor** 434
- de Broglie** 288, 312
- De Sauty bridge** 362
- dead-beat movement** 92, 420
- dead space** 185
- decay**, radioactive 526
- decay constant** 527*d*
- decrement**, logarithmic 92*d*
- defects of optical images** 267
- definition of physical quantity** 4
- deflection**, magnetic, of
 α -particles 522
 β -particles 522
cathode rays 469
- degradation of energy**, principle of 223
- degree Celsius** 184*d*
- degree of freedom** 210
- demagnetization** 461
- density**,
charge 347
effect of temperature on 188
number 196*d*, 201*t*, 378, 404
relative 71*d*
- deoxyribonucleic acid (DNA)** 132
- dependent variable** 25
- depletion layer** 477
- depth of field** 264
- depth of focus** 265
- derivatives**, table of 23
- derived quantity** 4
- detection**,
 α -particles 519
 β -particles 519
electromagnetic waves 285
 γ -rays 285, 519
- deuterium** 526, 529
- deviation**,
by a prism 257
by a prism of small angle 258
method for deriving lens equation 260
minimum 108, 257, 271
without dispersion 270
- diagram**, free-body 39, 65
- diamagnetism** 459, 463
- diameter of atom** 126, 136, 512, 517
- diameter of molecule** 136, 203
- diameter of telescope aperture** 268, 279, 311
- diathermanous** 243
- dichroism** 316
- dielectric behaviour** 366
- dielectric constant**, and
capacitors 365
Coulomb's Law 367
Gauss's Law 367
- dielectric strength** 366*d*, *t*
- diesel cycle** 223
- differential air thermometer** 243
- differential amplifier** 497
- differential thermocouple thermometer** 241
- diffraction**, 116
and interference 110, 294
by a circular aperture 307
by an edge 307
by a mirror 268
by a rectangular aperture 307
by a slit 307, 309
effect on double slit interference 294
electron 288, 312
Fraunhofer 304, 309
Fresnel 304, 306
 γ -ray 524
X-ray 135, 311, 508
- diffraction grating**, 117
mechanism of 304
overlapping orders 305
reflecting 306
resolving power 306
spectrum 305
transmission 304
- diffusion**,
gaseous 200, 201, 217
Graham's Law of 200
- diffusion cloud chamber** 158, 520
- diffusion coefficient** 202*d*
- digital electronics** 500
- dimensional analysis** 20
- dimensions** 5
uses of 19
- dimer** 234
- diode**
junction,
characteristic 478
construction 476
use as a rectifier 456
light-emitting 479
photo 479
point-contact 477
thermionic 473
Zener (voltage regulator) 479
- dip**, angle of 429
- dipole**,
electric 341, 342, 366
electromagnetic moment 407*d*, 409
magnetic 407
- direction of vector** 17
- direct vision spectroscope** 270
- directly heated cathode** 473
- disappearing filament pyrometer** 249
- discharge**,
glow 373
Townsend 373
- discharge of capacitor through resistor** 393
- discharge through gases** 373, 508
- disintegration**, radioactive 526
- disintegration energy** 528
- dislocations** 132, 144
- disorderly flow** 74, 161, 163
- dispersion**, 101
angular 269
by grating 305
by prism 269
of sound 326
of water waves 102
without deviation 270
- dispersion forces** 132
- dispersive power** 269*d*
- displacement**,
angular 56
electric 368*d*
in s.h.m. 83
law (Wien) 248
linear 28*d*
-time graphs 29, 85
- distortion of wavefront** 303
- distribution**,
Maxwellian 198
statistical 215
- distribution of energy in black body spectrum** 248
- diverging lens** 258, 261
- domain**, magnetic 463
- donor impurity** 380
- doping of semiconductor** 380
- Doppler effect** 119, 326
- Doppler shift** 121
- dose equivalent** 524
- dot product of vectors** 19
- double refraction** 317
- double slit** 112, 294
- doublet**, achromatic 268, 270
- drain** (transistor) 480
- drain conductance** 482*d*
- drain resistance** 482
- drift speed**,
and diffusion 201
and *Hall* effect 404
of electrons in metals 374, 378
- drops**, shape of 153
- drop-weight method** 160
- ductile fracture** 142

- ductility 143
Dulong and Petit 194
Dunnington 471
dust-tube 212, 333
dynamic equilibrium 227, 391, 473
dynamo,
 theory 428
 winding 433
dynamometer, 422
 joulemeter 423
 wattmeter 423, 452
- E**
 ϵ_0 , measurement of 361
ear, sensitivity of 324
Earth,
 age 530
 gravitational field 172, 352
 magnetic field 401, 429
 mass 172
 rotation 12, 38, 121
Earth inductor 429
earthing 359
earthquake waves 97
ebonite, electric charge 336
eddy currents,
 damping 420
 direction of 430
 heating by 430
 in transformer 431
edge, diffraction at a straight 307
effective current 442
effective dose 524
efficiency, of
 γ -ray detector 520
 transformer 431
efficiency, thermal 222*d*
Einstein,
 mass-energy equivalence 290, 509,
 512, 529
 photoelectric effect 504, 510
 random walk problem 135
 theory of relativity 2, 122, 168, 290
elastic collision 52, 198
elastic limit 142
elasticity, moduli of 138
electric calorimetry 191, 385
electric charge 336, 372*d*
electric circuit 389
electric current 372
electric current density 374*d*, 379,
 404
electric deflection,
 α - and β -particles 522
 γ -rays 522
 cathode rays 469
electric dipole,
 electric field of 341
 energy of 342
 induced 366
 moment of 342*d*
 torque on 342
electric displacement 368*d*
electric displacement flux 368*d*
electric field, 339*d*, 411
 and potential 350
 energy storage in 363
 in electromagnetic wave 282
 of charged conductor 348
 of cylindrical charge 347
 of dipole 341
 of point charge 340
 of spherically symmetric charge
 347
 within a conductor 346, 347
electric field flux 344*d*, 368
electric field line 340
electric mobility 374*d*, 379, 380
electric specific charge 468, 470
electric wind 355
electrical conductivity 374*d*; *t*
electrical energy, transmission of
 386
electrical forces 337, 367
electrical potential 349*d*
electrical resistivity 374*d*; *t*
electrical resonance 94, 452
electrical standards 394
electrification by rubbing 336
electrodes 468, 473, 508
electrolysis, use for measuring N_A
 136
electrolytic capacitor 360
electromagnet, material for 461
electromagnetic induction 424
 laws 426
 applications 430
electromagnetic interaction 3, 127,
 339, 400, 410
electromagnetic moment 407*d*, 408,
 465
electromagnetic oscillations 94, 453
electromagnetic radiation 282
electromagnetic spectrum 284*t*
electromagnetic waves, 282
 properties 285, 288, 507, 523
 speed 107, 289, 361, 416, 539
electrometer,
 attracted disc 356
 d.c. 338, 364, 520
 gold leaf 256
electromotive force, 388*d*
 calculation of induced 424
 comparison of cell 396
 in rotating coil 428
 of standard cell 394
 thermoelectric 185, 391, 397
electron gun 473
electronic charge 337, 471
electronic mass 468
electronic rest-energy 510, 511
electrons,
 emission of 468, 473, 503, 523
 identification of 468
 ionization by 521
 magnetic deflection of 469
 orbital 514
 speed of 469, 471
electron mobility 379
electron-photon interaction 510
electrons, diffraction of 288, 312
electrons in atoms,
 Bohr description 514, 515
 wave mechanical description 514,
 515
electrons in metals,
 electrical conduction by 373, 378
 thermal conduction by 236
electronvolt 350*d*
electroscope,
 charging by induction 356
 gold leaf 356
electrostatic induction 346, 359, 367
electrostatic voltmeter 356
element, atomic number of 509
elliptical polarization 102
emission,
 field 468
 photoelectric 503
 secondary 468
 spectrum 271, 515
 thermionic 473
emissivity,
 spectral 245*d*, 247
 total 247
emitter 484
end-correction,
 air columns 329, 332
 metre bridge 398
energy,
 α -particle 521
 availability of 224
 β -particle 523
 binding 510
 biofuel 536
 chemical 43, 385
 degradation of 223
 equipartition of 135, 211
 free surface 151*d*
 geothermal 536
 hydroelectric 534
 internal 45, 189*d*, 205, 207, 385
 kinetic 44*d*
 nuclear 43, 528
 photon 286, 504, 524
 potential 44*d*
 elastic 44, 146, 420
 electric 349*d*, 363
 gravitational 44, 173
 intermolecular 127, 187, 233
 magnetic 408, 439
 solar 534
 strain 146
 surface 151*d*
 tide 535
 wave 535

- wind 536
 - zero-point 181
 - energy conservation law** 43
 - energy consumption** 533
 - energy conversions** 45
 - energy current density** 202*d*
 - energy density** 146
 - energy distribution in black-body,**
spectrum 248
 - energy levels in atom** 376, 515
 - broadening of 376
 - energy loss in**
 - hysteresis cycle 141, 461
 - transformer 431
 - transmission 386
 - energy resources** 532
 - energy storage,**
 - capacitor 363
 - inductor 438
 - stretched wire 146
 - energy transfer** 45
 - engines, heat** 210, 221
 - entropy,**
 - changes during *Carnot* cycle 221
 - degradation of energy 224
 - macroscopic ideas and 218
 - microscopic ideas and 219
 - the second law and 219
 - equation of motion** 36, 57, 85
 - equation of state,**
 - ideal gas 197, 200, 208, 234
 - real gas (*van der Waals'*) 234
 - equations, checking dimensions of** 19
 - equilibrium,**
 - conditions for
 - rotational 66
 - translational 66
 - mechanical, types of 69
 - statistical 215
 - thermal 180
 - thermodynamic 204
 - equipartition of energy** 135, 211
 - equipotential surface** 173, 350
 - equivalence of mass and energy** 509
 - equivalent focal length** 262
 - erecting telescope** 279
 - errors** 13
 - escape speed** 174
 - evaporation** 227
 - excess pressure across a curved**
surface 155
 - excess pressure methods for γ** 159
 - exchanges, *Prévo*s't theory of** 246
 - excitation potentials** 515
 - exclusion principle** 3, 376
 - exitance,**
 - spectral radiant 245*d*, 246
 - total radiant 245*d*
 - expanding Universe** 121
 - expansion, thermal** 187
 - expansion of gas, work done during** 205, 207
 - expansivity,**
 - cubic 188*d*
 - linear 187*d*
 - experimental uncertainty** 13
 - explosive collisions** 52, 198
 - exponential (e^x)** 22
 - decay and growth 392, 437, 527
 - exposure meter** 506
 - exposure rate** 524
 - exposure time** 264
 - extensive quantity** 219
 - extraordinary ray** 317
 - extrinsic semiconductor** 380, 381
 - eye-lens** 276, 277, 278, 279
 - eyepiece** 270, 276
 - eye-ring** 279
- F**
- f*-number** 264
 - fact** 4
 - factor,**
 - absorption 245*d*, 246
 - power 452*d*
 - reflection 245*d*
 - Fajans*** 525
 - farad, the** 358*d*
 - Faraday* constant** 10, 136*d*, 539*t*
 - Faraday's* ice-pail experiment** 347
 - Faraday's* laws of electromagnetic induction** 426
 - fatigue fracture** 145
 - fatty acid** 136
 - feedback** 454
 - acoustic 499
 - factor 497
 - negative 496
 - positive 487, 499
 - ferrite** 430, 461
 - ferromagnetic material** 430, 436, 449, 461
 - ferromagnetism** 463
 - Féry's* radiation pyrometer** 248
 - Fick's* Law** 201
 - fibre optics** 255
 - field,**
 - conservative 46, 173, 337, 427
 - electric 339, 411
 - flow 74, 78, 411
 - force 3, 169, 400, 411
 - gravitational 169, 352, 411
 - magnetic 400, 410, 411
 - field coil** 433
 - field effect transistor (f.e.t.)** 481
 - field emission** 468
 - field strength,**
 - electric 339*d*
 - gravitational 170*d*
 - magnetic 465*d*
 - filament** 473
 - film, thin, colours in** 299
 - film badge** 525
 - filter** 249
 - first law of thermodynamics** 194, 205, 210, 219
 - fission** 528
 - five-fourths power law** 193
 - fixed points of I.P.T.S.** 183
 - Fizeau,***
 - measurement of small distances 301
 - speed of light 290
 - Fleming's* Rules,**
 - left-hand motor 401, 425
 - right-hand dynamo 426
 - flexural rigidity** 150
 - flight** 77
 - flint glass** 269
 - flip-flop** 493
 - flotation, principle of** 72
 - flow,**
 - continuous 190, 192, 226
 - irrotational 74
 - laminar 74, 161, 164
 - streamline 74, 161, 164
 - turbulent 74, 76, 163, 167
 - fluctuation, proportional** 217
 - fluid,**
 - flow of non-viscous 74
 - flow of viscous 161
 - motion through viscous 165
 - pressure in moving 75
 - pressure in static 71
 - structure of 129, 152, 161
 - fluidity** 163*d*
 - fluorescence** 271, 474, 519
 - flux,**
 - electric displacement 368*d*
 - electric field 344*d*, 368
 - linkage (magnetic) 424*d*
 - magnetic 401*d*, 424, 426, 428
 - flux density, magnetic** 401*d*, 410, 428
 - flyback** 474
 - flywheel** 61
 - focal length,**
 - lens 259*d*, 261, 263
 - mirror 266
 - thin lenses in contact 262
 - focal plane** 259
 - focal point** 100, 101, 256
 - focus, principal** 259
 - forbidden zone** 377
 - force, 35**
 - action-reaction 35
 - between parallel conductors 357
 - central 169, 337
 - centrifugal 40
 - centripetal 40, 516
 - coercive 461
 - conservative 46, 173, 337, 427
 - electric 337, 367
 - field 152
 - frictional 35
 - impulsive 49
 - intermolecular 127, 139, 152, 162, 198, 232, 234
 - on moving charge 400, 469

- magnetic 415
 moment of 65*d*
 normal 35
 restoring 85
 superposition 35
 tensile 35, 138
- force constant of spring** 44, 86*d*
- forced convection** 191, 193
- forced oscillations** 93, 454
- forces**
 combinations of 18, 65
 systems of 67
- Fortin barometer** 12, 73
- Foucault** 290
- Fourier's Law** 202
- Franck-Hertz experiment** 514
- Franz and Wiedemann Law** 239
- Fraunhofer diffraction** 304, 309
- Fraunhofer lines** 269, 273
- free expansion** 216
- free fall** 38
- free oscillation** 82, 91, 456
- free path, mean** 201*d*
- free surface energy** 151*d*
- free-body diagram** 39, 65
- freedom, degrees of** 210
- frequency,**
 angular 85, 441
 audible range 325
 beat 115, 327
 comparison of 90, 327, 332, 475
 cut-off 506
 measurement of 331
 natural 93, 456
 of an a.c. 441
 of s.h.m. 84
 of a wave motion 97*d*, 101, 119,
 284, 324
 of resonance 94, 452
 threshold 505
- Fresnel biprism** 296
- Fresnel diffraction,**
 defined 304
 examples 306
- friction,**
 as a damping force 91
 coefficient of 36*d*
 tidal 62
- fringe systems, classification of** 294
- fringes,**
 contour 300
 diffraction 307
 interference 112, 294
 of constant inclination 299
 of constant thickness 300
 white light 297
- full radiation** 244
- full-wave rectification** 457
- fundamental frequency** 330, 332
- fundamental interval** 181
- fundamental mode of vibration** 330
- fundamental quantity** 4
- fusion,**
 microscopic description 130, 225
 nuclear 529
 specific latent heat of 225*d*
- G**
- gain,**
 closed-loop 496
 open-loop voltage 495
- galvanometer,**
 adaptation as ammeter or
 voltmeter 394
 ballistic 420, 421
 dead-beat 93
 moving-coil 418
 moving-magnet 418
 sensitivity,
 charge 420*d*
 current 419*d*
 thermocouple 443
- gamma rays,**
 absorption 519
 as annihilation radiation 511
 detection 285, 519
 diffraction 524
 energy 524
 ionizing effect of 519
 pair production by 510
- gas, electron** 236, 378
- gas constant, the universal molar**
 197*d*, 208
- gas thermometer** 181, 184, 186
- gases,**
 and vapours 227, 229
 atomicity 211
 conduction of electricity through
 373, 515
 equations of state for,
 ideal 197, 200, 208, 234
 real 234
 ideal 181, 197*d*, 198*d*, 207, 231
 internal energy 207, 233
 kinetic theory 198
 laws 197
 liquefaction 235
 molar heat capacity 207*d*
 real 231
 speed of sound in 148, 327
 thermodynamics 204
- gates, logic** 487
- Gaussian surface** 345
- Gauss's Law** 170, 345, 411
 and Coulomb's Law 345
 applications of 171, 345
 for dielectrics 367
 for gravitation 170
- Geiger and Marsden** 511
- Geiger-Müller tube** 519
- generators,**
 a.c. and d.c. 428, 433
 Van de Graaff 355
- geothermal gradient** 536
- germanium** 377, 405
- Germer and Davisson** 288
- glass as a disordered solid** 129
- glass transition temperature** 134
- glow discharge** 373
- gold, freezing point** 183, 249
- gold-leaf electroscope** 356
- gradient,**
 of a graph 25
 potential 351*d*
 temperature 237*d*
 velocity 162*d*
- Graham's Law of Diffusion** 200
- grain boundaries** 132
- grains** 132
- graphite** 529
- graphs, 17, 25**
 errors 17
 gradient 25
 intercept 25
- grating, diffraction** 117, 304
 resolving power 306
 spacing 117, 304
 spectrum 305
- gravitation, Newton's Law of** 169,
 411
- gravitational constant G ,** 169*d*
 constancy of 169
 measurement 176
- gravitational effect of spherical
 shell** 171
- gravitational field** 3, 169, 352, 411
- gravitational field flux** 170*d*
- gravitational field line** 170
- gravitational field potential** 173*d*
- gravitational field strength** 170*d*
- gravitational force unit** 38*d*
- gravitational interaction** 3, 35, 127
- gravitational mass** 12, 37, 88, 169
- gravitational potential energy** 44,
 173
- gravity, acceleration caused by,**
 determination 38, 57, 83
 effect of Earth's rotation on 12, 38
 inside the Earth 173, 176
 standard 38, 73*d*
 variation with altitude 172
 variation with latitude 38
- gravity, centre of** 69
- gray** 524*d*
- greenhouse** 243
- grid as brightness control** 473
- ground state** 514, 524
- grounded emitter** 484
- growth and decay curves** 392, 437,
 527
- guard-ring, for**
 capacitor 357, 360
 heat conduction 241
- gyration, radius of** 57*d*
- gyroscope** 62
- H**
- half-cycle average** 23, 442

half-life period 527*d*
half-wave rectification 442, 457
Hall coefficient 404*d*
Hall effect 336, 403
Hall p.d. 404
Hall probe 405
hard X-rays 507
hardness 143
Harker and Chappuls 185
harmonic 330
harmonic motion,
 combinations of 90, 475
 damped 91
 simple 82, 139
hazards, radiation 524
head of liquid 73
heat, 189*d*
 atomic 194
 latent 225
 radiant 243
 specific latent 225*d*, 226*t*
heat capacity, 189*d*
 molar 190*t*, 194*d*, 207*d*, 211
 principal 194, 207
 specific 189*d*, 190*t*
heat engine 210, 221
heat flow, and probability 217
heat loss in calorimetry 190, 192, 247
heat transfer coefficient 238
heating effect of a current 378, 385
Hebb's method 332
Heisenberg, indeterminacy principle
 4
helical spring 86, 139, 141
helicopter 51
helium,
 liquefaction 235
 relation to α -particle 523
 speed of sound through 182
 vapour pressure 182
Helmholtz coil system 413
 use in magnetic deflection 471
Henning's method 226
henry, the 411, 437*d*
Hertz 514
hertz, the 84*d*, 538
Heuse 208
Heyl 177
high pressures, gas behaviour at
 197, 232
hole, positive 379, 381
bolow conductor 337, 346, 352
hologram 286
holography 286
Hooke's Law, 139, 142
 as cause of s.h.m. 87
 microscopic cause of 139
horizontal component of Earth's
 field 429
hot wire ammeter 418
human ear, audible range of 324
humidity, effect on speed of sound
 328

Huygens' construction,
 diffraction, 106, 303
 diffraction grating 304
 dispersion 107
 reflection 107
 refraction 107, 108
 shock wave 123
 total internal reflection 108
hydrofracturing 536
hydrogen,
 Bohr theory of atom 515
 bomb 132
 energy levels in atom 376, 516
 spectrum of 515
 variation in molar heat capacity
 213
hydrometer 73
hydrostatics 71
hyperbolic fringe locus 113, 294, 297
hyperbolic orbit 175
hypothesis 4
hysteresis,
 magnetic 461
 display of curve 462
 energy loss by 431, 461
 mechanical 141

I

ice-pail, Faraday's 347
ice-point 182*d*
ideal gas,
 difference between principal
 molar heat capacities 208
 equation of adiabatic change 209
 equation of state for 197, 200, 208,
 234
 macroscopic definition 182, 197
 microscopic definition 198
 scale of temperature 181, 200
 thermometer 185
image, 256
 Boys' 263
 defects of 267
 formed by a wave 100, 101
 lens formation of 259
 mirror formation of 266
 real 101, 256
 virtual 100, 256
impact of waves on obstacles 118,
 303, 308, 316
impedance 444*d*
impedance matching 432
impulse,
 angular 61*d*, 420
 linear 48*d*
impurity,
 acceptor 377, 381
 donor 377, 380
incidence, angle of 252
incompressible flow 74
independent variable 25
index, refractive 101, 253*d*, 258, 326
indicator diagram 205, 231
indirectly heated cathode 473
induced charge 346, 353, 359, 428
induced current 424, 430
induced e.m.f. 424, 427, 428, 433, 436
inductance,
 and switches 440
 energy storage by 439
 mutual 439*d*
 self- 436*d*
 solenoid 437
 toroid 437
induction,
 charging by 346, 356
 magnetic 400
induction effect (intermolecular
 forces) 132
inductive circuits 437, 446
inductor, 436
 Earth 429
 energy of 438
 growth and decay of current in
 437
inelastic collision 52, 198, 378, 515
inertia, moment of 57*d*
inertial mass 12, 37*d*, 88, 169
infra-red,
 detection 243, 285
 properties 243, 285
 range 243, 284
infrasonic 323
instantaneous acceleration 29*d*
instantaneous angular velocity 31*d*
instantaneous linear velocity 28*d*
instantaneous power 47*d*, 385, 442,
 447, 448, 451
instruments, 418
 moving-coil 418
 moving-iron 418
 pivoted magnet 418
insulated gate f.e.t. 482
insulation strength 366*d*; *t*
insulators,
 electric 238, 336, 373, 374*t*, 376
 thermal 238*t*
integrals, table of 32
integrated circuits 500
integrator 498
intensity,
 electric field 339*d*
 magnetic field 465*d*
 of interference fringes 113
 of a wave 324
intensive quantity 219
interaction 2, 3*t*
 electromagnetic 3, 127, 339, 400,
 410
 gravitational 3, 35, 127
intercept on graph 25, 302
interference,
 conditions for 112, 292
 in thin films 298

of polarized light 292, 318
 two-source 112, 294, 296
interferometer, *Pohl's* 297
intermolecular attraction energy ϵ
 128
intermolecular forces 127, 139, 152,
 162, 198, 232, 234
intermolecular potential energy 127,
 187, 233
internal energy 45, 189*d*, 205, 207,
 385
internal reflection, 108, 254
internal resistance 386, 389, 396
**International Practical Temperature
 Scale (I.P.T.S.)** 11, 182, 186, 228,
 249
International System of Units (S.I.)
 9, 538
intrinsic semiconductor 377, 379
inverse square law,
 electrostatics 336, 411
 gamma rays 521
 gravitation 169, 411
 wave motion 105, 324, 521
inversion temperature 233*d*, 235
inverting amplifier 496
ionic binding 130
ionic crystals 133
ionization current 520
ionization potential 517
ionizing radiations 126, 285, 507, 519
ionosphere 285
ionospheric waves 100
ion-pair 519, 521
iron, magnetic properties of 431, 461
iron losses 431, 461
irradiance 246*d*
irreversible process 205, 214, 224
irrotational flow 74
isentropic change 218
isobar 509*d*
isobaric process 205, 206, 207
isochronous motion 83
isolated sphere, capacitance of 358
isothermal, critical 231, 235
isothermal bulk modulus 140*d*, 148
isothermal change 206, 208
isothermals,
 carbon dioxide 232
 ideal gas 232
van der Waals gas 235
isotopes,
 discovery of 509
 radioactive 529
isotropic 97, 319
isovolumetric 206, 207

J

Jaeger's method 159
Johnson noise 135
Joly's steam calorimeter 208
joule, the 42*d*, 538

Joule heating 385
Joule-Kelvin effect 233, 235
joulemeter 423
Joule's experiments,
 gas intermolecular p.c. 233
junction diode,
 characteristics 478
 construction 476
 uses 456
junction gate f.e.t. 481
junction transistor (b.j.t.) 483
Jupiter's satellites 175

K

Kater's pendulum 63
kelvin, realization of 11
kelvin, the 9*d*, 11, 182, 184
Kelvin temperature scale 181
**Kelvin's statement of the second law
 of thermodynamics** 210
Kepler's Laws 168, 174
Kerr electro-optic effect 319
kilogram, realization of 11
kilogram as standard mass 9*d*, 11
kilogram force 38*d*
kilogram weight 38*d*
kilowatt hour 386*d*
kinematics 28
kinetic energy, 44*d*
 change during collision 52
 rotational 59, 210
 translational 210
kinetic theory, see p. 137 for detailed
 references
Kirchoff's Law for radiation 246
Kirchoff's rules for electric circuits
 4, 389
krypton 9, 11
Kundt's dust tube 212, 333

L

lag in a.c. circuits 441, 446, 448*t*
lagging, thermal 237
laminar flow 74, 161, 164
laminations 430
Laplace (speed of sound) 327
laser 283, 293
 uses 286
latent heat,
 measurement 226
 microscopic explanation 128, 225
 specific 225*d*
lateral heat loss 237, 241
lattice,
 ion spacing in 133, 311, 508
 metal 378
law,
 nature of 3
 (see also subject concerned)
Lawton and Pimpton 337
lead in a.c. circuits 441, 448*t*, 449
leakage current 456, 477

least confusion, circle of 267
least distance of distinct vision 275
Le Chatelier's principle 152, 426
Lees' disc 242
left-hand motor rule (Fleming) 401,
 425
length, standard of 9*d*
lens,
 achromatic 268, 270
 action of 101, 259
 definitions for 259
 measurement of refractive index
 263
 power of 262*d*
 sign convention for 260
lens aberrations 267
lens blooming 299
lens combinations 262, 268
lens equation 259
lens-maker's formula 260
Lenz's Law 426, 428, 436
lift, aerodynamic 77
light,
 particle aspects of 286, 287, 504
 spectrum 284, 285
 speed of 289, 361, 416, 539
 wave aspects of 288
light-dependent resistor (LDR) 490
light-emitting diode (LED) 479
lightning 355
limit of audibility 324
limit of experimental error 16
limit of proportionality 142
line spectra,
 visible 269, 273, 515
 X-ray 508
linear absorption coefficient 521
linear charge density 348
linear expansion, thermal 187
linear expansivity 187*d*
lines,
 electric field 340
 fluid flow 74, 161
 gravitational field 170, 411
 magnetic field 400, 426
liquefaction of gases 235
liquid crystal display (LCD) 479
liquid phase,
 behaviour of molecules in 128, 129
 conduction in 236, 241, 373
 pressure exerted by 71*d*, 72
 surface effects of 151
Lissajous figures 90, 475
litre 198
Lloyd's mirror 296
load line 486
location of fringes 294
logarithm, natural 22
logarithmic decrement 92*d*
logic gates 487
logic level 488
London effect 132
longitudinal wave, 96

- graphical representation 98, 323
reflection 99, 111, 329
- long-range forces** 128
- long-range order** 129
- loop equation** 390, 392, 431, 434, 437, 449
- Lorentz's conductivity relationship** 239
- Lorentz's rotating disc** 394, 432
- lost volts** 389
- loudness of sound** 324
- loudspeaker** 333
- luminous intensity** 10
- Lummer and Pringsheim** 248
- Lyman series** 515
- M**
- Mach cone** 122
- Mach number** 123*d*
- McLeod gauge** 73
- macroscopic** 2, 139, 187, 197, 218
- Madelung constant** 133
- Magnadur** 461
- magnetic circuit** 464
- magnetic damping** 91, 430
- magnetic deflection of particles** 400, 469, 522
- magnetic dipole**, 407
electromagnetic moment of 407*d*, 409, 422, 465
torque on 407
- magnetic domain** 463
- magnetic field**, 400, 410
energy of 439
in electromagnetic wave 21, 282
line 400, 426
of Earth 401, 429
- magnetic field strength H** 465*d*
- magnetic flux** 401
- magnetic flux density B** , 401*d*, 410
measurement 405, 415, 428, 459
set up by circular coil 412
Helmholtz coil 413, 471
solenoid 413
straight wire 413
toroid 413, 437, 459
- magnetic force**,
on moving charge 400, 469, 522
on straight wire 9, 406, 415
- magnetic induction** 400
- magnetic materials** 461
- magnetic moment** 407
- magnetic pole** 408
- magnetic reluctance** 464*d*
- magnetic susceptibility** 460*d*
- magnetization** 465*d*
- magnetization curve** 461
- magnetizing force H** 465*d*
- magnetomotive force** 464
- magnetostriction** 464
- magnification**,
angular 275*d*
longitudinal 262*d*
transverse 262*d*, 266
- magnifying glass** 276
- magnifying power**, 275*d*
magnifying glass 276
microscope 278
telescope 278
- majority carriers** 381
- malleability** 143
- Maltese Cross experiment** 469
- Malus's experiment** 315
- Malus's Law** 319
- Malus's Theorem** 106
- manometric capsule** 331
- mark-space ratio** 452
- Marsden** 511
- mass**,
and energy 290, 509, 512, 529
centre of 68*d*
conservation of 74
contrasted with weight 37
gravitational 12, 37, 88, 169
inertial 12, 37*d*, 88, 169
measurement of 12, 37
molar 9, 196, 328
orders of magnitude 13
relative molecular 196*d*
relativistic 468, 510
rest 468, 510
standard of 9
- mass number** 509*d*, 525
- mass spectrograph** 513
- mathematical relationships** 21
- matter, phases of** 129
- maximum permitted dose level** 524
- maximum power theorem** 386
- Maxwellian distribution** 198
- Maxwell's cork-screw rule** 411, 427
- Maxwell's electromagnetic radiation theory** 2, 282, 289, 361
- Mayer's equation** 208
- mean free path** 201*d*
- mean value of mathematical function** 23
- measurement** 11
- mechanical energy** 43, 45
- mechanics**,
classical 2, 34
Newtonian 2, 34
quantum 2, 514, 515
relativistic 2, 34, 468
statistical 2, 216
- medium for wave propagation** 96, 120, 122, 282, 322
- Melde's experiment** 331
- melting** 130, 225
- meniscus, shape of** 153
- mercury thermometer** 181, 186
- metal rectifier** 375, 456
- metallic bonding** 131
- metallic solid solution** 133
- metals**,
electrical conduction 373, 378, 383, 405
electrons in 131, 505
thermal conduction 236
- method of dimensions** 19
- metre, realization of** 11
- metre, the** 9*d*
- metre bridge** 398
- Michelson** 290
- micrometer** 11
- microphone** 333
- microprocessor** 502
- microscope** 278
- microwaves** 284, 285, 289, 314, 316
- Millikan**,
determination of e 471
determination of the *Planck* constant 505
- minimum deviation** 108, 257, 271
- minority carriers** 381
- mirrors**,
aberrations 267
equation 266
parabolic 267
sign convention 266
- mobility, electric** 374*d*, 379, 380
- model** 4, 513
- moderator** 529
- modes of vibration**,
air columns 329
stretched strings 331
- modulus of elasticity** 139*d*
- molar** 10
- molar heat capacity** 1905, 194*d*, 207*d*, 211
- molar volume** 198
- mole, realization of** 11
- mole, the** 9*d*
- molecular attraction energy ϵ** 128
- molecular complex** 232
- molecular crystals** 133
- molecular mass, relative** 196*d*
- molecular velocities** 198
- molecule**, 127*d*
size of 136, 203
- molecules, forces between** 127, 139, 152, 162, 198
- moment**,
electric dipole 342*d*
electromagnetic 407*d*, 408, 465
- moment of a couple** 67*d*, 406
- moment of a force** 56, 65*d*
- moment of inertia** 57*d*
calculation 57
determination 63
- moments, principle of** 66
- momentum**,
angular 61*d*
conservation of angular 61
conservation of linear 50, 510
force and rate of change of 50
linear 34*d*, 48, 510

photon 287
relativistic 510
momentum current density 203*d*
monostable multivibrator 493
monatomic molecule 211
monochromatic light 294, 301
Moon,
acceleration towards Earth 168
effect on tides 176
orbital angular momentum 62
moons, Jupiter's 175
Moseley 511
motion, see particular type (e.g. s.h.m.)
motor,
a.c. 434
d.c. 434
methods of winding 434
motor effect 400, 406, 425, 428
moving charge 372, 400, 469
moving-coil galvanometer,
as ammeter 394
as voltmeter 395
construction 418
use as ballistic galvanometer 420, 421
moving-iron galvanometer 418
multimeter 399
multivibrator,
astable 492
bistable 493
monostable 493
multiple slit interference 117, 304
musical note 325
mutual conductance 482*d*
mutual inductance 439*d*
mutual induction 431, 439
mutually exclusive events 215

N

natural frequency 93, 456
natural logarithm 22
n-channel f.e.t. 481
near point 275, 276
negative charge 336, 373, 403, 469
Nernst's metal block calorimeter 191
Neumann's Equation 426, 439
neutral axis 149
neutral equilibrium 69
neutral temperature 392
neutral surface 149
neutrino 523, 524
neutron, discovery of 526
neutron source 526
newton, the 36*d*, 538
Newtonian fluids 163, 203
Newtonian mechanics 2, 34
Newton's corpuscular theory 107, 290
Newton's experimental law of impact 52
Newton's formula (lenses) 263

Newton's Law of cooling 193
Newton's Law of gravitation 169, 411
Newton's Laws of motion 34
Newton's rings 294, 301
Nicol prism 317
nodal lines 112, 294
node 111
nodes,
distance between adjacent 112
in air columns 330, 333
on strings 331
using light 289
noise,
acoustical 325
electrical (*Johnson*) 135
non-conservative force 46, 427
non-inverting amplifier 497
non-Newtonian fluids 163
non-viscous fluids 74, 161
normal adjustment 276
normal contact force 35
note, musical 325
n-p-n transistor 483
n-type semiconductor 380, 381, 403
nuclear charge 337, 511, 512
nuclear dimensions 126, 337, 511
nuclear disintegration 518, 526
nuclear fission 528
nuclear force 3, 337, 512
nuclear fusion 529
nuclear reaction 526
nuclear stability 512
nuclear transmutation 526
nucleon 509, 512
nucleus,
binding energy 510, 512
size 126, 337, 511
nuclide 509
number,
atomic 509*d*, 526
mass 509*d*, 526

O

objective lens 277, 278, 279, 280
object, virtual 261, 262
ohm, the 374*d*, 538
ohmic resistor 375
Ohm's Law,
experimental test 375, 397
statement 375
oil film colours 299
oil film experiment 136
open air columns 329, 330
open-loop voltage gain 495
operational amplifier 495
optic centre 259
optical activity 318
optical lever 188, 419
optical path 293, 298, 300
optical pumping 283

optics,

geometrical 252
physical 282
orbit 168, 174, 516
parking 175
orbital electrons 514
orbitals 515
order of interference 305
orderly flow 74, 161, 164
orders of magnitude 13*f*
ordinary ray 317
ordinate 26
organ pipe 329
orientation effect,
dielectric materials 366
intermolecular forces 131
paramagnetic materials 463
oscillation, centre of 63
oscillations,
damped 91, 93, 454
electrical 454
forced 93, 454
free 82, 91, 456
simple harmonic 82, 139
torsional 87, 89, 148, 409
oscillator 499
oscillator circuit 453, 486
oscilloscope 473
Otto cycle 223
output characteristic 484
overtone 330
oxide-coated cathode 473

P

pair annihilation 511
pair production 511
parabolic mirror 268
parallel,
capacitors in 362
resistors in 384
resonant circuit 451, 452
parallel axes theorem 59
parallelogram of vectors 18
parallel-plate capacitor,
air filled 359
measurement of ϵ_0 361
use in *Millikan's* experiment 472
with dielectric 365, 367
parallel-side air films 294, 298
paramagnetism 459, 463
paraxial ray 259, 266
parking orbit 175
partial pressures, Dalton's Law of 200
particle current density 201*d*
particle-wave duality 287, 504
partition 215
pascal, the 71*d*, 538
Pascal's principle 72
Paschen series 516
Pauli exclusion principle 3, 376

- p-channel f.e.t.** 481, 483
peak values (a.c.) 428, 442
Peltier effect 388, 391
pendulum,
ballistic 12, 52
Barton's 93
compound 62, 88
Kater's 63
simple 57, 87
torsion 87, 89, 148
penetration, by
radioactive emissions 519, 423
X-rays 507
percentage error 16
perfect gas,
macroscopic definition 182, 197
microscopic definition 198
periodic motion 82
periodic table 509, 511
permanent gas 232
permanent magnetism 461
permanent set 142
permeability,
absolute 460*d*
relative 460*d*, 464
permeability constant 410*d*, 417
permittivity,
absolute 366*d*
relative 365*d*, 366*t*
permittivity constant 337*d*, 361
perpendicular axes theorem 59
Perrin 469
phase angle 93, 104, 113, 294, 449
phase change on reflection 99, 111, 293, 329
phonons 236
phosphor 474
phosphorescence 271, 474
photocell 506
photodiode 479
photoelasticity 318
photoelectric effect, 503
laws 504
quantum explanation 504
photoelectrons 468, 503
photomultiplier tube 520
photon,
energy 286, 504, 524
momentum 287
wave behaviour 287
photon-electron interaction 510
photosphere 273
pinch-off voltage 481, 482
pitch, viscosity of 163
pitch of a note 325
Pitot static tube 76
Planck constant 287, 504, 508, 539
Planck's quantum theory 248, 504
plane of polarization 102, 314, 316
plane of vibration 102, 314
plane-polarized light,
description 283, 314
detection 318
production 316
uses 318
planetary motion, *Kepler's Laws of*
168, 174
plasma 129
plastic flow 143
platinum resistance thermometer
185, 186
Plimpton and Lawton 337
p-n-p transistor 483
Pohl's interferometer 297
point charge 336, 340, 341, 353
points, electrical discharge 355
Poiseuille's equation 165, 167
Poisson's ratio 147
polar molecule 131, 366
polarimeter 318
polarization,
circular 102
elliptical 102
plane (linear) 102, 314
polarization of dielectric material
132, 366
polarizer 316
polarizing angle 315*d*, 318
polaroid 316
pole,
magnetic 408
of a mirror 266
pollution 534
polycrystalline 132
polygon of forces 66
polymers,
amorphous 134
crystalline 134
porous-plug experiment 233
positive charge 336, 372, 403
positive hole 379, 381, 403
positive rays 470, 508
positron 338, 510, 523
potential,
and field strength 350
contact 391, 476
divider 384
due to a point charge 353
due to a sphere 353
electric 349*d*
gravitational 173*d*
inside a conductor 352, 354
ionization 517
stopping 505
potential barrier at n-p junction
477
potential difference,
electric 350*d*, 373*d*
comparison 396
measurement 395, 396, 397
related to electric field 350
potential energy, 44*d*
elastic 44, 146, 420
electric 349, 363
gravitational 44, 173
intermolecular 127, 187, 233
potential gradient 351*d*
potentiometer 395
power,
electric 385, 423
factor 452*d*
in a.c. circuits 442, 447, 448, 451
magnifying 275*d*
mechanical 47*d*
of a lens 262*d*
of a wave 104
transmission of 386
Poynting vector 21
prefixes 10
pressure, 71*d*
at a point in a liquid 72
atmospheric 73
differences across a liquid
interface 155
effect on the speed of sound 328
exerted by,
ideal gas 197, 199
liquid 72
real gas 232, 234
solid 72
radiation 287
standard 73*d*, 182, 228
variation caused by sound wave
322
pressure law 197
Prévost's theory of exchanges 246
primary cell 394
primary quantity 4
principal axis 259, 266
principal focus 259
principal molar heat capacities of a gas 207*d*, 211
principal specific heat capacities 194
principle 4
principle of moments 66
principle of superposition 35, 99, 109, 292, 303, 337
principle of virtual work 157
prism,
achromatic 270
measurement of angle 271
minimum deviation produced by
108, 257, 271
Nicol 317
small-angled 258
probability,
concepts of 215
disorder and 215
heat flow and 217
of radioactive decay 526
process,
adiabatic 209, 327
irreversible 205, 214, 224
isobaric 205, 206, 207
isothermal 206, 208
isovolumetric 206, 207
reversible 204, 214, 388
progressive waves, 96
comparison with stationary 111

projectiles 30
 proof-plane 355
 propagation of wave motion 96, 282, 322
 proportionality, limit of 142
 proton 126, 471, 509, 516
 pseudovector 19, 55, 66, 71, 401
 p-type semiconductor 381, 403
 pulsatance 85, 441
 pulse code modulation 501
 pulse production 491
 pupil of the eye 280, 311
 pure spectrum 270
 pyrometer, 186
 disappearing filament 249
 total radiation 248

Q

quadratic equations 21
 quality factor (*Q*-factor) 95*d*, 455
 quality of sound 325
 quanta, distribution of 217
 quantity, physical 4
 quantization,
 in Bohr theory 516
 of electric charge 337, 472
 of energy in electromagnetic radiation 286, 504
 quantum 504
 quantum mechanics 2, 514, 515
 quantum number 288, 515
 quantum theory 248, 504
 quartz fibre suspension 177
 Quincke's tube 332
Q-value 95, 455

R

R, the universal molar gas constant 197*d*, 208
 radar 100
 Doppler effect 121
 nature of 100, 285, 289
 radial heat flow 240
 radial magnetic field 406, 419
 radian 22*d*
 radiant exitance 245*d*
 radiation,
 black-body 244
 dose 524
 exposure 524
 hazards 524
 measurement of temperature by 181, 182, 186, 248, 249
 pressure 287
 pyrometer 186, 248, 249
 radioactive 519
 spectrum of 273, 284, 285, 305
 thermal 243
 radio waves 284, 285, 314
 radioactive dating 530
 radioactive decay constant 527*d*

radioactive equilibrium 528
 radioactive series 518
 radioactivity 518
 radioisotopes 529
 radiomicrometer, *Boys'* 243
 radiotherapy 529
 radium 518, 526
 radius of curvature of lens surface 263
 radius of gyration 57*d*
 ramp voltage 499
 random error 13
 random process,
 Brownian movement 135
 radioactive disintegration 526
 range of forces 128, 232
 range of radioactive decay products 519, 523
 rarefaction 323
 rate of reaction 128
 ratemeter 519
 ratio of principal molar heat capacities 209*d*, 211, 212, 327
 rationalization 337, 348, 411
 Rayleigh's criterion 310
 rays 97, 252, 283
 ordinary and extraordinary 317
 reactance,
 capacitive 448*d*, 451
 inductive 446*d*, 451
 reaction,
 chain 529
 nuclear 526
 rate of 128
 reaction energy 528
 reaction cross section 528
 reactor
 breeder 529
 nuclear 529
 thermal 529
 real and apparent depth 254
 real image 101, 256
 rectification of a.c. 455
 rectifier 375, 456
 rectifier bridge circuit 420, 443
 rectilinear motion 29
 rectilinear propagation of light 252, 303
 red shift 121
 reduced mass 210
 reflectance 245*d*
 reflecting galvanometer 419
 reflecting telescope 268, 279
 reflection,
 and Doppler effect 121, 289
 as cause of stationary waves 111, 289, 329, 331
 at curved surfaces 266
 at plane surfaces 99, 252
 Huygens' construction for 107
 laws of 107, 252
 phase change on 99, 111, 293, 329
 polarization by 315

 total internal 108, 254
 reflexion factor 245*d*
 refracting telescope 268, 278, 280
 refraction,
 at curved surfaces 101, 259
 at plane surfaces 100, 107, 252
 double 317
 Huygens' construction for 107
 laws of 101, 326
 of rays 252
 of waves 100, 252
 Snell's Law 101, 253, 326
 through prisms 108, 257, 258, 270
 through successive media 253
 through thin lenses 259
 refractive index,
 absolute and relative 101, 253*d*, 326
 and speed of light 108, 253
 by apparent depth 255
 by critical angle 256
 by spectrometer 271
 of material of a lens 263
 of material of a prism 255, 271
 refrigerator 222
 regenerative cooling 235
 Regnault 208
 reinforced metals 147
 reinforced plastics 147
 relative aperture 267, 268, 310
 relative biological effectiveness 524
 relative density 71
 relative permeability 460*d*, 464
 relative permittivity 365*d*, 366*t*
 relative refractive index 101, 253*d*, 326
 relative velocity 29*d*
 relativistic k.e. 510
 relativistic mass 468, 510
 relativistic mechanics 2, 34, 468
 relativistic momentum 510
 relay 461, 506
 reluctance, magnetic 464*d*
 remanence 461
 reservoir capacitor 456
 resilience 143
 resistance, 373*d*
 absolute measurement 432, 433
 and Ohm's Law 375
 internal 386, 389, 396
 measurement of 398
 microscopic cause 378
 small, determination of 397
 temperature coefficient of 382*d*, 383
 thermal 238
 thermometer 185, 186
 resistance offered by fluid to motion 166
 resistances,
 comparison of 396, 397, 398
 in series and parallel 383
 resistivity, 374*d*, 374*t*

- temperature coefficient of 382*d*, 383
- resistor**,
as converter of electrical energy 378, 385
- resolution of vectors** 18
- resolving power**,
of eye 311
of grating 306
of telescope 280
Rayleigh's criterion 310
- resonance**,
amplitude 94, 452
energy 94, 452
in parallel circuit 452
in series circuit 452
of air columns 329
of engineering structures
of strings 331
sharpness of 93
- resonance tube** 333
- rest mass** 469, 510
- restitution**, coefficient of 52*d*
- restoring force** 85
- resultant of vector combination** 18
- retentivity** 461
- retina**, image formation on 276
- reverberation** 326
- reversible pendulum** 63
- reversible process** 204, 214, 388
- Reynolds's number** 164*d*
- right-hand rules**,
Fleming's dynamo 426
grip 411
Maxwell's corkscrew 411, 427
- right-handed axes** 282
- rigid body** 55
- rigidity modulus**, 139, 141*d*, 149
comparison with viscosity 163
- ripple tank** 100, 102, 112, 116, 119, 326
- rise**, capillary 157, 159
- rocket** 50, 51
- Roemer** 290
- Röntgen** 507
- root mean square** 23
- root-mean-square current** 442
- root-mean-square speed** 199
- rotating coil**, e.m.f. in 428, 434
- rotation**,
of Earth (and g) 12, 38
of a rigid body 55
of plane of polarization 318
- rotational dynamics** 55
- rotor** 435
- Routh's rules** 4, 59
- Rowland**,
anchor ring 459
diffraction grating 304
experiment to detect B 410
- Royds** 523
- rule** 4
- Russell rules** 525
- Rutherford**,
and *Royds* 523
nuclear atom 511
- Rydberg constant** 515, 516, 539
- S**
- Satellites** 175
- geosynchronous 176
- saturated vapour** 227
- saturation of magnetic material** 461
- saturation vapour pressure**,
explanation 227
measurement 228
- Saturn's rings** 121
- Savart (and Biot) Law** 410
- saw-toothed wave form** 474
- scalar product of vectors** 19
- scalar quantity** 17
- scale of temperature** 180
- scaler** 519
- scattering**,
 α -particles 511
light 118, 217, 317
polarization and 317
wave motion 118
- scintillation counter** 520
- screw gauge** 11
- search coil** 415, 428, 459
- Searle**, conductivity measurement 241
- second**, realization of 11
- second**, the 9*d*
- second law of thermodynamics** 210, 219
- secondary circuit** 431, 439
- secondary emission of electrons** 468
- secondary quantity** 4
- secondary wavelets** 106, 116, 303
- Seebeck effect** 185, 380, 388, 391
- self-diffusion** 201
- self-inductance** 436*d*
- semiconductor** 373, 377, 379
extrinsic 380, 381
intrinsic 377, 379
- sensitivity**, of
ear 324
galvanometer 319*d*
potentiometer 397
- series**,
capacitors in 363
cells in 389
circuit (a.c.) 449
resistors in 383
resonant circuit 452
- series-wound dynamo** 433
- series-wound motor** 434
- shear**,
angle of 141, 163
modulus 139, 141*d*, 149, 163
strain 139, 141
stress 139, 141
- shearing**, rate of 163
- shearing force** 149
- shell, electron** 126, 515
- shielding**,
absence of gravitational 177
electric 346
- shock wave** 122
- shunt galvanometer** 394
- shunt-wound dynamo** 433
- shunt-wound motor** 434
- S.I. units** 9, 538
- siemens** 374, 538
- sign convention**,
real-is-positive 260, 266
vector product 19, 31
- silicon** 380, 383, 483
- simple harmonic motion**, 83*d*
and *Hooke's* Law 83, 139
angular 83*d*
approximate 87
combination of 90
damped 91
dynamics of 85
energy approach to 89
energy in 88
exact 86
kinematics of 83
- simple pendulum** 57, 87
- simultaneous events** 215
- sine law** 22
- single slit diffraction** 307, 309
- sinusoidal wave**,
equation for 103
no distortion in 103
- skater's angular momentum** 62
- skidding and toppling** 40
- skin friction** 166
- sky, colour of** 118, 217, 317
- slip rings** 433
- slope of a curve** 25
- small signal forward current**
transfer ratio 485*d*
- small-angled prisms** 258
- smoothing capacitor** 458
- Snell's Law** 101, 253, 326
and *Huygens'* construction 107
- soap bubble** 155
- soap film**, draining 300
- Soddy rules** 525
- sodium**,
absorption by 247
spectral lines 269, 273
- soft iron** 431, 461
- solar cell** 534
- solar constant** 246*d*
- solar furnace** 534
- solar panel** 534
- solar system** 168, 175
- solar year** 12
- solenoid**,
magnetic field 413
self-inductance 437
- solid phase**

- behaviour of molecules in 129, 140, 141
 electrical conduction 373, 378
 molar heat capacity 194
 strength 142
 thermal conduction 236
 thermal expansion 187
- solid solution**, metallic 133
- Sommerfeld* 407
- sonometer** 332
- sound**,
Doppler effect in 119, 123, 326
 intensity of 324
 loudness of 325
 no dispersion of 102, 326
 recording 333
 reproduction 333
 spectrum 324
 speed in gases 327
 wave nature of 97, 322
 wave properties of 325
- source (transistor)** 481, 482
- space charge** 473
- spark counter** 520
- specific** 189
- specific charge of α -particle** 523
- specific charge of electron** 468, 470
- specific heat capacity**, 189*d*
 measurement 191
 gases 189, 207
 solids 191
 variation with temperature 193
- specific volume** 232
- spectral series** 515
- spectrograph** 271, 513
- spectrometer** 270, 271
 X-ray 313
- spectroscope** 271
 direct vision 270
- spectrum**,
 absorption 247, 271, 274
 α -particle 521
 atomic 271, 515
 band 273
 β -particle 522
 black-body 244, 248
 continuous 274
 electromagnetic 284
 emission 271, 515
 γ -ray 524
 line 269, 273, 508, 515
 mass 513
 pure 270
 solar 273
 sound 324
 X-ray 508
- speed**,
 average 28*d*, 199
 instantaneous 28*d*
 most probable 198
 of light 289, 361, 416, 539
 of sound 327
 root-mean-square 23, 199
- speeds, Maxwellian distribution of** 198
- sphere, capacitance of** 358
- spherical aberration** 267
- spherical charge distribution** 347, 353
- spherical lens** 259
- spherical mass distribution** 172
- spherical mirrors** 266, 268
- spherical wave front** 97, 105
- spin** 61, 463
- spinhariscope** 520
- split-ring commutator** 433
- spring**,
 and *Hooke's* Law 139
 and s.h.m. 86
 energy stored by 44
- spring balance** 12, 37
- spring constant** 44, 86*d*
- stability of equilibrium** 69
- standard boiling point** 228*d*
- standard cell** 394
- standard gravity** 38, 73*d*
- standard pressure** 73*d*, 182, 228
- standard temperature** 198
- standards, electrical** 394
- standards of measurement** 9
- standing waves**, 99, 110
 compared with progressive waves 111
 equation for 112
 in light 289
 in sound 327, 329
- starting resistance (motor)** 435
- state** 215
- state, equation of**,
 ideal gas 197, 200, 208, 234
 real gas 234
- stationary waves** (see standing waves)
- statistical descriptions** 215
- statistical equilibrium** 215
- statistical mechanics** 2, 215
- statistical nature of radioactive decay** 527^e
- statistical processes** 216
- stator** 435
- steady-flow calorimetry** 192, 208, 226
- steady state** 192, 226, 237
- steam point** 182*d*
- Stefan-Boltzmann constant** 247*d*, 539
- Stefan's Law** 247
- stiffness** 142
- Stokes's Law** 165, 166, 167
 use in *Millikan's* experiment 472
- stopped pipes** 329, 330
- stopping down** (optical instrument) 267, 279
- stopping potential** 505
- straight wire**, magnetic field of 413
- strain** 138*d*, 139, 140, 141
 energy 146
 types of 139
- streamline flow** 74, 161, 164
- strength of solids** 142
- stress** 138*d*, 139, 140, 141
 effect on polarized light 318
 types of 139
- stress-multipliers** 144
- stress-strain graphs**,
 glass 144
 polythene 144
 rubber 144
 single metallic crystal 143
- stretched strings** 103, 110, 331
- stretching of a wire** 141, 142
- string, vibrating** 331
- stroboscope** 331
- strong (nuclear) interaction** 3
- strontium-90** 519
- structure of matter** 127
- sublimation** 130, 146, 225
- subsonic** 323
- substitution method for resistance** 398
- subtractor** 498
- sugar** 318
- Sun**,
 chromosphere 273
 photosphere 273
 source of energy 529
 spectrum 273, 274
 temperature of surface 246, 248
- superconductivity** 382
- supercooling** 158, 235
- superficial expansion** 188
- superheating** 158, 235
- superposition, principle of** 35, 99, 109, 292, 303, 337
- superposition of forces** 35
- supersaturation** 158, 235
- supersonic boom** 123
- surface charge density** 348*d*, 355
- surface energy** 151*d*, 153
- surface tension**, 151*d*
 effect on vapour pressure 157
 measurement of 159
 molecular explanation of 152
 variation with temperature 151, 152, 160
- susceptibility, magnetic** 460*d*
- suspended-coil galvanometer** 418
- suspension constant** 45, 148, 177
- sweep of electron beam** 474
- switch, transistor** 487
- switches and inductance** 440
- symbol, use of** 6
- synchronization** 474
- synclastic surface** 156
- systematic error** 13, 302
- Système Internationale d'Unités* 9, 538

T

Tacoma Narrows bridge 95

- target (X-ray tube)** 507
telescope,
Cassegrain 279
Coudé 279
 diffraction in 268, 310
Galilean 279
Newtonian 279
 normal adjustment 276
 reflecting 268, 279
 refracting 268, 278, 280
 resolving power of 280
 terrestrial 278
telephoto lens 265
telephoto ratio 265
temperature 180*d*
 absolute zero of 181
Boyle 233, 235
Celsius interval 184*d*
 critical 231, 235
 inversion 233, 235
 kinetic interpretation of 199
 measurement of 186
 neutral 392
 standard 198
 statistical interpretation of 217
temperature coefficient of resistivity
 382*d*, 383
temperature gradient 237*d*, 328
temperature interval 184
temperature scales 180
 thermodynamic 222
tensile strain 138, 139*d*, 142
tensile strength 142
tensile stress 138, 139*d*, 142
tension,
 state of 35, 152
 surface 151*d*
tension in a rotating hoop 41
terminal velocity 165, 166, 471
tesla 401*d*, 538
theorem of parallel axes 59
theory, nature of 4
thermal (heat) capacity 189*d*
thermal conductance 238
thermal conduction 217, 236
thermal conductivity 237*d*, 238*t*
 measurement 241
 mechanisms 236
thermal diffusion 200, 201, 217
thermal efficiency 222*d*
thermal equilibrium 180, 204
thermal expansion 187
thermal insulation 238*t*
thermal resistance 238
thermal resistance coefficient 238
thermal runaway 479
thermion 468
thermionic emission 473
thermistor 382
thermocouple 181, 185, 391, 397
thermodynamic equilibrium 204
thermodynamic temperature scale
 181, 222
thermodynamics,
 first law 194, 205, 210, 219
 second law 210, 219
 zeroth law 180
thermoelectricity 388, 391
thermometers,
 compa. ed 184, 186
 in practice 184
thermometric property 180, 181
thermonuclear reaction 529
thermopile 243
thermoplastics 134
thermosetting 134
thickness of an oil film 136, 299
thin film interference,
 parallel-sided 298, 299
 wedge 300, 301
thin lens equation 259
thin lenses in contact 262
thin prism 258
Thomson's canal rays 508
Thomson's method for specific charge 470
Thomson's model of the atom 514
thorium series 518
three forces at a point 66
three-phase system 458
threshold frequency 505
threshold of audibility 324
tical 461
tides, theory of 176
timbre 325
time,
 measurement 12
 standard of 9, 11
time base 474
time constant,
 capacitive 393*d*
 inductive 438*d*
time period 83*d*, 441
tone 325
toroid,
B inside 413, 437, 459
 self-inductance of 437
 use for studying magnetic materials 459
torque, of a
 current loop 407
 motor 434
 twisted wire 148
torsion oscillations 87, 89, 148, 409
torsion wire 44, 148, 177
torsional pendulum 87, 89, 148
total internal reflection,
 rays 254
 waves 108
toughness 143
tourmaline 316
Townsend discharge 373
tracer elements 529
track, banking of 40
trajectory equation 30, 342
transfer characteristic 485
transfer of thermal energy 217, 236, 243
transformation, radioactive 526
transformer 431
transistor, 480
 action 481, 482, 484
 as current amplifier 485
 as oscillator 486
 bipolar 480, 483
 characteristics 481, 483, 484
 construction 480, 482, 483
 field effect (f.e.t.) 481
 junction 483
 unipolar 480, 482
 uses 480
transition region 477
translational motion 55, 200
transmission of electrical energy 386
transmutation 526
transport phenomena 201
transverse waves 96, 99, 103, 282, 314
travelling wave 96, 111
triangle of forces 66
triggered bistables 494
triggering,
 negative-edge 494
 positive-edge 494
triple point 181, 230*d*, 235
truth table 488
tube, resonance 333
tube of flow 74
tungsten filament 473
tuning fork 327
tuning of electrical circuit 94, 453
tunnel through the Earth 176
turbulent flow 74, 76, 163, 167
twisting of spring segments 141
Tycho Brahe 168
Tyndall effect 317

U

- U-value** 238
ultimate tensile stress 142
ultrasonic waves 100, 323
ultraviolet light 284, 285, 503
uncertainty principle 4
undamped s.h.m. 82
uniform circular motion 32, 39
uniform temperature enclosure 244
unipolar transistor (f.e.t.) 480, 482
unit cell (crystal) 129, 133, 312
units,
 derived 10
 fundamental 9
 realization of 11
 S.I. 9, 538
universal gravitation, law of 169
universal molar gas constant 197*d*, 208
Universe 121
unsaturated vapour 227, 231
unstable equilibrium 69

- upper fixed point** 181
- uranium**,
 dating by 530
 fission 528
 series 518
- U-tube**, liquid oscillations in 87
- V**
- vacuum**,
 permeability of 410
 permittivity of 337
- valence band** 376, 381
- valence electrons** 126, 380
- valency** 126
- Van de Graaff** 355
- van der Waals**,
 equation of state for real gases 234
 residual forces 131
- vaporization** 227
- vapour pressure**,
 dependence on surface curvature 157
 measurement 228
 saturation 227
- vector quantity** 17
- vectors**,
 addition and subtraction 18
 multiplication 18
 resolution 18
- velocity**,
 angular 31*d*
 average 28*d*, 199
 critical 163, 165
 drift 374, 378, 404
 escape 174
 instantaneous 28*d*
 mean 28*d*, 199
 most probable 198*d*
 of electromagnetic waves 289, 361, 416, 539
 of sound 327
 orbital 174
 relative 29*d*
 root-mean-square 199
 terminal 165, 166, 471
 wave 97
- velocity gradient** 162*d*
- velocity-time graphs** 29, 85
- Venturi meter** 76
- vernier** 12
- vertical component of Earth's field** 429
- vibration**, plane of 102, 314
- vibrations**,
 damped 91, 92, 454
 electrical 454
 forced 93
 free 82, 91, 456
 of air columns 329
 of strings 331
 simple harmonic 83*d*, 139
 torsional 87, 89, 148, 409
- virtual image** 100, 256
- virtual object** 261, 262
- virtual work**, principle of 157
- viscosity**,
 coefficient of 141, 163*d*, 203
 gases and liquids compared 161
 microscopic description 161
- viscous flow** 161
- viscous resistance to motion** 165
- visible light** 252, 285
- vision**, least distance of distinct 275
- visual angle** 275*d*
- volt** 349*d*, 388, 538
- voltage**,
 pinch-off 481, 482
 ramp 499
- voltage amplifier** 496
- voltage comparator** 497
- voltage-follower circuit** 497
- voltage sensitivity** 419*d*
- voltmeter** 388
- voltmeter**,
 a.c. 418, 443
 calibration 396, 399
 d.c. 395, 418
 moving-coil 394, 420
 valve 418
- volume**,
 critical 232
 molar 198
- volume flux** 75*d*
- von Laue** 311, 508
- vulcanization** 134
- W**
- Walton** 526
- waste disposal** 525
- water**,
 hydrogen bonding in 132
 triple point of 181
 waves on 97, 102
- watt** 47*d*, 538
- wattless current** 452
- wattmeter** 423, 452
- wave equation** 103
- wave function** 514
- wave mechanics** 2, 514, 515
- wave motion** 96, 282, 322
- wave number** 104*d*
- waveform** 98, 325
- wavefront** 97, 282
- waveguide** 283
- wavelength** 97*d*
- wavelets** 106, 116, 303
- wave-particle duality** 287, 504
- waves**,
 classification 98, 282, 322
de Broglie 288, 312
 electromagnetic 282
 frequency 97*d*, 115, 119, 284, 324
 ionospheric 100
 longitudinal 96, 111, 323, 329
 power 104
 progressive (travelling) 96, 111
- properties 99
 sinusoidal 103
 sound 97, 322
 standing (stationary) 99, 110, 112, 289, 327, 329
 transverse 96, 99, 103, 282, 314
 velocity 97*d*, 289, 327
- wavetrain** 98
- weak interaction** 3
- weber** 401*d*, 538
- wedge fringes** 300, 301
- weight** 12, 37, 170
- Weston cadmium cell** 394
- wetting of surfaces** 155
- Wheatstone bridge** 376, 397
- whiskers** 143
- white light fringes** 294, 297, 307
- Wiedemann and Franz Law** 239
- Wiener's experiment** 289
- Wien's displacement law** 248
- Wilson cloud chamber** 158, 520
- wind**, effect on velocity of sound 120, 328
- wire, stretching of** 142, 146
- work**, principle of virtual 157
- work done**,
 by force 42*d*, 44, 174, 351
 by ideal gas in reversible process 205, 208, 210
 by torque 59, 342, 408
 in stretching a wire 146
- work function energy** 473*d*, 473*t*, 505
- work function potential** 473*d*, 473*t*
- work-energy theorem** 44, 60, 75
- work-hardening** 145
- X**
- X-ray crystallography** 135, 312, 508, 511, 524
- X-ray diffraction** 135, 311
- X-ray spectra** 508, 511
- X-ray tube (Coolidge)** 507
- X-rays**,
 absorption 507
 ionizing effect 126, 507
 production 285, 507, 511
 properties 284, 285, 507
- Y**
- yield point** 142
- Young's experiment** 112, 294
- Young's modulus** 139*d*, 141
- Z**
- Zener diode** 479
- Zener effect** 479
- zero**, absolute (temperature) 181
- zero of potential** 46, 173, 350
- zero point energy** 181
- zeroth law of thermodynamics** 180
- zinc**, f. p. of 183
- zones of a lens** 267

Since publication, *Essential Principles of Physics* has been an invaluable reference work for both A level and post A level students of physics.

This reset new edition has been completely revised to bring it in line with syllabus changes. It has

- new material in 24 chapters, much dealing with engineering applications and properties of materials
- a new section on *Energy Resources*, and
- a rewritten and greatly expanded chapter on *Electronics* to bring it completely up to date with current developments.

JOHN MURRAY



ISBN 0-7195-4566-8



9 780719 545665

07-DQT-492