Physics

CONCEPTS & MODELS

Wenham Dorling Snell Taylor

GATESHEAD COLLEGE
10019088

PHYSICS

concepts and models

GATESHEAD COLLEGE
530
WEN
10019088
KNOWLEDGE CENTRE

E. J. Wenham is Head of the Physical Science Department, Worcester College of Education. He has for many years been closely involved in curriculum development of physics at school and college level both in the UK and in many other parts of the world. He has acted as consultant for UNESCO, OECD, and the British Council. From 1962–6, Mr Wenham was Associate Organizer for the Nuffield O-Level Physics Project and he is presently concerned with the revision of the same project.

G. W. Dorling is Senior Lecturer in the Physical Science Department, Worcester College of Education. He has been closely involved in in-service work for teachers of physics both in the UK and abroad. He has worked on syllabus revisions of O-level physical science and for the ONC, and was also a member of the consultative committee of the Nuffield Combined Science Project. He is an examiner in A-level physics for one of the examination boards.

J. A. N. Snell is Senior Physics Master at St Edward's School, Oxford. Previously he has taught at Eastbourne College and Downside School. He is particularly interested in project work in applied physics and technology at school level and has developed programmed courses in practical electronics for school use.

B. Taylor is in the Physics Department at Rugby School. He has been Science Adviser to Leicestershire Education Authority and was Senior Physics Master at St John's College, Southsea. He has been particularly concerned with the design of apparatus and laboratories. Mr Taylor has run many in-service courses for teachers both in the UK and abroad. He is on the subject committee for A-level physics for one of the examination boards.

PHYSICS

concepts and models

Advisory board:

Dr A. D. C. Grassie University of Sussex

Professor E. F. W. Seymour University of Warwick

Dr A. J. Walton The Open University

E. J. WENHAM Worcester College of Education

G. W. DORLING Worcester College of Education

J. A. N. SNELL St. Edward's School, Oxford

B. TAYLOR Rugby School



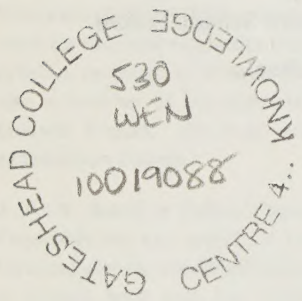
ADDISON-WESLEY PUBLISHERS LIMITED
London Reading, Massachusetts
Menlo Park, California Don Mills, Ontario

GATESHEAD
TECHNICAL
COLLEGE
LIBRARY



B 18710

WORLD STUDENT SERIES EDITION



Copyright © 1972 by Addison-Wesley Publishers Limited.
Philippines copyright 1972 by Addison-Wesley Publishers Limited.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior written permission of the publisher. Printed in Great Britain.

PREFACE

The primary purpose of this book is to provide pre-university and college students with a readable, coherent course of study in the fundamental science of physics. Such a course must reflect the nature of the subject itself, presenting physics both as a process and as a structure. The process is one of concept development and model building; the structure is provided by a closely inter-connected fabric of ideas. This is reflected in the "unit" approach which has been adopted, and in the concern to show that physics is involved in the solution of open-ended problems about the physical world.

It is assumed that students using the text are familiar to some extent with the ideas and content of such courses as the Nuffield O-level Physics course and that elementary mathematics is studied concurrently. Nevertheless, all important concepts are developed from first principles, reliance being placed not so much on a thorough understanding of specific material as on a sound appreciation of the methods and ideas with which physics is concerned.

It is assumed, too, that a student using this book would wish to understand the concepts so thoroughly that he will be able to apply them effectively. Some open-ended problems for solution are therefore offered at the end of each unit. As with problems in physics itself, some of these have no clear-cut solutions but provide ample opportunity for discussion and speculation.

Teachers will observe that the units and chapters of the book may be arranged in several different ways and so provide a variety of routes through the material. Even so, it is hoped that the introductory unit, which sets the scene for the others, will always be studied first. The final unit is a case-study of the development of man's ideas about the atom and its nucleus, and mirrors the first unit in its concern for the process of physics. Teachers will also observe that the second unit treats dynamics from the "momentum first" approach whilst fields are treated together in Unit 7. After a relatively brief treatment of basic thermal physics in Unit 3, Unit 4 provides an introduction to a model for matter and to the science of materials. A generalized treatment of the phenomena of vibrations and waves in Unit 5 is followed by a unit which is concerned with the basic phenomena of electricity with special reference to the circuit and to various circuit elements. The relationship between electricity and matter is considered in Unit 8.

Units of the Système International are used throughout with, where it is expedient, the addition of certain decimal fractions and multiples of those units which have special names (e.g., the kilowatt-hour) and of some units which are defined in terms of the best available experimental values of certain physical constants (e.g., the electron-volt and the unified atomic mass unit). Since it is our hope that the readers of this book will wish to refer to the original papers, at least in certain areas, some familiarity with these additional units is helpful.

Naturally any such text as this has been strongly influenced by the series of distinguished books on physics published in the last decade. Special mention must be made of the publications of the Nuffield Foundation Science Teaching Project, the Physical Science Study Committee and the Harvard Project Physics. And it would be wrong not to recognize the debt owed to such authors as Arons, Bennet, Rogers, Sears and Zemansky, Warren and, of course, Feynman.

The authors owe a special debt of gratitude to Professor E. F. W. Seymour (University of Warwick), Dr A. D. C. Grassie (University of Sussex) and Dr A. J. Walton (the Open University) who together constituted an advisory panel. They read the manuscript in successive drafts and their comments and suggestions were of the utmost value. The comments of Miss R. M. Grant also contributed much to the final manuscript. Nevertheless the authors must accept full responsibility for any errors whether of fact or interpretation which may still exist in their work.

Our thanks are also due to Mrs Sheila Pargeter and Mrs Doris Wenham for many hours of typing, to Miss B. Jones for help with the preparation of the index and to the members of our families for their infinite patience.

January 1973

E. J. Wenham
G. W. Dorling
J. A. N. Snell
B. Taylor

CONTENTS

UNIT ONE: MODELS

1 Introduction

1.1 Robert Boyle 2
1.2 Torricelli's experiment 3
1.3 Laws 5
1.4 Theories and models. 5
1.5 The continuation of the work of Boyle . . 8
1.6 The discovery of Boyle's law 9

2 A Model for a Gas

2.1 Pressure and temperature 12
2.2 Assumptions within the model 12

UNIT TWO: MOTION, FORCE, AND ENERGY

3 Introduction

3.1 Change and rate of change 18
3.2 Vector and scalar quantities 19
3.3 The addition of vectors 20
3.4 Comparison of the two methods 21
3.5 Vectors and navigation 22
3.6 Signs and vectors 22
3.7 Resolution of vectors 23
3.8 Summary 23

4 The Investigation of Interactions: Some Experimental Techniques

4.1 The reduction of friction in the laboratory . 24
4.2 Methods for measuring speed 26
4.3 An investigation of interactions 28
4.4 The rule begins to emerge 31
4.5 Extending the rule 31
4.6 The new rule 32
4.7 The concept of mass 33
4.8 A conservation rule 33

5 Momentum and Mass

5.1 The law of conservation of momentum . . 34
5.2 Momentum as a vector 34
5.3 Isolated systems 34
5.4 Making measurements of mass 35
5.5 The unit of mass 36
5.6 Inertial mass and gravitational mass . . . 36
5.7 The cause of momentum change 36

6 Force

6.1 Acceleration 39
6.2 The unit of force. 40
6.3 Newton's third law of motion 41

7 The Effects of Forces: Work and Energy

7.1 Work 44
7.2 The unit of work. 45
7.3 Distance 46
7.4 Power 46
7.5 Energy. 46
7.6 Calculations of energy transfer (work done) when a mass is lifted in a gravitational field . 48

7.7 The calculation of the energy transferred when the shape of an elastic body is changed . . 49
7.8 The calculation of the energy transferred when the speed of a mass is changed 50
7.9 The general case 51
7.10 The quantitative examination of some energy transfers 52
7.11 The conservation of energy 54
7.12 Conservative and dissipative forces . . . 54
7.13 The energy stored in a rotating body . . . 54
7.14 Energy "book-keeping" during collisions. The classification of collisions 56
7.15 The kinetic energy remaining after an inelastic collision 57
7.16 The transfer of kinetic energy resulting from elastic collisions 57
7.17 The condition for maximum energy transfer in an elastic collision 59
7.18 A scientific toy 60

8 Frames of Reference

8.1 The conservation of energy in different frames of reference 62
8.2 Interactions in two dimensions 62
8.3 Inertial frames 62
8.4 Motion along a curved path 64
8.5 Motion in a circle 65
8.6 Circular motion and frames of reference . . 66
8.7 Spaceflight and weightlessness 67

UNIT THREE: THERMAL PHYSICS

9 Temperature

10 Temperature, Heat, and Work

11 Heat Capacities

12 A Second Law?

12.1 The second law 86
12.2 The Carnot cycle 87
12.3 The efficiency of a Carnot engine 88
12.4 The thermodynamic temperature scale . . 89
12.5 The thermodynamic scale and the ideal gas scale 90
12.6 The place of thermodynamics 90

UNIT FOUR: A MODEL FOR MATTER

13 A Model for a Gas: A Second Look

13.1 A simple mathematical analysis of the model 94
13.2 An equation of state 97
13.3 Some implications of the model 98
13.4 Molecular speeds 103
13.5 Temperature and kinetic theory 104
13.6 Avogadro's number 105
13.7 The atomic model 106

14 A Model for a Solid

14.1 A possible relationship between a force and the separation of the particles 109
14.2 The physical properties of solids 110
14.3 Further investigation of a model for the solid state 114

15 Crystalline Structure

15.1 Three-dimensional close packing 117
15.2 Orderly arrangements 118
15.3 Crystal structure and the properties of solids . 119
15.4 Elastic deformation of an ordered structure . 119
15.5 Solids as ordered structures 120
15.6 Interference patterns and crystals 121
15.7 Investigating the structure of real solids . . 124

16 Applying the Model for Matter

16.1 Surface energy and surface tension . . . 129
16.2 Plastic deformation 131
16.3 The thermal and electrical properties of metals 134
16.4 Conclusion 136

UNIT FIVE: VIBRATIONS AND WAVES

17 Periodic Motion

17.1 Introduction 144
17.2 Describing periodic motion 144
17.3 Isochronous oscillation 146
17.4 Sinusoidal motion 147
17.5 Phase 149

18 The Motion of an Oscillator

18.1 Velocity and acceleration of a sinusoidal oscillator 152

18.2 Computing the motion of a simple oscillator 155
18.3 The equation of motion of a simple harmonic oscillator 158
18.4 The simple pendulum 159
18.5 The energy of a simple harmonic oscillator . 163

19 Damped Oscillations

19.1 Damped oscillations 166

20 Forced Oscillations

20.1 Forced oscillations 173
20.2 Resonance 176

21 Coupled Oscillators and the Transmission of Energy

21.1 Coupled oscillators 183
21.2 A line of coupled oscillators: propagation of a longitudinal vibration 186
21.3 Wave machines, and the mathematical description of travelling sinusoidal waves. . . 189
21.4 A string of particles: propagation of a transverse vibration 191

22 Waves on Water

22.1 Waves on water 194

23 Wave Behaviour

23.1 Reflection of waves 199
23.2 Stationary waves 202
23.3 Wave phenomena in two dimensions . . . 205
23.4 Particles, waves, and models 213
23.5 Some interference phenomena 213
23.6 Coherence. 218

UNIT SIX: ELECTRICITY

24 Introduction

25 Steady Direct Currents

25.1 The simple circuit 235
25.2 Potential difference 238
25.3 Resistance. 241
25.4 The resistance of wires 244

26 Power, Energy, and Electromotive Force

26.1 Power 248
26.2 Electromotive force 249
26.3 Energy changes 251
26.4 Matching source and load 253
26.5 Kirchhoff's second law 255
26.6 Solving circuit problems 257

27 Some Useful Circuits

27.1 The potential divider 261
27.2 The slide wire potentiometer 263
27.3 Bridge circuits 265
27.4 Combining resistances 266

28 Electric Charge

28.1 Electric charge 268
28.2 Capacitors 271
28.3 Discharge of a capacitor 273
28.4 Dielectrics and capacitors 275

29 Using Circuit Elements

29.1 Some circuit elements 278
29.2 Circuit elements in combination 281

30 Electromagnetism

30.1 Electromagnetic forces 288
30.2 Electric motors 295

31 Electromagnetic Induction

31.1 Electromagnetic induction 301
31.2 The disc dynamo. 305
31.3 Induction without movement 305

32 Alternating Currents

32.1 Alternating current 313
32.2 Inductors in a.c. circuits 318
32.3 Capacitors in a.c. circuits 321
32.4 The transformer 328
32.5 Motors and alternating currents 331

UNIT SEVEN: FORCES, FIELDS, AND RADIATION

33 Fields of Force

33.1 Introduction 340
33.2 Fields of force 341
33.3 The gravitational field 342
33.4 The electrostatic field 343

33.5 The magnetic field 346
33.6 Field strength directions 346
33.7 Mapping other fields 349
33.8 Lines of force 349
33.9 Faraday and lines of force 350
33.10 The magnetic field and moving charges . . 351
33.11 A unit for current 352
33.12 The behaviour of bodies under the influence of a field of force 353

34 The Inverse Square Law

34.1 Variations in field strength 355
34.2 Flux and the inverse square law 356

35 The Electric Field

35.1 The electrostatic force between two charged particles 358
35.2 A model for the electric field 360
35.3 A consequence of the flux model 361
35.4 Using the flux model 362

36 The Gravitational Field

36.1 The force of gravity 364
36.2 Weightlessness 365
36.3 Newton's test of the inverse square law . . 366
36.4 The measurement of the gravitational constant (G) 367
36.5 Comparison of electric and gravitational fields 367

37 The Magnetic Field

37.1 Measuring magnetic fields 369
37.2 Exploration of magnetic fields 370
37.3 The field at the centre of a flat coil . . . 371
37.4 The magnetic field of a current-carrying solenoid 372
37.5 The field outside a long solenoid 374
37.6 The flux model and magnetic fields . . . 374
37.7 The value of the magnetic force constant . . 374

38 Potential Energy in Fields of Force

38.1 Introduction 376
38.2 Gravitational potentials in the field around the Earth 377
38.3 Potential changes and potential gradients . . 379
38.4 Potential at a point in the Earth's gravitational field—the general case 380
38.5 Self-energy 381
38.6 A potential well 381
38.7 The special case of circular motion . . . 382

39 Potential in the Electric Field

39.1 Definitions and measures 385
39.2 A further look at field potentials 388
39.3 Energy stored in an electric field 389
39.4 Electric potential due to a dipole 390

40 Electromagnetic Induction

40.1 Induction in a magnetic field 392
40.2 The electromotive force induced by a changing magnetic flux 394
40.3 The direction of the induced e.m.f. . . . 396
40.4 The energy stored in a magnetic field . . . 396
40.5 The self-inductance of a solenoid 398
40.6 An alternative form of Faraday's law of electromagnetic induction 398

41 Maxwell's Model of the Electromagnetic Field

41.1 Introduction 401
41.2 Maxwell's equations 402
41.3 Loops of magnetic flux 402
41.4 A search for a fourth equation of the electromagnetic field 404
41.5 The directions of the induced electric and magnetic fields 404
41.6 Electromagnetic oscillations and electromagnetic waves 407
41.7 The special theory of relativity 414

UNIT EIGHT: ELECTRICITY AND MATTER

42 Electric Currents

42.1 Introduction 426
42.2 Evidence for an ionic model 427
42.3 Summary 431

43 The Electron

43.1 The electronic hypothesis 433
43.2 Measuring the charge 434
43.3 Millikan's experiment 435
43.4 The conduction of electricity through gases at low pressure 437
43.5 The nature of the cathode rays 439
43.6 Some applications of thermionic emission . 442

44 A Model for an Electric Current

44.1 The Drude theory of electric current in metals 445
44.2 Thermal conductivity 447
44.3 The force on a current-carrying wire in a magnetic field 449
44.4 The Hall effect 449

45 Ionization and Energy Levels

45.1 The experiment of Franck and Hertz . . . 453
45.2 An application of the Franck-Hertz tube . . 455

46 Electricity, Matter, and Light

46.1 Photoelectricity 457
46.2 Energy levels 458
46.3 Summary 462
46.4 The Sun's spectrum 463
46.5 Absorption of ultraviolet radiation in the atmosphere 463
46.6 The stimulated emission of radiation . . . 463
46.7 The helium-neon laser 464

UNIT NINE: ATOMIC STRUCTURE—A CASE HISTORY

47 The Atomic Model

47.1 Introduction 472
47.2 The early atomists 472
47.3 The atomic theory 475
47.4 The genesis of quantum ideas 477
47.5 X-rays 478
47.6 The electron—the first chip 480
47.7 A crisis develops—the photoelectric effect . 482
47.8 Another predicament for the wave theory . 482
47.9 Planck's quantum theory 483
47.10 Albert Einstein 483
47.11 Waves? Particles? 484

48 The Nuclear Atom

48.1 The discovery of radioactivity 485
48.2 Subsequent developments 486
48.3 The structure of the atom 487
48.4 Scattering 488
48.5 Other models of the atom 489
48.6 Ernest Rutherford 489
48.7 The scattering experiments of Geiger and Marsden 490
48.8 The nuclear atom emerges 491
48.9 Angles of scattering 493
48.10 Further testing of the nuclear atom model . 493
48.11 Atomic number 494
48.12 Models of the atom 495

49 The Nucleus

49.1 The composition of the nucleus 497
49.2 The artificial transmutation of an element . 500
49.3 Natural transmutation 502
49.4 The search for a neutral particle 503
49.5 The discovery of the neutron 504
49.6 The nuclear structure of elements 505
49.7 Alpha-particle emission 506
49.8 Isotopes 506
49.9 Beta-particle emission 507
49.10 Gamma radiation 508
49.11 The behaviour and detection of neutrons . . 509
49.12 Modern alchemy—a dream come true . . 510
49.13 Epilogue 510
49.14 Notes on the cloud chamber photographs . 510

Appendices

Appendix 1 522
Appendix 2 523
Appendix 3 524
Appendix 4 525
Appendix 5 526
Answers to selected problems 529
Glossary of symbols 531
General bibliography 533
Index 535

UNIT ONE
Models

1 Introduction
2 A Model for a Gas

Chapter 1
INTRODUCTION

1.1 Robert Boyle
1.2 Torricelli's experiment
1.3 Laws
1.4 Theories and models
1.5 The continuation of the work of Boyle
1.6 The discovery of Boyle's law

Many of the well-known laws of physics are associated with the names of scientists. Any study of gases soon introduces the student to the laws of Boyle, Charles, and perhaps of Gay-Lussac. Planetary astronomy leads to the laws of Kepler at an early stage in its development. Much elementary optics is based on Snell's law. Electricity demonstrates its importance with any number of laws—due to Coulomb, Ohm, Kirchhoff, Ampère, Faraday, Lenz. It is an easy task to add to the list.

Some of these laws are quite remarkable for the degree of perception shown by their discoverers. Who but Kepler could have deduced his three laws of planetary motion from the mass of data supplied by Tycho Brahe? Even in our age of computers this still ranks as a magnificent achievement.

But, when one has marvelled at the skill of these scientists in reducing a mass of observed behaviour into a succinct law of behaviour; when one has admired their implicit belief in a physical world whose behaviour can be expressed in the regularities described by these laws, one may still be surprised at the remarkable dullness of so many of these laws. Can a relationship between electric current and potential difference have been the culmination and the driving force of the life of Georg Ohm? Can a relationship between the pressure and the volume of a gas sample have been the goal of Robert Boyle? Physics is, at times, presented as if it were.

1.1 ROBERT BOYLE

Robert Boyle (1627–91), a contemporary of such men as Christopher Wren and Samuel Pepys, wrote an account of many of his experiments in a book which he called *New Experiments, Physico-Mechanicall, Touching the Spring of the Air and its Effects.* Written in the style of a letter to his nephew it begins:

> "Receiving in your last letter from Paris a desire that I would add some more experiments to those I formerly sent you over; I could not be so much of your servant as I am, without looking upon that desire as a command; and consequently, without thinking myself obliged to consider by what sort of experiments it might the most

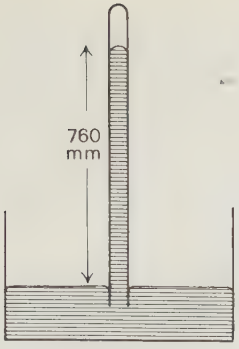


Fig. 1.1

acceptably be obeyed. And at the same time, perceiving by letters from some other ingenious persons at Paris, that several of the Virtuosi there were very intent upon the examination of the interest of the air; in hindering the descent of the quicksilver in the famous experiment touching a vacuum; I thought I could not comply with your desires in a more fit and seasonable manner, than by prosecuting and endeavouring to promote that noble experiment of Torricellius; and by presenting your Lordship an account of my attempts to illustrate a subject, about which its being so much discoursed of where you are, together with your inbred curiosity and love of experimental learning, made me suppose you sufficiently inquisitive."

A comprehensive account of Boyle's experiments which he recorded in this book is to be found in the *Harvard Case Histories in Experimental Science* from which this quotation is taken. Reading the account of Boyle's work on the air, it soon becomes clear that he was motivated not merely by a desire to satisfy the whims of his young nephew. The work was stimulated by the previous experiments of Torricelli to which Boyle refers.

1.2 TORRICELLI'S EXPERIMENT

Only a few years before Boyle's experiment with air, Torricelli, a pupil of Galileo, had shown that if a long tube, closed at one end, was filled with mercury and inverted with the open end immersed in a bowl of mercury, the column of mercury in the tube did not exceed about 760 mm in height (see Fig. 1.1). The space which appeared above the mercury appeared to be "empty."

The phenomenon could not be denied. But Torricelli's explanation of why the mercury should behave in this way was at the centre of controversy during the middle of the seventeenth century. To quote Boyle, Torricelli believed that:

> "the falling of the quicksilver in the tube to a determinate height, proceedeth from the aequilibrium, wherein it is at that height with the external air, the one gravitating, the other pressing down with equal force upon the adjacent mercury."

This view was determined by Torricelli's *theory* (or *model*) of an earth surrounded by a "sea of air" which exerts a pressure on us and on our surroundings, living, as we are, at the bottom of this "sea."

What had led Torricelli to perform the experiment with the mercury? No one knows for sure but there is some evidence to suggest that he had evolved his model of the atmosphere *before* doing his famous experiment. The fact that water could not be pumped up higher than about 11 m using a single lift pump was known to Galileo who mentions it as an unsolved problem in his book *The Dialogues concerning Two New Sciences*. Torricelli himself was a keen disciple of Galileo. It seems likely that his theory was proposed as a solution to this problem. His experiment with the mercury thus becomes an attempt to verify his solution by experiment.

Despite the experiment, the theory was not generally accepted. Many of the critics could not understand why, if the theory was acceptable, the mercury in a barometer tube did not fall when the barometer was sealed off from the surrounding air. However, Boyle was firmly convinced of the correctness of Torricelli's theory and set out, in a sequence of famous experiments, to demonstrate that a number of deductions which might be made from the theory could, in fact, be verified by experiment. For example, one of his first experiments was to use a pump,

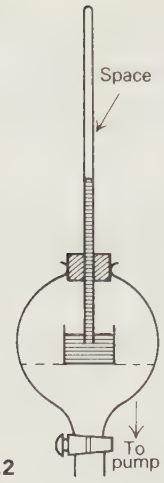


Fig. 1.2

designed by Robert Hooke, to remove the air from the top of the bowl of mercury (Fig. 1.2).

He reasoned that, if Torricelli's model of the atmosphere was satisfactory, the mercury in the tube should fall since there would be less air pressure to support the column. This is a simple experiment to perform today. But Boyle found it so difficult that he felt called upon to give a detailed description of his technique.

So we see that Boyle was led towards his famous law, not by a desire to find how the pressure and volume of a gas varied but by a need to verify a theory or model of the behaviour of the air. His motive was to verify some ideas then current as to *why* things behaved as they did, and not simply to find out *how* they behaved. But, as we shall see, further evidence about the behaviour of the physical world came to light as a result of his investigations.

Since this is not a book about the history of physics, some reason must be given for considering the work of Robert Boyle. Does the law by which Boyle's name is celebrated indicate the sort of questions he asked about the physical world? History strongly suggests that it is not. It suggests that Boyle was not so much interested in answering questions about how the world behaves but in why it behaves as it does. This is typical of the way in which most of science, and thus physics, was evolved. While experiment, which occupies so much of a scientist's time, is concerned with answering "How?", that question is itself posed by theories and models which have been proposed to answer "Why?".

Of course, no one is going to ask "Why?" before there is some clear reason for doing so. What the history of physics seems to show is that the question "Why?" usually arises at a very early stage in the investigation of new phenomena and that attempts to answer it are at once made by proposing theories or models which it is hoped represent the underlying causes. Such theories suggest new experiments which may both serve to test the theory and bring new facts to light. With this in mind, let us re-examine the work of Boyle (Fig. 1.3).

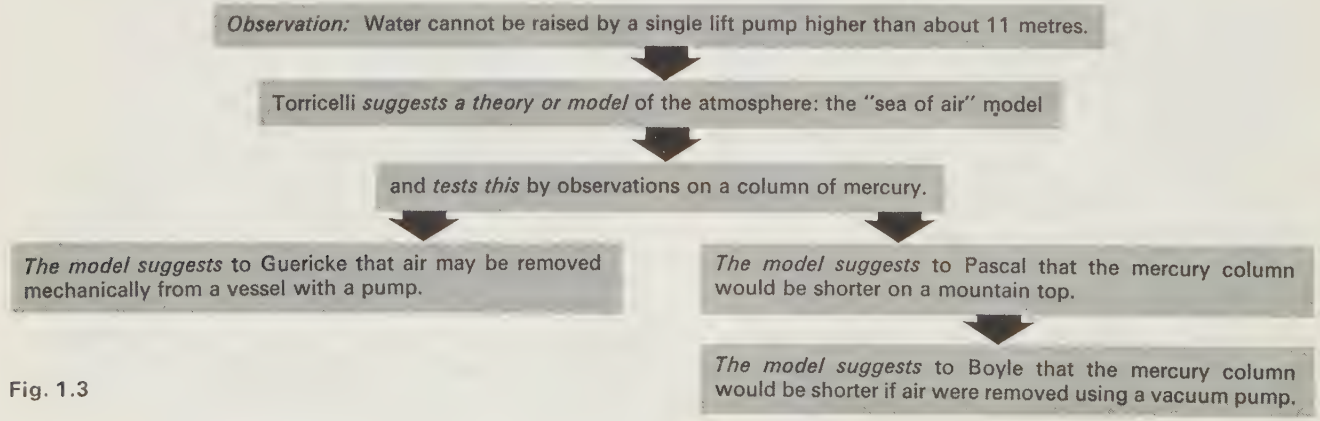


Fig. 1.3

1.3 LAWS

We are now in a position to consider the meaning of this word as it is used by scientists. At its simplest, a *law* is a summary of observed, measurable behaviour. Philosophers have discussed at length the impertinence of scientists who assume that what is true for a necessarily limited number of observations made in the past is true for all similar observations both now and in the future. But, as R. A. R. Tricker points out "So far as practical action is concerned, we are driven to disregard the problem, as in fact, we always do." (Tricker, 1965.)

Kepler's laws of planetary motion offer a good example. They summarize many observations of the positions of the planets relative to each other. Their formulation depends solely on these observations and not on any preconceived theory as to why the planets behave as they do. This is clearly borne out by Kepler's original intention that they should support theoretical speculation of an entirely different sort from the Newtonian theory of gravitation which they now help to confirm.

There are many laws in physics of the same kind. They include:

The relationship between the angle of incidence which a light beam makes with a mirror and its consequent angle of reflection. This may be expressed in the simple law that the angle of incidence is equal to the angle of reflection.

The relationship between the extension of a loaded spring and the load itself (Hooke's law).

The relationship between the activity of a radioactive substance and time. This law is rather more complex than those cited above but, nevertheless, the relationship can be expressed by the simple exponential law.

In all these cases, the law describes a simple mathematical relationship between two measurable quantities. The laws do not depend upon any preconceived notions about causes. They merely summarize the answers to the appropriate "How?" question. (For example, how does the extension of the spring change with load? How does the activity of a radioactive substance change with time?) Such laws may seem dull and lacking in interest, but they imply that the physical world behaves in a coherent fashion. This implication is basic to the sciences.

1.4 THEORIES AND MODELS

If we believe that the physical world behaves in the coherent way suggested by such laws, we are driven to ask "Why?". It is man's attempts to answer this second question which have led to that body of knowledge which we know as science.

Our knowledge of the behaviour of the physical world can never be complete. It follows that our knowledge of the causes of that behaviour can never be certain. The best we can do is to guess (intelligently) some likely cause; to build imaginative pictures of the processes which give rise to this observed behaviour. These imaginative theories or models, as these pictures are called, have been the essential feature of scientific speculation throughout its recorded history. They are imaginative adventures of the mind and they bring the essential ingredient of creativity into science. If science today differs from that of the Ancient Greeks, it is not in the building of models or of theories that that difference is to be found. Much of early Greek speculation is now unacceptable to science in general and to physics in particular. This is not because we no longer speculate, but because we now lay down fairly clear rules about what are, and what are not, acceptable theories and models in the field of science. We shall return to this problem when we attempt to build a model for the structure of matter.

The wave theory of light is an excellent example of a modern theory. This suggests that light energy is transported from place to place in a wave-like manner. It can be shown that light, if regarded as a wave, can be reflected and refracted according to the laws which are found to apply to light. So a wave provides us with a satisfactory model for light—satisfactory in many respects; indeed, it might be preferable to talk of the wave *model* for light rather than of the wave *theory* of light. However, the two words are often used as though they were interchangeable.

It must by now be clear that the word *model* does not mean quite the same thing to a physicist as it does to the man in the street. When we talk of a *model ship* we usually mean a scaled-down copy of the real thing. The model ship is *like* the real ship. When physics took over the word *model*, the essential feature of the everyday usage which was preserved was in the word *like*. Very often, in physics we are dealing with processes which arise from underlying causes which are not directly perceived. This is true of the behaviour of light. No one can *see* a wave of light. The only sort of waves which we can see are those which might be generated in a stretched rope, a "slinky" spring, or on the surface of a sheet of water, whether in a ripple tank or the ocean itself. We can experiment with these waves and examine their characteristic properties. Even more important, we can write down mathematical expressions to describe their behaviour and make verifiable deductions from those expressions. When we talk of light energy travelling in waves, we mean that the energy is travelling *like* waves travelling down a rope or over water. We consider that a wave model is a good model of the way in which light behaves. It is in this sense that the scientist uses the word *model*. Let us consider one or two examples.

a) Newton's theory of gravitation

The theory of universal gravitation was proposed by Newton in an attempt to account for the observed motions of the planets. The essence of the theory is that all bodies attract all other bodies and, in their turn, are attracted by all other bodies because they possess mass. This force of attraction is called the force of gravity. It can account for the observed motions of the planets; for the attraction of all terrestrial bodies to the earth; for the behaviour of comets, of earth satellites, of space ships, and even for the formation of the galaxies. In so far as this force cannot be directly experienced, it becomes an act of speculation to say that the planets move round the sun because a force attracts them towards the sun just as (*like*) a force acts on a mass which is whirled round one's head on an attached string. In this sense, Newton's theory is a model for the behaviour of the planets (or of earth satellites, or whatever is being considered). But there is an important sense in which this model is different from any we have previously considered. In order to incorporate fully the experience summarized in the three laws of Kepler the force between a planet and the sun must vary inversely as the square of the separation, r, between the planet and the sun. The forces on masses which are attached to strings do not behave in this way. Furthermore, the force must be proportional to the product of the mass of the planet, m_p, and the mass of the sun, m_s. In full

$$F \propto \frac{m_p m_s}{r^2}.$$

All this is incorporated in Newton's theory of gravitation. The model is, to a considerable degree, a mathematical model. Certain features of this model can best be expressed in mathematical terms. This is a characteristic of a physical model and, as physics has developed, so its models have become more and more mathematical in nature.

But Newton took matters further than this. If the movement of the planets depends on some force of attraction between the planets and the sun—a force which depends on their mass—then this force ought to be evident between *all* bodies. This is what Newton proposed. He suggested that the force holding the planets in their orbits was the same force as that which attracted all earthly bodies to the earth. It is for this reason that Newton's theory is often referred to as the theory of universal gravitation. He had extended the familiar terrestrial force of gravity out to the edge of the solar system and beyond.

The history of the development of this theory is one which has received very detailed attention from the historians of science. It is often used to illustrate the relationship between observation, theory, and law. By carefully disregarding some of the planetary theories which have not proved successful, it is possible to represent in Newton's work a clear-cut relationship between these three entities. The argument is often propounded as follows: careful observations made by Tycho Brahe were summarized in

laws of planetary motion by Kepler. This led Newton to propose a hypothesis (or tentative theory) to account for these laws. From this theory he made new deductions which, on being justified, led to his hypothesis attaining the status of a theory.

It is difficult to assess the usefulness of such an analysis. It almost certainly does not reflect the historical development of the theory—too much is left unsaid. However, with hindsight, it is often tempting to re-structure scientific developments into such a "logical" order. It is doubtful whether such an ordering is justifiable. Theories and models are rarely propounded on a basis of purely objective fact. They must be seen against the current background of scientific thought.

We have tried to suggest in the previous discussion of the work of Boyle that theory (speculation) and the making of measurements proceed side by side, each acting on the other: a theory suggests some worth-while observations. The results modify the theory. The modified theory, in its turn, suggests new observations.

Certainly, Tycho Brahe and Kepler, influenced still by the scholasticism of the Middle Ages, had their own theories, and these were derived from Greek mathematical speculation. The theories did not prove fruitful, however, and it is a tribute to the work of these men and an object lesson for science, that their experimental work had value because their own theories were not allowed to influence the objectivity of their observations and analyses.

Newton's theory was successful, but not because he followed that pseudo-logical process which some commentators suggest. Newton's tremendous contribution was the insight into what is a fruitful theory. It is to Newton that we owe the view that a theory should be sufficient—sufficient to explain observed facts, and to propose new observations. The ability to know what is sufficient is one of the marks of genius.

If one wanted a story to show how Newton came to propose his theory of universal gravitation, one could do worse than consider the legend of Newton and the apple. Told in its usual form (how Newton sat under a tree and an apple fell to the ground nearby and so he discovered the law of gravity) it has often been laughed at. But think of it again. Newton is sitting in a Cambridge college garden, on a sunny day in late autumn, pondering new mathematical ideas he has just worked out to describe motion under a central force. This is the problem for which his laws of motion have been evolved. He has calculated that Kepler's laws must mean that the planets are acted upon by some force proportional to their masses. All this follows from his new dynamics, but it does not constitute a complete theory. What could that force be? An apple falls to the ground nearby. He stares at it, unheedingly for a moment. And then the flash of inspiration. Of course!

There is no more evidence to support this version of the story than there is to support the more commonly expressed view of Newton's careful and logical analysis of the problem. But as an allegory, it described the process of the evolution of theories (when these have been more fully documented) much better than does the usual picture.

Before leaving this famous theory, let us consider one further aspect of the traditional story. It contains within it a suggestion that a theory is essentially different from a hypothesis erected on the basis of a single successfully established deduction. While the status of the theory is clearly enhanced once a proposed consequence proves successful, one swallow does not make a summer. The status of a theory is further and further enhanced as it is shown experimentally to have further and further validity. But, no matter how many deductions are successfully tested, the theory can never be proved to be true in the absolute sense.

One needs only to show that one deduction is contrary to experience to invalidate, or, at least, considerably to limit, the theory. This view has been expressed very cogently by the philosopher Karl Popper. In a discussion of the way in which a scientific theory can be distinguished from other types of theories or beliefs he has written:

> "But I shall certainly admit a system as empirical or scientific only if it is capable of being *tested* by experience. These considerations suggest that not the verifiability but the falsifiability of a system is

to be taken as a criterion of demarcation. In other words: I shall not require of a scientific system that it shall be capable of being singled out, once and for all, in a positive sense; but I shall require that its logical form shall be such that it can be singled out, by means of empirical tests, in a negative sense: it must be possible for an empirical scientific system to be refuted by experience." (Popper, 1959.)

This is written with the careful choice of words common to philosophers, but the meaning is clear. Popper is attempting to distinguish scientific theories in the same way that Newton distinguished forces—by considering their unique properties. And just as we turn Newton's first law of motion round so that it tells us what is, or what is not, properly described as a force, so we can use this passage to tell us what is, and what is not, a meaningful scientific theory.

b) Rutherford's model of the atom

As another example of a theory in physics, let us consider the model of an atom whose origin is usually ascribed to Lord Rutherford. This is a specific case of an attempt to visualize something which is unseen and unseeable in terms with which we are more familiar. We shall consider shortly some of the evidence which led to the firm establishment of a picture of matter as being built up of small, chemically indivisible particles (atoms). As further experimental evidence accrued, it became necessary to extend this theory and to visualise atoms as having some sort of structure.

On the basis of his work in radioactivity, Rutherford suggested that the atom was "a system in which the greatest part of the mass of the atom was concentrated in a minute nucleus." He conceived that such an atom would have "a positive central charge Ne and be surrounded by a compensating charge of N electrons." (e is the electronic charge and N a positive integer). In making deductions from his theory the N electrons were considered as simply "a sphere of electrification containing a charge $-Ne$ supposed uniformly distributed throughout a sphere of radius R."

Here we have a purely pictorial representation of the atom. It allows both qualitative understanding of certain atomic events, and quantitative mathematical treatment. By "understanding" we mean that acceptance of the model allows us to offer an explanation of certain observed behaviour—in this case, the scattering of alpha particles.

The important thing to remember about a pictorial model such as this is that it is no more than a model. It provides an interpretation of the way atoms behave. Too much must not be read into the picture. This is a model of the atom seen "with the eyes half-closed." For example, you cannot ask of this model "what is the structure of the surrounding sphere of electric charge?" The model was not proposed to answer that question and, as Rutherford shows, it is irrelevant to the problem in hand.

Eventually, the detail in the model may be improved and a structure may be suggested both for the sphere of the electric charge and for the central nucleus. But this aspect of the model will remain unnecessary in interpretating Rutherford's experiments. We will return to this point later when we are considering models for the structure of matter.

1.5 THE CONTINUATION OF THE WORK OF BOYLE

In order to illustrate the process by which physics builds models and theories of the physical world, we shall consider one model—that of the structure of matter—in detail. But before doing so, and as a prelude, let us look again at the work of Boyle as he continued his investigations into the behaviour of gases. These led to the celebrated law.

Soon after the publication of his book on *The Spring of the Air*, a number of attacks were launched against his advocacy of Torricelli's theory. One of these, by Linus, is particularly interesting because of the way in which Boyle dealt with the argument. Boyle's explanation of the behaviour of the mercury column in Torricelli's classic experiment involved the support of the column by the weight of the external air. Since the pressure of the air is applied directly to the surface of the mercury in the surrounding dish, the

theory necessarily attributes to that air a certain resilience, or springiness, in resisting the crushing force of its own weight. Hence the title of the book itself. Linus objected to this interpretation, saying that whilst he agreed that air had both weight and springiness, these were not sufficient to support the mercury column. He suggested that there was also some agency within the space above the mercury which held up the mercury column.

Boyle's reply to this appeared in an appendix to the book and was introduced thus:

> "The other thing, that I would have considered touching our adversary's hypothesis is, that it is needless. For whereas he denies not, that air has some weight and spring, but affirms, that it is very insufficient to perform such great matters as the counterpoising of a mercurial cylinder of 29 inches as we teach that it may; we shall now endeavour to manifest by experiments purposely made, that the spring of the air is capable of doing far more than it is necessary for us to ascribe to it, to salve the phenomena of the Torricellian experiment."

The important point to note here is Boyle's statement that Linus' additional hypothesis is "needless." He intends to show that his own theory is sufficient. This is an important feature of all theories in physics since the time of Boyle. Such theories never, or should not, incorporate more elements than are sufficient to provide some explanation of the matter in hand. Especially has this to be borne in mind when dealing with those areas of speculation best described in terms of models. We saw in considering Rutherford's model of the atom how easy it is to ascribe more features to the model than can either be justified or needed for the immediate experimental evidence which Rutherford was considering.

1.6 THE DISCOVERY OF BOYLE'S LAW

The apparatus which Boyle used to test his ideas was the same as that still to be seen in many teaching laboratories today (see Fig. 1.4). He poured mercury in at the open end until the difference in levels was about 760 mm. This meant that the compressed air in the short closed limb was supporting both the pressure of the atmosphere acting on the top of the mercury in the open limb and a column of mercury whose pressure was equal to that of the atmosphere. Boyle noted that the volume of the compressed air was about one half of its volume when the levels in the two columns were the same, i.e. when the air was under atmospheric pressure alone. In his discussion of the result of this experiment he points out that the smaller the space into which the air is compressed the greater its pressure (or *spring*). He goes on

> "So that here our adversary may plainly see, that the spring of air, which he makes so light of, may not only be able to resist the weight of 29 inches (760 mm), but in some cases of above one hundred inches (250 cm) of quicksilver . . . ."

Boyle took a sequence of readings relating the compression of air to the pressure upon it, but seems to have been unable to see any mathematical relationship between the results. It was not in fact Boyle, but Richard Townley, who first detected in Boyle's readings, the relationship which we now know as Boyle's law, namely that the volume of a given mass of gas is inversely proportional to the pressure upon it.

This discovery so captured Boyle's imagination that the dispute which has led to it seems to have been almost forgotten. Boyle himself, in further experiments, seems to have shown that the law remains

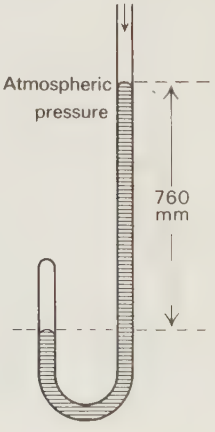


Fig. 1.4

true when the pressure on the gas sample is reduced below atmospheric. He also seems to have tested the effect of temperature and to have realized at least implicitly that to his law must be added the words "at constant temperature."

By his investigations into the "spring of air" and his establishment, at least experimentally, of the law which bears his name, Boyle not only confirmed the Torricellian theory but considerably extended it. He tried to produce a model of the air which would account for its compressibility. In fact he produced two, saying he was not willing "to declare peremptorily for either of them against the other." His first model consisted of "a heap of little bodies, lying one upon another, as may be resembled to a fleece of wool."

His second was of particles whirled around in the "subtle fluid" which was thought at that time to fill all space. Boyle said of this model:

> "it imports very little, whether the particles of air have the structures requisite of springs, or be of any other form (how irregular soever) since their elastic power is not made to depend upon their shape or structure, but upon the vehement agitation."

Here is yet another example of the way Boyle showed himself to be one of the forerunners of modern physical thinking. He would neither commit himself to one particular theory before he could produce experimental evidence to justify this comment nor would he ascribe to any one model more characteristics than he needed for the matter in hand (as when he said it matters little what the structure of the whirling particles is).

This seems an appropriate point to take up the problem of a present-day model for the structure of matter, and to use this to show how models can be built up in physics.

Chapter 2
A MODEL FOR A GAS

2.1 Pressure and temperature
2.2 Assumptions within the model

Experimental work with gases since Boyle's time has shown that, with the exception of their density, the physical properties of all gases are very nearly the same at atmospheric pressure.

They all conform to Boyle's law in the way their volumes change under applied pressure. They all expand as their temperature is raised at constant pressure and they all have (at low pressures) the same coefficient of expansion.

Since their properties seem to be largely independent of their chemical nature, we may expect that any model of a gas designed to explain these properties will be particularly simple. Let us try to build such a model.

The main criteria to be satisfied are:

a) the model should be as simple as possible;
b) the model should not be inconsistent with well established theories or models in related fields of study;
c) the model should be capable of further predictions which can be verified by experiment.

Criterion (b) immediately gives us a starting point for the model—we shall imagine the gas to be an assemblage of discrete particles which are not themselves capable of division. Throughout the period of recorded science, from the early Greeks to the present day, there has been speculation about the ultimate structure of matter. The recurring question has been whether matter can be subdivided without limit, or whether there are ultimate particles whose further subdivision would destroy the identity of the material. It would seem that the correct description of the structure of matter would depend on the outcome of such an experiment in subdivision. But even with the powerful machines of today which can take apart a nucleus, this idealized experiment has never been done. A piece of metallic copper, for example, has never been taken and subdivided until some threshold is reached, after which the subdivided particles no longer possess the properties we ascribe to copper. On the other hand copper gas does not behave like the solid material. The assumed particulate structure of matter is a model no less than any other. The assumption of this model enables us to explain any

number of phenomena, but the ultimate particles of matter remain so small that they will always be outside the realm of *direct* experience.

That the structure of matter is best described in terms of some ultimate particles was generally assumed by a number of scientists long before the model could be directly useful. Boyle assumed it in trying to give some explanation of the spring of air, but was unable to test either of his hypotheses.

Dalton was the first scientist to use the particulate model in a direct sense in order to provide a description of observed behaviour—in this case the laws of chemical combination. The development of the model in the hands of Gay-Lussac and Avogardo enabled the relative masses of the ultimate particles to be calculated. Furthermore, it was found necessary to classify these ultimate particles in terms of *atoms* and *molecules*.

The laws of chemical combination allow no other easy interpretation. So, in developing a model for the physical behaviour of gases, it would be unreasonable to assume anything other than a particulate nature for the gas.

The first feature of a gas that our model must explain is its high degree of compressibility and springiness. We have a choice between Boyle's two hypotheses. Either (a) the particles are in contact but themselves have an intrinsic springiness, like wool, or (b) the particles are far apart in a constant state of motion. This state of motion could account for the pressure exerted by a confined volume of gas by the collisions made by these particles with the side of the vessel.

Is there any other property of a gas which might guide one's choice?

2.1 PRESSURE AND TEMPERATURE

When the temperature of a sample of gas, confined to a given volume, is raised, its pressure also rises. The higher the temperature of the gas becomes, the greater the pressure it exerts on its container. The pressure of the gas is said to be temperature dependent. On the basis of the two proposed models, we must either assume that the intrinsic springiness or elasticity of the particles increases with temperature, or (model b) the particles move faster as the temperature rises.

It now seems possible to make a reasonable choice between the two models. For an independent piece of physical theory, now firmly established, suggests that heat is a measure of the energy transferred to or from a body by a gradient of temperature. A considerable body of experimental evidence can best be interpreted in terms of this theory. A model of a gas as widely spaced moving particles has much to commend it. It can be made consistent with this view of heat—for then the energy transferred to such a gas when it is heated becomes in the simplest model the kinetic energy of its moving particles.

This is an attractive possibility. The model we are proposing has yet to be tested but if it proves to be successful it would not only serve to explain the properties of gases, but also extend the scope of two other major theories: the particulate nature of matter and the identification of heat with energy transfer.

2.2 ASSUMPTIONS WITHIN THE MODEL

Before we seek to analyse this model mathematically, we must be clear about how much or how little we shall assume about the nature of these particles which are to be in a constant state of "vehement agitation." The pressure is assumed to result from the collisions between these particles and the walls of the container. Such particles must have mass, size and shape. They clearly attract each other to some extent, or the existence of liquid and solid forms of the same gaseous materials could not be explained. If the particles are to bounce off each other and the walls, there must also be some sort of repulsive force when they get close to each other, and to the molecules of the walls.

We might be able to analyse such a model mathematically, but at best, it would be extremely difficult. Can we, as a first approximation, make the model simpler?

Let us consider the difference in volume between gases and liquids. Every 1 cm^3 of water gives rise to 1600 cm^3 of steam. The space occupied by the gas is largely "empty." It should not be an unreasonable

approximation to assume that the particles themselves occupy a negligible proportion of this space.

And what of the inter-molecular forces? The mathematical analysis will be considerably simplified if we can assume that these are of very short range compared with the average separation of the particles of the gas and with the dimensions of the vessel. Then, the forces will only act during the actual process of collision. One outcome of this is that the pressure exerted by a gas is independent of the exact nature of the material of the walls of the container, since the forces between wall and gas particles only act for a small proportion of the total time. This fact is justified by experience.

The gas pressure is not only independent of the nature of the container, but also of its shape. To achieve this, the motion of the particles must be evenly distributed over every direction on any normal short-time average. We may imagine that this random distribution is maintained by internal collisions within the gas and in the collisions with the walls by particles which are moving in random directions as they approach the walls. A random distribution of directions must be taken to imply a random distribution of speeds as well, from the very high to the very low. However, we may assume that there is some time-independent average speed. The model itself has been partly suggested by the energy theory of heat. If the energy transferred to or from a gas as a result of a temperature gradient is to be interpreted as the kinetic energy of the motion of the particles, then the total kinetic energy must presumably remain constant if the temperature of the gas remains constant.

The constancy of the total kinetic energy of the moving particles in the gas, also means that either all the collisions are perfectly elastic, or that over any normal short time period, inelastic collisions are equally balanced by hyper-elastic collisions. In our model we shall take the simplest view and assume that the collisions are elastic.

So, summarizing, we are proposing a model of a gas as an assemblage of particles in a state of constant motion. The particles themselves have mass but occupy a volume which is negligible compared with the total volume of the gas. Their velocities are distributed at random, but with a well-defined average numerical value. All collisions with the walls will be assumed to be elastic.

One further word about what the model does not assume. It does not assume any structure for the particles except that they must be elastic. It is a matter of complete indifference as to whether they are what a chemist would call atoms or molecules. For this reason, we shall adhere to the term particle.

Further development of this model demands a certain familiarity with that behaviour of particles which is the proper study of dynamics. This will occupy Unit 2; moreover the considerations given to the ideas of pressure and of temperature deserve to be taken further. This will be done in Unit 3.

PROBLEMS: Unit One

1.1 The model of a gas as an assemblage of elastic particles moving about in all directions can give an adequate explanation of how a real gas, enclosed in a container, can exert a pressure. Explain how this model can also account for the pressure of the atmosphere, which is not in a closed container.

1.2 Imagine you are an air-molecule in the tyre of a motor-car. Describe *your* view of each of the following:

a) keeping the car supported,
b) more air being pumped into the tyre,
c) some petrol "molecules" coming in through the tyre valve with the air,
d) cooling down after a long run on a hot summer day,
e) a puncture.

1.3 Descartes proposed a very simple model to explain the phenomenon of seeing. He suggested that the eyes shot out rays which, on touching an object, gave rise to the sensation of sight.

a) Describe briefly (half a page) what you understand by the use of the word *model* when it is used in this context.
b) Suggest one experiment or observation you could make which could support or refute this model.
c) It has been suggested that all that experimental observation can do is to show if a model is *not* correct. Experiments can never provide absolute confirmation of the correctness of a model. Give a reason for accepting this view.

1.4 This extract is taken from *Introductory Physical Science* Chapter 8 (The Atom Model of Matter), by the IPS Group of EDC, Prentice-Hall, Englewood Cliffs, 1967.

"To help visualize what we meant by 'grainy or discrete' in describing radioactivity, we used a spray can spraying droplets of paint. Such a picture must not be taken too literally. Of course, we know that a piece of polonium does not look like a spray can, and we do not believe for a moment that the little particles that produce tracks in the cloud chamber are made of paint. We could equally well have thought of a sprayer of insecticide; and again we do not believe that what ever blackened the photographic plate in Becquerel's experiment is the same stuff that we use to kill bugs. We use a specific example or analogy as an illustration for a more abstract idea. The droplets of paint really stand for some kind of small particles, each of which can make a Geiger counter click or leave a track in a cloud chamber. If there are more particles, more tracks will be left; but the individual tracks are not affected by the number of particles. This abstract description of the radioactive process in terms of particles is an example of a 'model' or theory. This particle model clearly accounts for the increase in the blackening of a photographic plate with increasing exposure time and the decrease of the blackening as the distance from the source is increased.

But for a model, such as the particle model of matter, to be really useful, it has to offer us more than merely a convenient way of summarizing and accounting for facts we already know; a model must enable us to make predictions. Here is an example of a very simple model and a prediction we can make from it: Suppose someone hands you a sealed tin can. You shake it and hear and feel something slosh around inside. From this simple experiment of shaking the can,

you form a mental picture—a model—of what is inside. You conclude that the can contains a liquid. You have no idea what colour the liquid is or what it tastes or smells like, but you feel sure it has a property characteristic of liquids—it sloshes around inside a container. From this model (that the can contains something like a liquid) you can make a prediction: If you punch a small hole in the bottom of the can, liquid will drip out."

The following questions develop some of the ideas expressed in the extract you have just read.

a) Summarize in a sentence what this extract suggests as the main characteristics of a good model.

b) Every model has three ways in which it will relate to the group of phenomena of which it provides an underlying explanation:
   i) its most important relation is the extent to which it holds properties *in common* with the phenomena it describes,
   ii) nevertheless, the model will have properties which have no analogy with the phenomena the model describes,
   iii) and lastly the model may have properties which are quite the reverse of those exhibited by the phenomena.

Give examples of each of these features in the paint spray model of the radioactive process. You may use ideas from the extract but you should also add ideas of your own. As a hint, you may like to consider such things as colour of the particles, hardness of the particles, the way they would spread out with increasing distance from the source (do radioactive radiations behave in the same way?).

c) Explain how the particle model of the radioactive process can account for the observations made on the radioactive process in lines 20–24.

d) The extract suggests that a good model should not only account for present observations but should also predict some new properties. Can you suggest any additional properties of the radioactive process not mentioned in the extract which the model suggests?

e) The second paragraph offers an example of the way a model can make a prediction which can be tested. Suppose in the example chosen, you punched a hole in the bottom of the can and no liquid dripped out. What further conclusions might you draw and how would you set about testing them?

1.5 This is a question about describing different kinds of scientific statements. Here are some possible descriptions of such statements:

"states an experimental fact"
"makes an hypothesis"
"quotes a scientific law"
"is a rough estimate"
"is a deduction from earlier statements".

You are asked to give a brief description of each of the numbered statements in the passage below. Your descriptions should use phrases like those given above; you may use, combine or adapt the phrases above or invent others of your own.

1) If we assume that in a gas the atoms have a radius of about 10^{-10}m and a mean separation of about ten atomic diameters . . .
2) . . . it is clear that an alpha particle in traversing several centimetres of the gas must encounter some thousands of atoms of gas.
3) Only a minute fraction of such encounters, however, produce any appreciable deflection of the alpha particle.
4) It is difficult to avoid the conclusion that the greater part of the atomic volume is effectively empty.

Oxford and Cambridge Examination Board
Nuffield A-level Physics Examination (1970)

1.6 Sir Humphry Davy, a famous scientist who worked at the Royal Institution, London, during the early years of the nineteenth century, once said "Nothing leads to the advancement of knowledge as the application of a new instrument."

There are many instances in the history of science where this is true. Find out, and write a short note on, how

a) the telescope influenced the development of the present day model of the solar system,
b) heat engines profoundly influenced the theory of heat,
c) electron microscopes have influenced models of the structure of matter, both physical and biological.

BIBLIOGRAPHY: Unit One

In this unit, an historical survey of the work of Robert Boyle has been used to reflect some very general ideas about the nature of physical enquiry.

Conant, J. B. (ed), (1957). *Harvard Case Histories in Experimental Science*, Harvard University Press.

Provides a more detailed survey of the work of Robert Boyle. The final chapters of Volume I are also important for the account they give of the emergence of the particulate model of matter in the hands of Dalton, Avogadro, Gay-Lussac, etc. (Extract reprinted by permission.)

For an account of the status of models and theories in physics reflected in this unit you should read

Hesse, M. B. (1963). *Models and Analogies in Science*, Sheed and Ward.

A longer work by the same author (1961) *Forces and Fields*, Nelson, may be of interest to some. The first chapter deals with matters taken up in the smaller book.

Popper, K. R. (1959). *The Logic of Scientific Discovery*, Harper Torchbooks, Harper and Row. (Extract reprinted by permission.)

Tricker, R. A. R. (1965). *The Assessment of Scientific Speculation*, Mills and Boon.

Brings together a variety of modern views. In depth, it goes beyond the intention of the present unit, but the early chapters will be found to give useful background.

The kinetic theory of gases, Newton's theory of Gravitation, and the Rutherford model for the atom are developed in detail in Units 4, 7, and 9 respectively. Useful references to these developments will be found at the end of the appropriate units.

UNIT TWO
Motion, Force, and Energy

3 Introduction
4 The Investigation of Interactions: Some Experimental Techniques
5 Momentum and Mass
6 Force
7 The Effects of Forces: Work and Energy
8 Frames of Reference

Chapter 3
INTRODUCTION

3.1 Change and rate of change
3.2 Vector and scalar quantities
3.3 The addition of vectors
3.4 Comparison of the two methods
3.5 Vectors and navigation
3.6 Signs and vectors
3.7 Resolution of vectors
3.8 Summary

What is the nature of fire and heat? How do birds fly? Why is the sky blue? How is lightning caused? What is the description of the motion of the planets? Speculations on the Universe and the phenomena it exhibits have been made since the earliest times and our own immediate environment has produced many puzzles. The solutions of many of these have been found to rest in an understanding of motion, force, and energy. The development of such an understanding is essential if one is to study a physical science.

Comparatively little progress had been made towards understanding nature until mathematics had been developed sufficiently and men like Galileo and Newton had established, in the seventeenth century, our ideas on force and motion. Their outstanding contribution to the subsequent rapid increase of knowledge was recognized by writers such as Immanuel Kant: "When Galileo rolled balls down an inclined plane a new light burst upon all investigators of Nature"; and Albert Einstein, who wrote of Newton: "His clear and wide ideas will forever retain their significance as the foundation on which our modern conceptions of physics have been built." This was a generous comment from the man who overthrew much of that foundation.

This unit deals with the topics of motion, force, mass, and energy, and starts by considering momentum, a word often misused in everyday language. The use of some other terms must be clarified first, however.

3.1 CHANGE AND RATE OF CHANGE

The study of physics often involves the measurement of such quantities as length, time, temperature, energy, etc., and frequently we need to consider what happens when that quantity changes. This occurs so often that it is useful to adopt a "shorthand" notation. For instance, if the symbol l represents the length of a spring, then Δl is a shorthand way of stating *change of length*. The spring may change in length from 0.30 m to 0.35 m when the force acting on it is increased. In this case $\Delta l = 0.35 - 0.30 = 0.05$ m.

The rule for calculating changes is:

$$\text{change} = \text{final state} - \text{initial state}. \qquad (3.1)$$

If this rule is applied to the situation when the force is reduced and the length goes from 0.50 m to a final length of 0.45 m then $\Delta l = 0.45 - 0.50 = -0.05$ m. The significance of the sign is that it indicates that the change is a decrease of length.

When a change takes place it always takes some time to go from the initial to the final state. The temperature of the water in the cooling system of a car engine may change from 293 K to 353 K (i.e. from 20°C to 80°C) during the 5 min from starting up. The temperature rises at an average rate of 12 K per min since the change of 60 K occupied a time of 5 min. It is sufficient to write for

$$\frac{\text{change of temperature}}{\text{change of time}}$$

$$\frac{\Delta\theta}{\Delta t} = \frac{353 - 293}{5 - 0} = \frac{60}{5} = 12 \text{ K per min}$$

or better:

$$\frac{\Delta\theta}{\Delta t} = 0.2 \text{ K s}^{-1}.$$

At the end of the journey, the car is put into its garage and it is found that the engine cools from 353 K to 323 K (i.e. from 80°C to 50°C) in 20 min (1200 s).

$$\frac{\Delta\theta}{\Delta t} = \frac{323 - 353}{1200 - 0} = -\frac{30}{1200} = -\frac{1}{40} \text{ K s}^{-1}$$

$$= -0.025 \text{ K s}^{-1}.$$

In each of these examples, $\Delta\theta/\Delta t$ is a shorthand form for *rate of change of temperature* and the sign indicates whether an increase or a decrease of temperature is involved.

3.2 VECTOR AND SCALAR QUANTITIES

When the direction as well as the magnitude of a quantity must be quoted we are dealing with a *vector* quantity. The simplest example of a vector is displacement. This involves two pieces of information—the distance between two points and also the direction of one point from the other. For example when a pilot of an aircraft states his position he is giving his displacement from the airport (or some other stated landmark). He might say, "My position is 100 km due east of Rome."

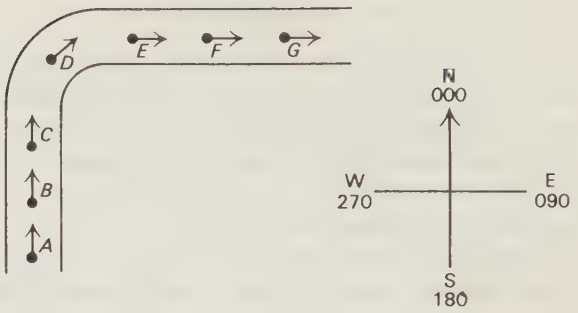


Fig. 3.1

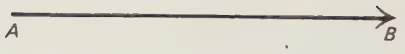
Another example of a vector is velocity, which is concerned with a speed and a direction. If his navigation is to be successful, a pilot needs to know the wind velocity, i.e. both the magnitude of the wind speed and its direction.

Scientists make a distinction between speed and velocity. Speed is an example of a scalar quantity and it is usually adequate to describe the rate at which a car is moving along a road, as a reading from a speedometer without any reference to direction since this is fixed by the road. Figure 3.1 helps to make the distinction. Thus a car travelling along the road shown can do so with a constant speed of 50 km h^{-1} all the way from *A* to *G*. But although its speed is constant, its velocity is not, at any rate not all the time. At *A*, *B*, and *C* its velocity is 50 km h^{-1} in a direction 000 (due north); at *D* its velocity has changed to 50 km h^{-1} heading 045 and it continues to change until it reaches *E* when it has become 50 km h^{-1} heading 090 and then continues to be constant through points *F* and *G*. This is a case in which the vector is changing, not by changing its magnitude but by changing its direction.

A vector is thus a quantity that has both a magnitude and a direction associated with it. It can be represented by an arrow drawn to scale so that its length gives the magnitude and its orientation gives the direction of the quantity.

3.3 THE ADDITION OF VECTORS

Suppose an aircraft starts at point A and flies until it reaches B where it has a displacement of 100 km, 090 from A. Figure 3.2 shows how it can be represented.



Scale :– 1 cm ≡ 20 km

Fig. 3.2

It then turns and flies until it is at C which has a displacement of 50 km, 045 from B. This part of the flight can be represented by Fig. 3.3.

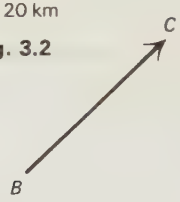


Fig. 3.3

These two flights can be added by means of a vector diagram drawn to scale. Note that the head of one vector is joined to the tail of the next (Fig. 3.4).

Now this aircraft could have reached C by flying along the route AC (Fig. 3.5). Consequently, flight AB plus flight BC is equivalent to flight AC, or $\boldsymbol{AB} + \boldsymbol{BC} = \boldsymbol{AC}$. Note the shorthand $\boldsymbol{AB}$, etc., which indicates the vector AB, etc. $\boldsymbol{AC}$ is called the *resultant* of $\boldsymbol{AB}$ and $\boldsymbol{BC}$. A resultant is obtained by adding vectors using the triangle method outlined and not merely by adding the two distances together. By taking a measurement off the triangle, $AC = 140$ km, which is clearly different from the addition of the two scalars 100 km and 50 km.

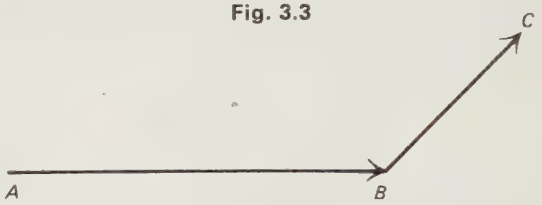


Fig. 3.4

It would be possible for the pilot to arrive at C by flying the vector 50 km, 045 first, arriving at D, and secondly to fly along the vector 100 km, 090 from D (Fig. 3.6).

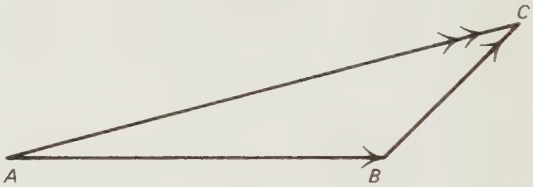


Fig. 3.5

Adding these vectors gives a triangle similar to ABC and the same resultant (Fig. 3.7).

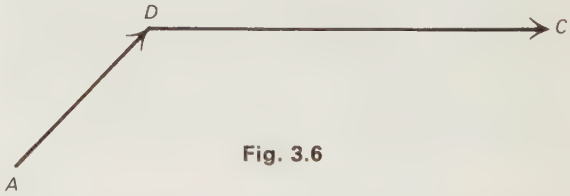


Fig. 3.6

The resultants are the same, since if Figs. 3.5 and 3.7 are superimposed, a parallelogram is obtained (Fig. 3.8).

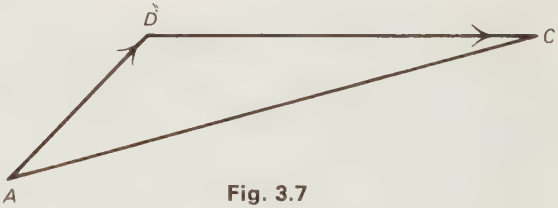


Fig. 3.7

Thus there are two alternative ways of adding vectors which really amount to the same process.

1. The triangle method: Draw the vectors to scale to form two sides of a triangle. The head of the first vector must be joined to the tail of the second. The resultant is the third side of the triangle and its direction is from the tail of the first vector to the head of the last one (e.g. Figs. 3.5 and 3.7).

2. The parallelogram method: Draw the vectors to scale so that they form adjacent sides of a parallelogram. The resultant is the diagonal drawn from the corner with two tails to the corner with two heads—as in Fig. 3.9(a) not as in Fig. 3.9(b).

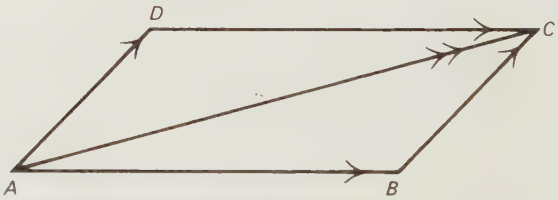


Fig. 3.8

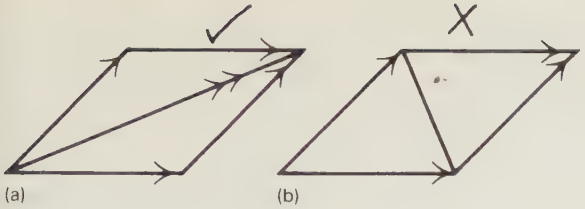


Fig. 3.9

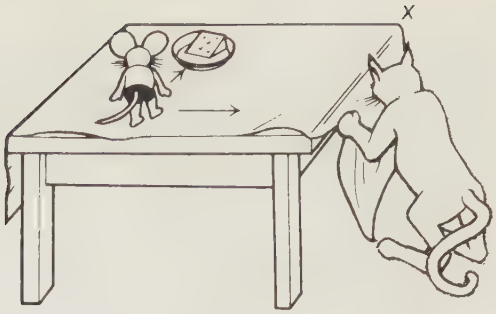
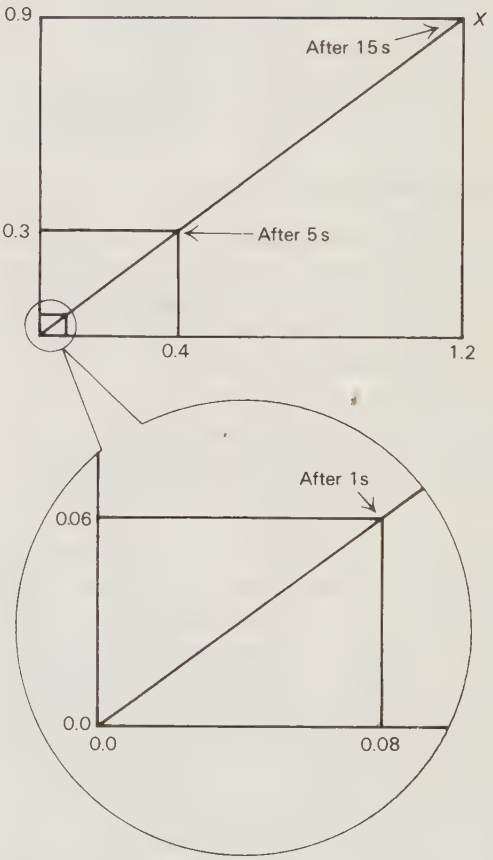


Fig. 3.10

3.4 COMPARISON OF THE TWO METHODS

It may appear strange to use the parallelogram method since it involves drawing two seemingly unnecessary extra lines in order to obtain a resultant. Yet there are many cases when the parallelogram method is more appropriate as it allows one to understand more easily what is happening. In general, the triangle method should be used when one vector acts first, followed by the second, whereas the parallelogram method may be best when both vectors act during the same period of time, or act at the same point.

How could two displacements occur at once? Consider the cat Tom, and the mouse Jerry (Fig. 3.10). Jerry is on a table and walking across it towards a piece of cheese. While this is happening, Tom pulls the tablecloth from left to right. Thus Jerry is moving in two directions at the same time. In 5 s he may travel 0.3 m towards the cheese and 0.4 m to the right towards Tom, who knows about vectors and expects his prey to arrive at the corner X, since the table measures 0.9 m wide by 1.2 m long. Figure 3.11 shows the positions of Jerry, relative to the table, after 1, 5, and 15 s. The 1 s case is of particular importance as it deals with displacement per second, or velocity. Thus the rules of vector addition apply to velocities. Jerry has a velocity of 0.06 m s^{-1} across the table and 0.08 m s^{-1} along the table. This gives him a resultant velocity of 0.10 m s^{-1} towards the corner X, as the diagonal of the circled diagram is 5 cm long which represents a distance of 0.10 m.



Scale:– 1 cm ≡ 0.2 m
(1 cm ≡ 0.02 m for the circled part)

Fig. 3.11

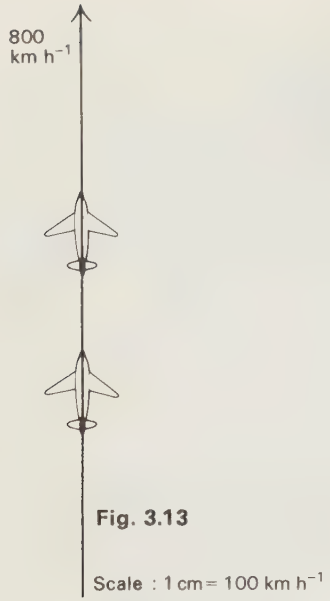


Fig. 3.13

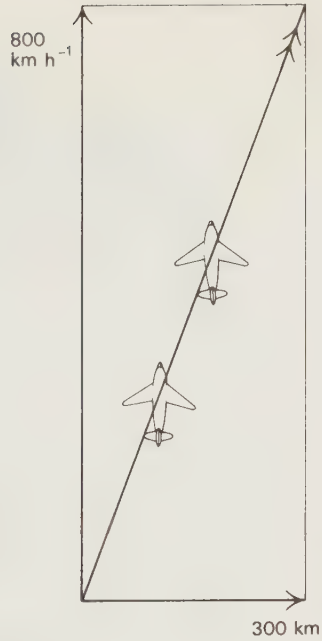


Fig. 3.12

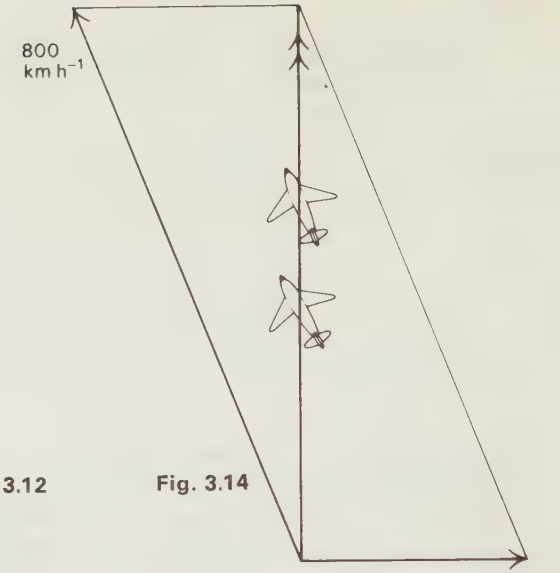


Fig. 3.14

3.5 VECTORS AND NAVIGATION

An aircraft pilot makes use of the displacement vector when he states his position, and velocity vectors and the rules of vector addition when deciding upon the direction to head the aircraft in order to arrive at a given destination. Suppose the aircraft has a speed of 800 km h^{-1} through the air and that at first there is no wind blowing (Fig. 3.12). If he points the nose of the aircraft due north (or on a heading 000), his track over the ground is also in the direction 000.

If he now finds himself in a jet stream (a high speed westerly wind that occurs at altitudes of between 7 km and 13 km. Wind speeds can exceed 300 km h^{-1}) which is blowing with a velocity of 300 km h^{-1} from 270, his resultant motion relative to the ground will be the result of having two velocities at the same time. While the aircraft is moving through the air at 800 km h^{-1}, the air is carrying it eastwards at 300 km h^{-1}.

The resultant can be found by drawing to scale a velocity parallelogram (Fig. 3.13). The resultant velocity is 854 km h^{-1} in a direction 020, and although the aircraft is heading (or pointing) due north, it is moving over the ground in a direction 020 and at a speed of 854 km h^{-1}.

However, the pilot wishes to arrive at a destination which is due north of his starting point. His resultant velocity must be in a direction 000, and a velocity parallelogram can give the direction he must head, to achieve this. By heading the aircraft in a direction 338 as shown in Fig. 3.14, he will move over the ground due north and at a speed of 742 km h^{-1}.

3.6 SIGNS AND VECTORS

Let the two vectors $\boldsymbol{AB}$ and $\boldsymbol{BC}$ be equal in magnitude, and suppose they are added as in Fig. 3.15; the resultant is given by: $\boldsymbol{AB} + \boldsymbol{BC} = \boldsymbol{AC}$. Consider what happens when $\theta = 0$. The triangle becomes a line as in Fig. 3.16, and the resultant of the two vectors is zero since A and C coincide. Thus, $\boldsymbol{AB} + \boldsymbol{BC} = 0$ or $\boldsymbol{AB} = -\boldsymbol{BC}$ or $\boldsymbol{AB} = -\boldsymbol{BA}$ since $\boldsymbol{BC} = \boldsymbol{BA}$ in Fig. 3.16. The significance of a negative sign attached to a vector is that the direction of the vector is reversed, or turned through 180°.

This application of a sign is frequently used to indicate the direction of motion of an object. For example, suppose an escalator can move upwards or downwards at a speed of 0.6 m s^{-1}, and that we adopt the convention that the upward direction is + or positive, and the downward direction is − or negative. A man, who can walk up a stationary staircase at a

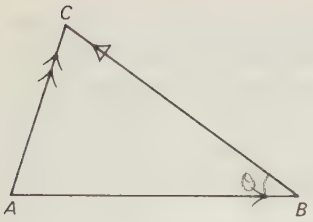


Fig. 3.15

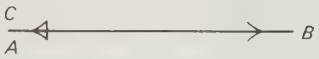


Fig. 3.16

speed of 0.4 m s^{-1}, is in a hurry and walks up the ascending escalator. This situation can be described as:

velocity of man relative to escalator = + 0.4 m s^{-1},
velocity of escalator = + 0.6 m s^{-1},
resultant velocity of man = + 0.4 + 0.6 = + 1.0 m s^{-1}.

If the escalator is reversed and the man continues to walk upwards the situation is then:

velocity of man relative to escalator = + 0.4 m s^{-1},
velocity of escalator = − 0.6 m s^{-1},
resultant velocity of man = + 0.4 + (− 0.6) = − 0.2 m s^{-1}.

Thus by changing the velocity of the escalator from + 0.6 m s^{-1} to − 0.6 m s^{-1}, i.e. by reversing it, the velocity of the man changes from + 1.0 m s^{-1} to − 0.2 m s^{-1}; the negative sign indicating that he is descending.

3.7 RESOLUTION OF VECTORS

It is often convenient to replace a single vector by two component vectors. For example, we may imagine a truck on rails being pulled by a man using a rope inclined at an angle θ to those rails (see Fig. 3.17a).

We may *resolve* the force F into two component vectors at right angles to one another, F_x along the rails and F_y at right angles to the rails (see Fig. 3.17b). We see at once that, if the law of vector addition is applied,

$$F = F_x + F_y.$$

Moreover we can see from the figure that

$$F_x = F \sin\theta$$

and

$$F_y = F \cos\theta.$$

Of the two component vectors, only F_x is effective in moving the truck since F_y is balanced by the opposing thrust of the rails. F_x is able to accelerate the truck to the constant velocity which develops when it is just equal to the frictional forces opposing it.

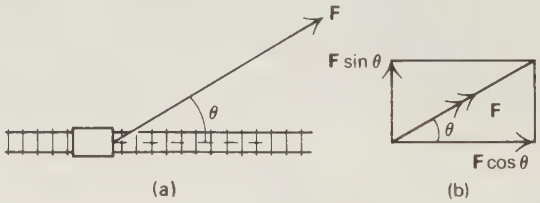


Fig. 3.17

These ideas concerning rates of change and vectors will be used often in the work that is to follow.

3.8 SUMMARY

Vector quantities possess both magnitude and direction and may be added by the application of the triangle (or the parallelogram) law. Such quantities will be represented by the appropriate letter printed in bold type face whilst the magnitude of the quantity will be represented by the same letter printed in ordinary type. For example the magnitude of the force F is represented by F.

It should be noted that the additions of vectors discussed in the examples given above can only be tested by experiment.

Chapter 4
THE INVESTIGATION OF INTERACTIONS: SOME EXPERIMENTAL TECHNIQUES

4.1 The reduction of friction in the laboratory
4.2 Methods for measuring speed
4.3 An investigation of interactions
4.4 The rule begins to emerge
4.5 Extending the rule
4.6 The new rule
4.7 The concept of mass
4.8 A conservation rule

When a scientist has a complex problem to solve he starts by seeing what factors are involved. He may then attempt to remove or at least reduce the effect of some factors; he may proceed by keeping other factors constant and see if there is a simple law which governs the behaviour of the remaining factors.

What could possibly be simple about anything as wild as a collision or an explosion? At first sight it seems unlikely that there is a simple relationship between the motion of objects before and after a collision.

The collision of two cars is a very complex situation; they are likely to be of different masses and to have different and even changing speeds. A simplification to aid the investigation would be to have two identical cars, each travelling with the same speed, but even this is complicated by the fact that an engine must be working to maintain this constant speed, since friction would otherwise cause a retardation. An idealized colliding system, as far as the scientist is concerned, is one in which no friction exists so that no engine is necessary.

4.1 THE REDUCTION OF FRICTION IN THE LABORATORY

A number of pieces of apparatus have been designed for experimenting with collisions in which the effects of friction have been reduced.

a) Trolleys on ball-bearing wheels

These are designed (see Fig. 4.1) so that two or more trolleys can be stacked and the mass moving varied. Some trolleys are equipped with a spring plunger that can be released by striking a peg. By this means the "explosion" of two trolleys can be investigated.

b) The linear air track

The length of the track is drilled with small holes from which air, supplied by a vacuum cleaner, is blown so supporting the vehicles clear of the actual track (see Fig. 4.2).

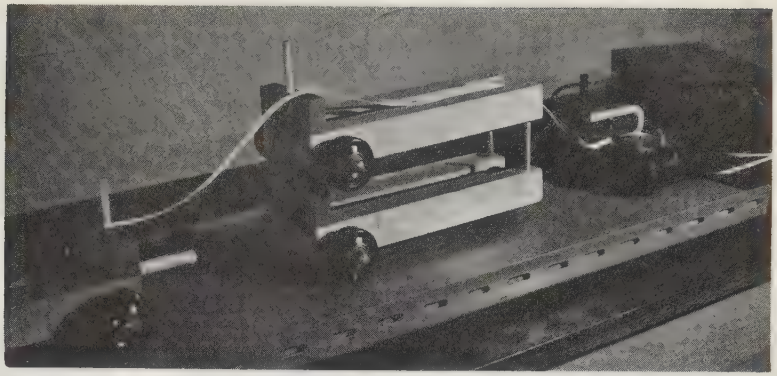


Fig. 4.1 Dynamics trolleys with timer. (Courtesy of Philip Harris, Ltd.)



Fig. 4.2 A linear air-track with vehicles. (Courtesy of Griffin and George, Ltd.)

c) Pucks on a flat glass plate

These can either be operated by air, supplied by an electrically driven pump carried on the puck itself (see Fig. 4.3), or by a small piece of solid carbon dioxide which continuously sublimes to the gaseous state (see Fig. 4.4). In both types, the gas pressure underneath the puck supports the puck clear of the glass plate.

d) The air table

Similar in operation to the air track, the air table allows motion and collisions in two dimensions to be studied (see Fig. 4.5).

An astronaut in orbit is in an almost ideal laboratory for the performance of experiments on frictionless motion. In Fig. 4.6 you can see his toothbrush moving freely across the craft from one side to the other without the support of wheels or gas pressure.

In all of these cases, one should question whether friction has been eliminated, or only reduced sufficiently for it to be neglected. No matter how well made are the wheels and bearings of a trolley, some friction will still remain. Experience tells us that a trolley which has been pushed on a horizontal table will eventually come to rest because friction gradually dissipates the trolley's energy. But what about the air track vehicles and the pucks? They certainly have no wheels yet experiment shows that they gradually slow down and lose energy. Where can the friction be? This time it is caused by viscosity, or gas friction, occurring mostly in the narrow gap between the vehicle and the track or between the bottom of the puck and the glass plate. Does the spacecraft provide the ideal frictionless environment? Again, it is gas, the atmosphere of the craft, that provides the friction or resistance to motion; this time, however, the resistance is very small indeed and it could be reduced even further by having a vacuum in the spacecraft. We will not be

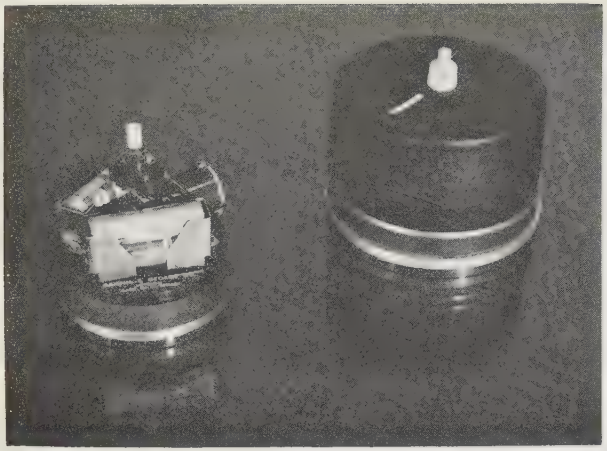


Fig. 4.3 An air-driven puck. (Courtesy of Griffin and George, Ltd.)



Fig. 4.4 Carbon dioxide pucks. (Courtesy of Philip Harris, Ltd.)

going to this extreme. We will use trolleys, air tracks, and pucks, and accept that in these cases the effect of friction may be neglected.

Problem The force of friction tends to retard a moving body. How could you arrange a trolley experiment so that the earth's gravitational pull on a trolley compensated for the frictional force?

Explain whether or not your method could be applied to the case of two trollies moving in (a) the same direction and (b) opposite directions.

4.2 METHODS FOR MEASURING SPEED

In order to determine speed a measurement of a distance and of the time taken to cover that distance is required.

If the speed of a car is being measured, it is sufficient to use a stopwatch and time it over a distance of a kilometre. A very fast car taking 20 s to cover this distance can be "hand timed" to an accuracy of $\pm$ 0.2 s. This represents an error of 1%, which is the worst that can be expected as a slower car will take longer and this can be measured to a better percentage accuracy. For example, 40 $\pm$ 0.2 s represents an error of $\frac{1}{2}$%.

In the laboratory, the distance over which a trolley or a puck is timed is comparatively short; it is not likely to exceed 1 m and in many experiments this distance will be covered in less than 4 s. Consequently, the best that can be attained with a stopwatch is an accuracy of 4 $\pm$ 0.2 s or 5%. A trolley going at twice the speed and timed over half the distance will have its speed measured to an accuracy of only 20% if a watch is used. Consequently, other methods are used for measuring the short times involved in laboratory experiments.

a) The ticker-tape timer

This is a familiar piece of apparatus that when driven from a 50 Hz a.c. supply makes a dot every 0.02 s on a piece of paper tape that is attached to a trolley. If the trolley is accelerating, the dots become further apart. The distance between any two dots is the distance travelled by the trolley in 0.02 s.

b) Photocell and lamp method (see Figs. 4.7 and 4.8)

A phototransistor is used to operate the starting and stopping of a dekatron scaler/timer which measures time in units of 0.001 s. When a light beam shines on the phototransistor the scaler does not count time

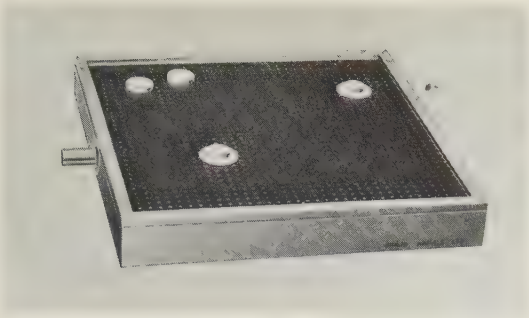


Fig. 4.5 Air table with pucks. (Courtesy of Philip Harris, Ltd.)

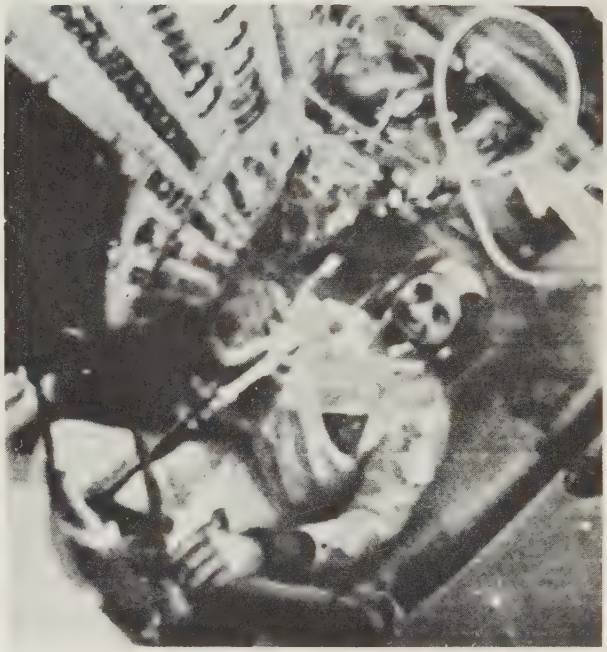


Fig. 4.6 A dynamics laboratory in space; TV transmission from Apollo 8 on December 22, 1968. (Courtesy of United States Information Service.)

units which are derived from a 1 kHz oscillator. If the beam is interrupted, the scaler starts to count and does so until the beam is restored. A card, 0.1 m long mounted on an air track vehicle can be arranged to interrupt the beam as shown in Fig. 4.7. Thus the time taken by the vehicle to travel 0.1 m can be measured.

The phototransistor can also be used to control the starting and stopping of a specially designed clock that can measure down to 0.01 s. Figure 4.8 shows how á card of known length, attached to a trolley this time, can be used with such a clock, and photocells.

c) Multiflash (stroboscopic) photography

This is a particularly versatile method of analysing the motion of a moving body. As can be seen from Figs. 4.10 and 4.14, a single photograph can record all the motions involved in simple and complex interactions, including motions that take place in two dimensions (see also Figs. 8.4 and 8.5). Multiflash photography involves the taking of a sequence of pictures, on a single piece of film, at regular intervals. This can be achieved in two ways:

1. The camera shutter is opened in the dark and the event to be recorded is illuminated by a rapidly



Fig. 4.7 A linear air-track with a timer-scaler. (Courtesy of Griffin and George, Ltd.)

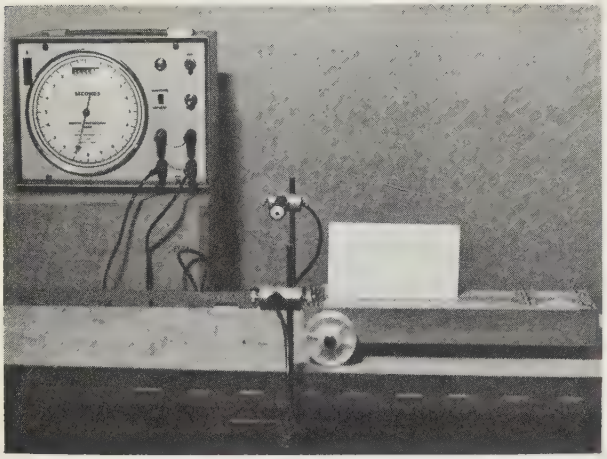


Fig. 4.8 Dynamics trolley with an electrically operated clock. (Courtesy of Griffin and George, Ltd.)

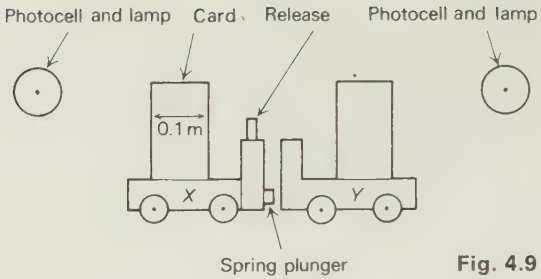


Fig. 4.9

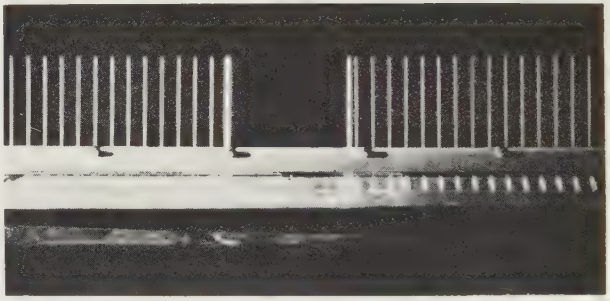


Fig. 4.10 (Photograph by B.T.)

flashing light which may be obtained from an electronic flash tube. The camera shutter is closed when the event is completed.

2. Alternatively, the trolleys or pucks can be illuminated by a steady light and the film is then exposed at rapid regular intervals by means of a mechanical shutter. This consists of a disc, having six equally spaced slots around its edge, that is rotated at a constant rate—typically 5 revolutions per second—in front of the camera lens. When the camera shutter is open, the film obtains a glimpse of the event to be recorded every time a slot passes.

4.3 AN INVESTIGATION OF INTERACTIONS

In order to discover a rule connecting the motion of objects before and after an interaction, some data must be obtained by doing experiments. The problem is first simplified by using two identical objects in a variety of interactions. The objects are either two trolleys or two air track vehicles, and, to illustrate their use all the various timing methods mentioned earlier are used.

Experiment A: An explosion

Two trolleys are initially at rest and in contact (Fig. 4.9). The spring plunger is released and the trolleys (X and Y) move apart, their individual speeds being measured by the card, photocell and clock method. The results are given in Table 4.1.

Table 4.1

	Trolley X	Trolley Y
Card length (m)	0.10	0.10
Transit time (s)	0.27	0.27
Final speed (m s^{-1})	0.37	0.37
Final velocity (m s^{-1})	−0.37	+0.37
Initial velocity (m s^{-1})	0.00	0.00

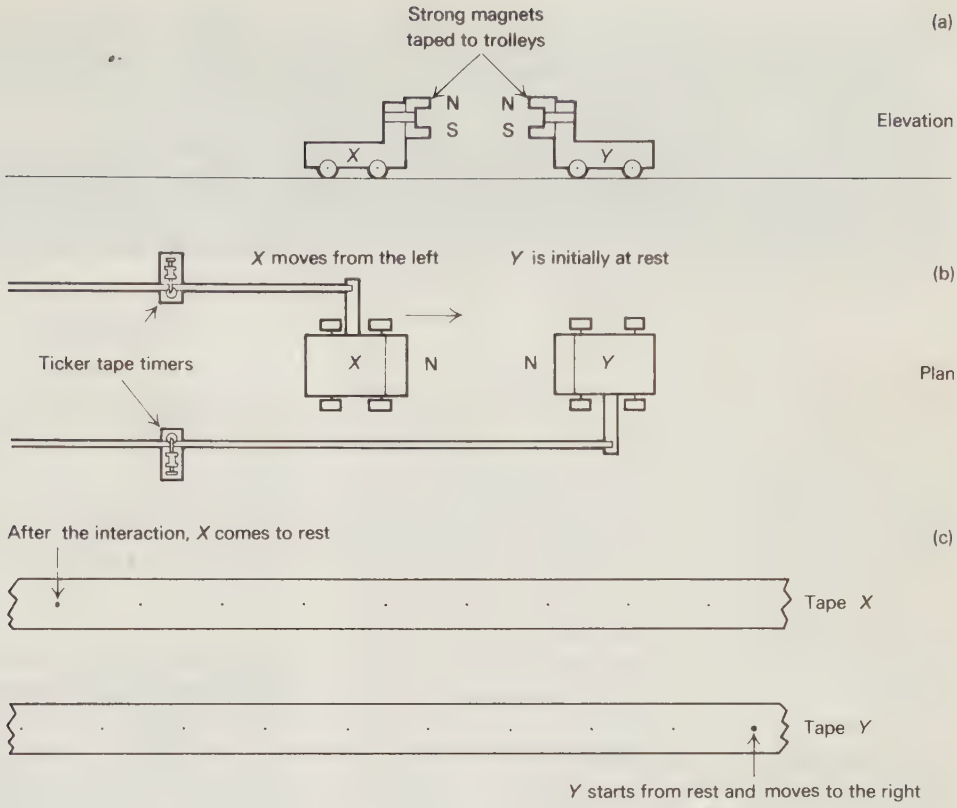


Fig. 4.11

As mentioned earlier, the sign indicates the direction of motion:

− to the left and + to the right.

Experiment B: A collision with a stationary object

Two air track vehicles P and Q are fitted with rubber band buffers. A white drinking straw is attached to each vehicle and these appear as the vertical lines on the multiflash photograph (Fig. 4.10). The space between two lines is proportional to the distance travelled in 1/30 s. Measurements from the photograph give velocities of the images of the straw. (Table 4.2.)

Table 4.2

	Vehicle P	Vehicle Q
Initial velocity (m s^{-1})	+0.066	0
Final velocity (m s^{-1})	0	+0.066

Experiment C: An interaction using magnets

The main difference between this and the previous experiment is that the interaction takes place *at a distance* by means of the repulsion between the magnets (see Fig. 4.11). The vehicles do not touch. The results taken from the tapes are shown in Table 4.3.

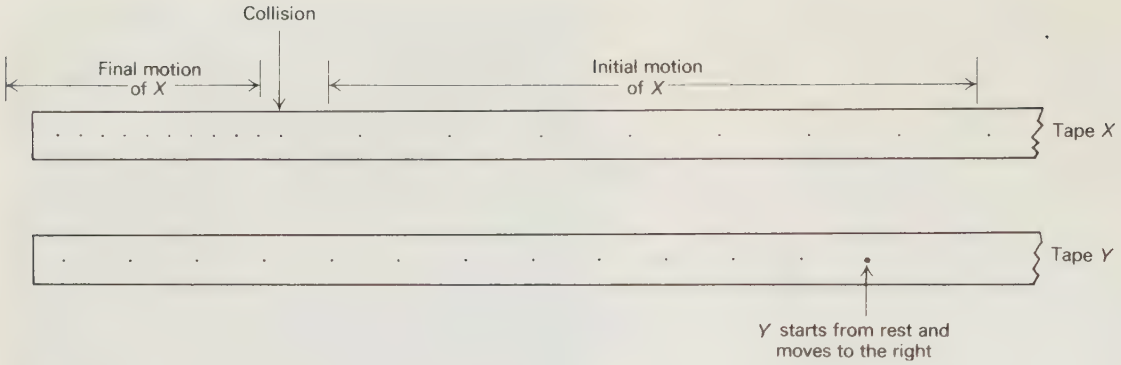


Fig. 4.12

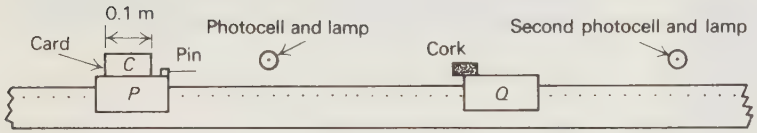


Fig. 4.13

Table 4.3

	Trolley X	Trolley Y
Initial velocity (m s^{-1})	+0.55	0.00
Final velocity (m s^{-1})	0.00	+0.55

Experiment D: A collision with a stationary object without using rubber or magnetic buffers

The arrangement of the previous experiment is used with the magnets removed so that a "hard" interaction occurs. The tapes obtained are as shown in Fig. 4.12.

This time, trolley X does not come to rest after the interaction but has a small final velocity. Table 4.4 shows the results given by the tapes.

Table 4.4

	Trolley X	Trolley Y
Initial velocity (m s^{-1})	+0.60	0.00
Final velocity (m s^{-1})	+0.15	+0.45

Experiment E: A collision between two bodies which stick together

In each of the previous experiments, the objects separate after the interaction. Although both are finally moving in experiment D, trolley Y moves away from X with a relative velocity of 0.30 ms^{-1}. In experiment E we examine the situation where separation does not occur. This is achieved by having a sharp pin attached to one air track vehicle that sticks into a small piece of cork fixed to the second vehicle, when the collision takes place (Fig. 4.13).

Vehicle P moves from the left and passes the first photocell. P hits Q, which is initially stationary, and both move off together past the second photocell. Table 4.5 shows the results obtained.

Table 4.5

	Vehicle P	Vehicle Q
Card length (m)	0.100	——
Transit time of P (s)	0.242	——
Transit time of $P + Q$ (s)	0.485	0.485
Initial velocity (m s^{-1})	+0.413	0.000
Final velocity (m s^{-1})	+0.206	+0.206



Fig. 4.14 (Photograph by B.T.)

Note: In practice, the two transit times can be taken with one scaler if it is operated by the two phototransistors connected in series. Then, if either beam is interrupted, the scaler starts. The vehicles must be travelling sufficiently slowly so that the scaler reading can be taken after the first transit and before the second one has commenced.

Experiment F: A collision between two moving objects

In the previous experiments, one object was at rest before the second one collided with it. This time, both are in motion moving towards one another for a head-on collision, after which each rebounds and travels back to its starting point. To record this motion, multiflash photography of air track vehicles is utilized. The reverse motion can be photographed without being superimposed on the forward one if a shutter is fitted in front of the track. This takes the form of a hinged flap, about 1 m long and 0.05 m high that is operated by hand when the interaction occurs. The lower parts of the drinking straws are then revealed and the upper parts concealed from the camera. A typical photograph taken by this method is shown in Fig. 4.14.

In the photograph, the space between two images is proportional to the distance travelled in 1/30 s. Measurements from it give the results shown in Table 4.6.

Table 4.6

	Vehicle *P*	Vehicle *Q*
Initial velocity ($m\ s^{-1}$)	+0.240	−0.150
Final velocity ($m\ s^{-1}$)	−0.120	+0.210

4.4 THE RULE BEGINS TO EMERGE

These experiments provide data about the motion of two similar objects before and after a wide variety of different interactions, including one in which they do not even touch. If the results of each experiment are examined we can see that the following rule applies: (Sum of velocities before interaction) = (Sum of velocities after interaction).

Note: Try adding speeds instead, by ignoring the signs, and see if the rule works in every case. You should decide that this rule only works if the velocities are treated as vectors.

4.5 EXTENDING THE RULE

We have obtained a satisfactory rule for the simplest possible case, namely two isolated, similar objects, but real life collisions are not always so conveniently arranged. To make our rule more useful, we will change a variable. We will keep the system isolated and experiment with objects that are not similar. Note that we are only changing one variable. This is the golden rule in scientific investigations. Other variables, such as friction, are not allowed to creep in and confuse the issue. The effect of such forces is treated separately.

So that the interacting objects differ by an obvious amount, the next experiments are performed with stacked trolleys. Thus one trolley can interact with two, two trolleys can interact with three, and so on.

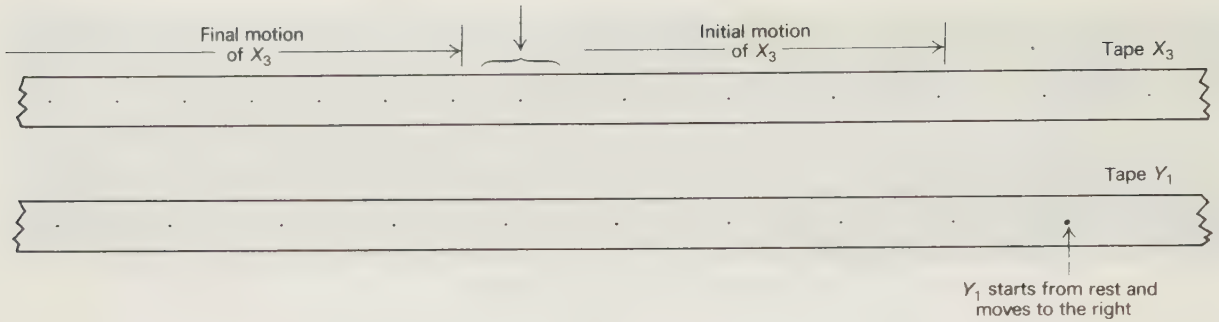


Fig. 4.15

Experiment G: An explosion between a single, and a double trolley

The arrangement for timing, etc., is the same as for experiment *A*. The symbol Y_2 refers to the double trolley. Table 4.7 gives the results.

Table 4.7

	Trolley X	Trolley Y_2
Card length (m)	0.10	0.10
Transit time (s)	0.22	0.44
Final velocity (m s^{-1})	−0.46	+0.23
Initial velocity (m s^{-1})	0.00	0.00

The sum of the velocities before interaction clearly equals zero. The sum of the velocities after interaction will only be zero if the velocity of the double trolley is doubled.

Experiment H: An explosion between a double, and a triple trolley

The trolleys are referred to as X_2 and Y_3 in Table 4.8 which summarizes the results obtained using the same technique as in experiment A.

Table 4.8

	Trolley X_2	Trolley Y_3
Card length (m)	0.10	0.10
Transit time (s)	0.16	0.24
Final velocity (m s^{-1})	−0.63	+0.42
Initial velocity (m s^{-1})	0.00	0.00

The sum of the velocities before interaction again equals zero. The sum of the velocities after interaction will only be zero if the velocity of X_2 is doubled and that of Y_3 is multiplied by three.

4.6 THE NEW RULE

From the data obtained by these last two experiments a modified rule emerges that applies to explosions between isolated objects of any size. Thus:

$$n_X u_X + n_Y u_Y = n_X v_X + n_Y v_Y, \qquad (4.1)$$

where the symbols have the following meanings:

- n_X and n_Y refer to the number of trolleys in stacks X and Y.
- u_X and u_Y refer to the initial velocities of X and Y.
- v_X and v_Y refer to the final velocities of X and Y.

Equation (4.1) is a statement of the rule that applies to all of the experiments A to H, but it requires further testing to see if it is true for *collisions* between objects of any size.

Experiment I: A collision between a triple trolley, X_3 and a stationary single trolley, Y_1

The arrangement is the same as for experiment D. The tapes obtained are as shown in Fig. 4.15. Table 4.9 gives the results taken from these tapes.

Table 4.9

	Trolley X_3	Trolley Y_1
Initial velocity (m s^{-1})	+0.70	0.0
Final velocity (m s^{-1})	+0.45	+0.75

If Eq. (4.1) is tested with these results we will see it is true for collisions as well as explosions.

4.7 THE CONCEPT OF MASS

In experiments G, H, and I the number of trolleys on each side of an interaction is significant in deciding the velocities after the interaction. In each case, the number of trolleys is a measure of the quantity of matter, or mass, involved. The mass of a body governs its behaviour during an interaction. For example, experiment G shows that after an explosion a body with a large mass moves off with a smaller velocity than one with a small mass. The larger mass has more *inertia* and tends to stay at rest. Also, experiment I demonstrates that, during a collision, a large mass undergoes a smaller change of velocity than does a small mass. (X_3 underwent a change of velocity of $0.45 - 0.70 = -0.25$ ms^{-1} whereas Y_1 suffered a change of 0.75 ms^{-1}.) Again we can say that the larger mass has more inertia as it tends to maintain its velocity. Mass can be defined as a measure of the inertia of a body, as well as being the quantity of matter in the body.

Newton summed up these findings in his first law, (Newton I) which states: "Every body remains at rest, or moves with constant velocity, unless acted on by an unbalanced force."

Remember that "constant velocity" implies that it is moving in a straight line and that its speed is constant.

The meaning of the term "unbalanced force" is best understood by reference to some examples. Consider a ball at rest on a billiard table. Two forces act upon the ball; the earth exerts a pull on it and the table supports it. These two balance. Then it is struck by a cue. The cue exerts a force on the ball and, as we shall see, the ball exerts an equal and opposite force on the cue. The force exerted by the cue on the ball is unbalanced as far as the ball is concerned. The ball is accelerated and so acquires a velocity. (At the same time, the cue is decelerated.) Subsequently the ball comes slowly to rest, because friction with the table now provides an unbalanced force.

Consider, too, a mass hanging from a thread. The thread supports it and the earth pulls upon it—there are two balanced forces acting. Now the thread breaks. Under the action of the now unbalanced pull of the earth, the mass falls towards the floor.

Or again, a motor car is accelerated to its maximum speed. The force exerted by the engine through the transmission is unbalanced until, at the maximum speed, it is exactly equal to the sum of all the frictional forces that exist. The forces are in balance as the car moves with constant motion.

4.8 A CONSERVATION RULE

Returning to Eq. (4.1), the term mass can now be used instead of "numbers of trolleys." If m_X and m_Y mean mass of X and mass of Y respectively, then the equation becomes:

$$m_X u_X + m_Y u_Y = m_X v_X + m_Y v_Y. \qquad (4.2)$$

This tells us that the sum of (mass × velocity) before an interaction is equal to the sum of (mass × velocity) after the interaction.

Chapter 5
MOMENTUM AND MASS

5.1 The law of conservation of momentum
5.2 Momentum as a vector
5.3 Isolated systems
5.4 Making measurements of mass
5.5 The unit of mass
5.6 Inertial mass and gravitational mass
5.7 The cause of momentum change

The quantity (mass × velocity), is so important that it is given a special name: *momentum*.

As we have seen, momentum is concerned with interactions between bodies and it is in this context that the idea is most fruitful. In everyday speech the word is sometimes used to describe motion, for example, a cyclist free-wheeling down a hill gains "momentum." This is, of course, perfectly true; but the speaker probably wishes to say no more than that the velocity of the cyclist is increasing. It would also be correct to say that the cyclist gains kinetic energy. Both momentum and kinetic energy depend on velocity but it would be an error to suggest that these words were nothing more than synonyms for velocity.

Momentum is important because in a collision or explosion within an isolated system, the total momentum is constant. We say that momentum is conserved.

5.1 THE LAW OF CONSERVATION OF MOMENTUM

This law states: "The total momentum of an isolated system of bodies, interacting only with each other, remains constant."

5.2 MOMENTUM AS A VECTOR

In deducing this law from the results of experiments, it was essential to assign directions to the various motions. Likewise in applying the law, it is vital to treat the product (mass × velocity) as a vector. Velocity is a vector and mass is a scalar that is acting as a multiplying factor. Consequently, momentum can also be classed as a vector.

5.3 ISOLATED SYSTEMS

An isolated system is composed of a number of masses that only interact among themselves. No external force such as friction "connects" them to other masses outside of the system. In all of our experiments it is clear which masses form the isolated system: trolley X and trolley Y form one isolated (nearly) system; vehicle P and vehicle Q form a perhaps better isolated system. We can say that these are isolated because, for instance, P only interacts with Q. It does not interact with the air track since

there is negligible frictional force between it and the air track. Consequently, the total momentum of P and Q is conserved.

But, suppose experiment E is performed with the holes under vehicle Q blocked, or covered with thin adhesive tape. There is now considerable friction between Q and the track, and P and Q are no longer the only members of an isolated system. When P interacts with Q it is also interacting with the air track through the medium of friction. Moreover, the track interacts with the bench and the bench. . . .

Next, suppose that the air track itself could be arranged to move freely without friction, perhaps by mounting it on special large air pucks. The isolated system now consists of P, Q, and the air track; and when P collides with Q, the initial momentum of P is transferred completely to the final momenta of P, Q, and the now moving air track. The law of conservation of momentum can be applied in this case and indeed to any case in which the members of the isolated system are clearly known.

What about the case of a car colliding with the back of a stationary car? Before impact the brakes would doubtless have been applied and a large frictional force would have acted. Could the law of conservation of momentum be applied here? We believe it could be, so long as all the interacting objects can be specified. In this case there are three involved; the car interacts with the stationary vehicle and also with the road (and therefore the Earth). The moving car does not give all of its momentum to the stationary one; while it is braking it is giving some momentum to the Earth—momentum it lost when the car originally accelerated from rest.

Does the law of conservation of momentum apply to a sprinter as he leaves his starting blocks? He is initially at rest and then suddenly gains some momentum; if the law is applicable an equal but opposite change of momentum must occur. Presumably the Earth undergoes this change and the Earth and the sprinter constitute an isolated system.

Some of these points may be illustrated easily with a clockwork or flywheel toy car and a piece of thin wood or card that can roll on double-ended knitting needles or other suitable rods. If the car, with its wheels already turning, is placed on the card the car will move forwards at the expense of the card moving backwards. If the car then collides with a small block fixed at one end of the card, the changes of motion will be obvious.

5.4 MAKING MEASUREMENTS OF MASS

The conservation of momentum law enables two masses to be compared. Applying Eq. (4.2) to the one-dimensional case,

$$m_X u_X + m_Y u_Y = m_X v_X + m_Y v_Y,$$

where u_x, etc., are magnitudes only. Rearranging,

$$+ m_X v_X - m_X u_X = - (m_Y v_Y - m_Y u_Y) \quad (5.1)$$

thus,

$$\frac{m_X}{m_Y} = - \frac{(v_Y - u_Y)}{(v_X - u_X)} = - \frac{\Delta v_Y}{\Delta v_X}. \quad (5.2)$$

The significance of the negative sign is that, in the case of an explosion, the objects move off in opposite directions, and in the situation where a collision occurs, one body undergoes an increase in velocity while the other suffers a decrease in velocity.

Equation (5.2) relates the masses of two bodies, X and Y, to the change of velocity each undergoes during an interaction. As an example, suppose that two different masses, X and Y, in the form of trolleys are initially at rest, "explode" apart, and that their velocities become $0.16\ \text{m s}^{-1}$ and $-0.96\ \text{m s}^{-1}$ respectively. Using Eq. (5.2),

$$\frac{m_X}{m_Y} = - \frac{(-0.96)}{0.16} = 6,$$

we see that the mass of X is six times the mass of Y.

As another example, consider an air track vehicle P moving at $0.080\ \text{m s}^{-1}$ which falls to $0.050\ \text{m s}^{-1}$ when P collides with vehicle Q. Q was initially at rest and ends up by having a velocity of $0.075\ \text{m s}^{-1}$. Applying Eq. (5.2):

$$\frac{m_P}{m_Q} = - \frac{\Delta v_Q}{\Delta v_P} = - \frac{(0.075 - 0)}{(0.050 - 0.080)}$$

$$= - \frac{0.075}{-0.030} = 2.5.$$

The mass of P is 2.5 times the mass of Q.

5.5 THE UNIT OF MASS

In both examples, we have found the ratio of two masses. So that we can speak of the mass of a particular object, it must be compared with some universally accepted standard mass such as the international prototype kilogramme. This is kept at the Bureau International des Poids et Mesures which is at Sèvres, a suburb of Paris. This is a truly international establishment and is built on land ceded by the French Government and declared to be international territory. The prototype kilogramme is a cylinder, having its height and diameter approximately equal, made of an alloy, 90% platinum and 10% iridium, which has a high density of 21 570 kg m^{-3}. This was adopted as the standard in 1889 and from it substandards have been made. Copy number 18 is held in Great Britain and from time to time it is returned to Paris for checking. Its mass has remained constant to within 1 part in 10^8. The alloy was chosen because it was thought to be the least liable to be affected by time and environment; experience has since shown this to be true. The objection to the use of platinum–iridium for a standard mass is that because of its high density (i.e. small volume), any slight abrasion will change the mass appreciably, consequently the greatest precautions are taken.

5.6 INERTIAL MASS AND GRAVITATIONAL MASS

In the earlier example, if trolley Y had been a copy of a kilogramme on wheels, then we could have stated that the mass of X is 6 kg. Values of mass obtained in this way, by comparing the inertias of two objects, are called inertial masses.

However, when copy number 18 is taken to Paris for checking it certainly does not undergo any interaction with the international prototype kilogramme!

It is common practice to use a pair of scales or some other form of pan balance to compare weights of bodies with a standard weight. In this process the gravitational pull of the earth on the body under test is compared with the gravitational pull of the earth on the standard body. Values of mass found in this way are called gravitational masses.

Now experiments (for example that celebrated experiment which Galileo may or may not have carried out at Pisa) show that a body with a large mass and a body with a small mass which are released together fall to the ground with the same acceleration. It seems that the Earth's gravitational pull on the larger mass is larger than that on the smaller but that the material which has to be accelerated is larger in just the same proportion. The resulting acceleration is the same for both.

> "The pull of the Earth's field is proportional to the amount of matter-to-be-pulled. The mass to be accelerated is proportional to the amount of matter-to-be-accelerated. We have no guarantee—except from the vital experiment which shows the acceleration the same for all—that these two kinds of 'amount of matter' are equal, or proportional, for all different materials."
>
> *Nuffield Physics* (*1966*)

We shall return to this question of the inertial and gravitational masses in Unit Six. Meanwhile we note that the fortunate equivalence of inertial and gravitational mass of a body enables us to compare masses by comparing weights with some form of balance: a vastly more convenient process than that of comparison by inertial effects.

However, the inertia method of measuring mass cannot be forgotten just because it is more difficult to perform. After all, how else could a mass be measured in a space craft? Spring balances and pan balances would read zero except when the rocket motors were on. This raises several questions: Can the inertia method be used in a moving space craft? Does the law of conservation of momentum apply in such a case? What is meant by "weightlessness"? These will all be considered later: for the moment let us examine the situation, not before and after a collision or explosion, but *during* it.

5.7 THE CAUSE OF MOMENTUM CHANGE

Consider Eq. (5.1):

$$m_X v_X - m_X u_X = -(m_Y v_Y - m_Y u_Y).$$

Since momentum is (mass × velocity), this equation in words is:

Final momentum of X − initial momentum of X = − (Final momentum of Y − initial momentum of Y).

Or more simply,

Change of momentum of X = − (Change of momentum of Y).

Or simpler still,

$$\Delta(mv)_X = -\Delta(mv)_Y. \tag{5.3}$$

This equation shows that, during an interaction, each mass undergoes an equal, but opposite, change of momentum. During that interaction each mass exerts a force on the other. It is these forces which produce the changes.

Chapter 6
FORCE

6.1 Acceleration
6.2 The unit of force
6.3 Newton's third law of motion

Is there a simple connection between the force which is applied, the time during which it acts and the change in momentum (mv)? In dealing with three such variables as these, the technique is to keep any one of the three constant and to examine how the other two are related. We will start by keeping the force itself constant.

Experiment J: To apply a constant force to a single trolley and to examine how the momentum changes with time

A constant force can be applied to a trolley by towing it with a thin length of rubber thread (elastic) kept at a constant stretch. With practice it is possible to keep the end of the elastic that is in the fingers, in the plane of the two front pegs of the trolley even though the velocity of the trolley is increasing. See Fig. 6.1.

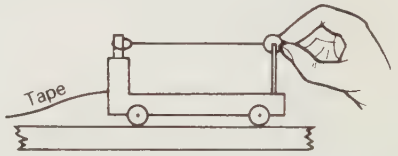


Fig. 6.1

The resulting motion of the trolley can be recorded with a ticker-tape timer and if the tape is cut into 10-space lengths which are stuck side by side on to card, the chart shown in Fig. 6.2 is obtained. The length of each strip gives the mean velocity during that particular period of 0.2 s (10 × 0.02 s). Examination of this chart shows that each step is of equal size which means that equal changes of velocity occur in equal periods of time, when the force causing the change is constant. It follows that the change in

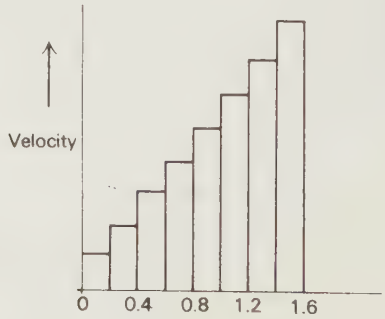


Fig. 6.2

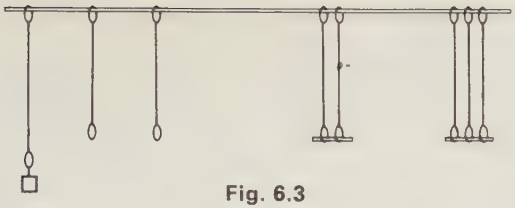


Fig. 6.3

velocity that occurs in, say, 3 s, is three times the velocity change occurring in 1 s. Thus, change of velocity is proportional to the time during which it is changing.

Or $\Delta v \propto \Delta t$ where $\propto$ means *is proportional to*. Since the mass moving (a single trolley) is constant it is also true that

$$\Delta(mv) \propto \Delta t$$

when the force is constant. Consequently the change of momentum is proportional to the length of time for which the force acts.

So far we have investigated momentum and time whilst keeping force constant. We must now apply a different, but nevertheless constant, force to the trolley and see what change of momentum occurs in a given time.

Experiment K: To apply different, constant forces to a trolley

Before doing the actual experiment with the trolley we must check the elastic threads to see how we may use them to give us three different forces. First we must see whether they are equivalent in their performance. We may do this very simply by supporting them side by side so that they hang vertically and then fastening a small mass to the lower end of one of them. That thread will stretch by some definite amount. Transfer the mass to each of the other threads in turn. What do you conclude if all stretch by the same amount?

Suppose next that you were to link two and then three of them together (Fig. 6.3), in parallel. What force would you expect to apply to them in order to stretch them by the same amount as in the first of these experiments?

Now we can be sure that, when they are used as in Fig. 6.1 with two and then three attached to the trolley peg in parallel, two and three *units of force* may be applied (taking the case in experiment J as one *unit of force*).

Tape charts produced as before, show a constant step size for a particular force. As expected the largest force produces the greatest step or change in velocity. Some typical results are shown in Table 6.1.

Table 6.1

Force	Average step size (mm)
1 unit	9.0
2 units	18.2
3 units	26.9

Allowing for experimental errors, this table shows that the step size is proportional to the force. This means that the force is proportional to the change of velocity and therefore to the change of momentum that results. Thus, $\Delta(mv) \propto F$. This can be combined with the result of experiment J by writing:

$$F\Delta t \propto \Delta(mv) .$$

Or, rearranging,

$$F \propto \frac{\Delta(mv)}{\Delta t} . \qquad (6.1)$$

This last relationship can be interpreted as "force is proportional to rate of change of momentum," which is incorporated in Newton's second law of motion: "Rate of change of momentum is proportional to the resultant force causing it, and it takes place in the direction of that force."

6.1 ACCELERATION

In all the cases leading to Eq. (6.1), the mass remained constant while the other quantities varied. Equation (6.1) can therefore be written:

$$F \propto m \frac{\Delta v}{\Delta t} . \qquad (6.2)$$

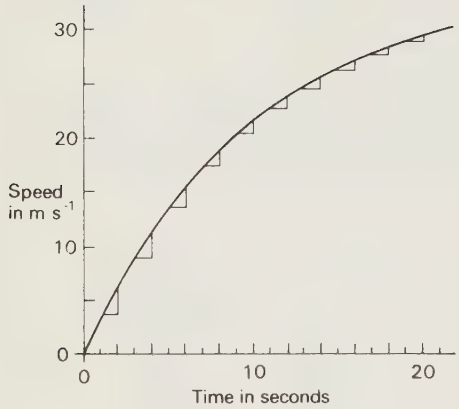


Fig. 6.4 A car starting from rest. Each slope "triangle" has a base, 1 s long and so the vertical side gives the change of velocity in 1 s. This acceleration is not uniform since it decreases as the car goes faster.

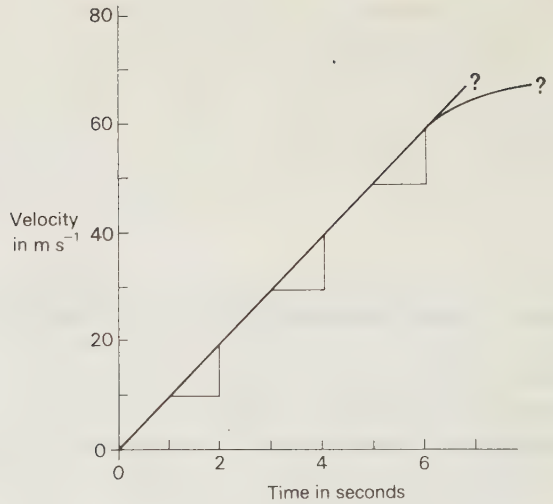


Fig. 6.5 A mass falling freely near the Earth's surface. Each slope "triangle" has the same vertical side and so the acceleration is uniform. It is just under 10 m s^{-2}.

The fraction $\Delta v/\Delta t$ is the *rate of change of velocity*, or *acceleration* (produced by a force acting on a mass). As an acceleration is a velocity divided by a time, its units are m s^{-1}/s or as is usually written, m s^{-2}. One should avoid the popular form metres per second per second since misunderstood, or understood literally, this results in the units $\frac{\text{m}}{\text{s/s}}$ or just metre.

Acceleration is a vector since it involves a vector quantity, velocity, and a scalar factor, time. Consequently the magnitude and the direction of an acceleration must always be specified.

It should be noted that there are cases of interactions when the mass is not constant. A rocket moves forward by interacting with the gas molecules produced when its fuel is burnt. The mass of fuel, and hence the mass of the rocket is continually diminishing.

The acceleration tells us by how much the velocity of a body changes in a second. Information about acceleration can conveniently be expressed on a velocity-time graph. Some examples are given in Figs. 6.4 and 6.5. It is interesting to note that experiments J and K show that a mass moving under the action of a constant force does so with uniform acceleration. The implication of the graph (in Fig. 6.5) is that since the falling mass is accelerating uniformly, then the gravitational force on it is constant and is independent of the velocity of the mass. It should be realized that the acceleration will only be uniform while the velocity of the mass is small enough for the resistance of the air to be negligible. As the air resistance becomes greater, the resultant force accelerating the mass becomes smaller until it is zero. The acceleration is then zero and the mass is falling at its maximum, or terminal velocity. If the graph had been continued further, it would have curved over just as the car's graph. The car is also tending to a maximum velocity when the force that the engine exerts through the wheels is exactly balanced by air resistance, etc.

6.2 THE UNIT OF FORCE

Equation (6.2) stated another way is:

Force is proportional to mass × acceleration

or,

$$F \propto ma. \qquad (6.3)$$

We have not yet discussed the units in which forces are measured. Scientists and engineers need a universal unit of force so that they can communicate useful information between themselves. Equation (6.3) does not lead directly to useful notions about how big forces are. It is like saying that the cost (c) of a bag of eggs is proportional to the number of eggs (n) in it. This information alone does not tell us how much money to hand over to the shopkeeper. So far, we can only write: $c \propto n$, but if we produce an equation: $c = pn$, where p represents the price of one egg, then c can now be expressed in terms of money.

Likewise, Eq. (6.3) can be written without the proportionality sign as: $\boldsymbol{F} = km\boldsymbol{a}$. The value of the constant k will decide the number that will describe the size of the force; k can have any value provided all scientists use the same value. For simplicity, the value chosen is $k = 1$. (One what? Try answering this when the unit of force is named.) The equation now becomes:

$$\boldsymbol{F} = m\boldsymbol{a}. \tag{6.4}$$

We have still not yet defined our unit of force, but we have the means, namely Eq. (6.4), to do so. If we make $m = 1$ kg and $\boldsymbol{a} = 1$ m s^{-2}, then the force is called 1 newton, an appropriate name for it!

One newton is the force that will give a mass of 1 kg an acceleration of 1 m s^{-2}.

A unit such as this one for force is classified as a derived unit, signifying that it is derived from the appropriate basic units. In the Système International d'Unités (S.I.) which was adopted by the Conférence Générale des Poids et Mesures in 1960 these basic units are the metre, the kilogramme, the second, the ampere, the kelvin and the candela. In the case of the newton, the appropriate basic units are the metre, the kilogramme and the second. You will find the international definitions of these and the other basic units, together with other information about the Système International in Appendix 1.

The equation

$$\boldsymbol{F} = m\boldsymbol{a}$$

can now be used for calculating the force acting in a given set of circumstances. For example, if a mass of 4 kg is found to have an acceleration of 3 m s^{-2} the force acting on it is 12 newtons (12 N).

Look at Fig. 6.4. What is the initial acceleration of the car? If the mass of the car is 800 kg, what force are the tyres exerting on the road during this acceleration? What is your estimate of the acceleration at the end of the tenth second? What is the force used to accelerate the car now? Why is this less than before?

Look at Fig. 6.5. The acceleration near the surface is 10 m s^{-2}. What is the force experienced by a body falling with this acceleration if its mass is 9 kg?

6.3 NEWTON'S THIRD LAW OF MOTION

During a collision between two masses they exert forces on one another. What are the magnitudes of these mutual forces?

Equation (6.4) may be written

$$\boldsymbol{F} = m\frac{\Delta \boldsymbol{v}}{\Delta t} \quad \text{or, in the limit} \quad F = m\frac{dv}{dt}$$

$$= \frac{\Delta(m\boldsymbol{v})}{\Delta t} \quad \text{or, in the limit} \quad F = \frac{d(mv)}{dt}, \tag{6.5}$$

$$= \frac{d\boldsymbol{p}}{dt}$$

where $\boldsymbol{p} = m\boldsymbol{v}$, the momentum.

This is an alternative and very useful form of Newton's second law.

Now the law of conservation of momentum has taught us that, in a collision between two bodies of mass m_1 and m_2 moving with velocities $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ each mass undergoes an equal but opposite momentum change

$$\Delta(m_1\boldsymbol{v}_1) = -\Delta(m_2\boldsymbol{v}_2).$$

The time interval during which the collision takes place is the same for the two masses so we may write

$$\frac{\Delta(m_1\boldsymbol{v}_1)}{\Delta t} = -\frac{\Delta(m_2\boldsymbol{v}_2)}{\Delta t}.$$

That is, the rates of change of momentum of the two interacting bodies are equal and opposite. Since,

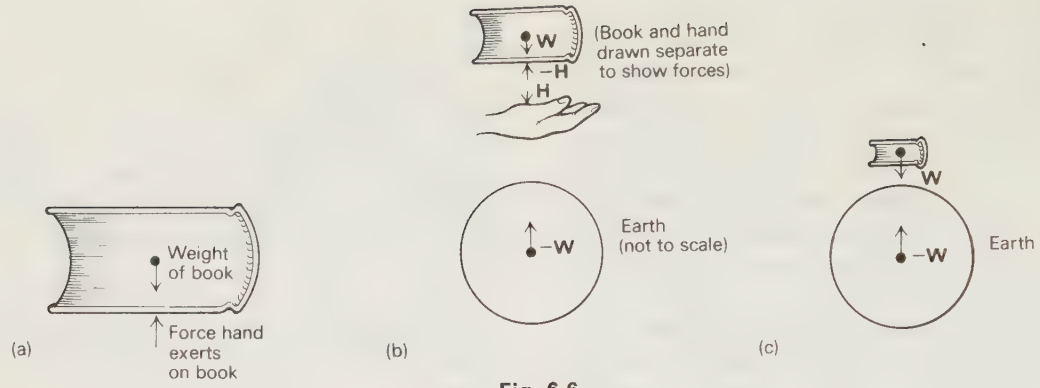


Fig. 6.6

from Eq. (6.5) the rate of change of momentum is a measure of the force acting, we see that, when two bodies interact, the force on one is equal and opposite to the force on the other.

This fact is stated in Newton's third law of motion: "If a force acts on a body, then an equal but opposite force must act upon *another* body." In his *Principia* of 1686, Newton gives examples to explain this law in a *scholium* or explanatory note.

> "If you press a stone with your finger, the finger is also pressed by the stone. If a horse draws a stone tied to a rope, the horse (if I may so say) will be equally drawn back towards the rope; for the distended rope, by the same endeavour to relax or unbend itself, will draw the horse as much towards the stone as it does the stone towards the horse, and will obstruct the progress of the one as much as it advances that of the other."

This seems to be a paradox, for if these forces are equal but opposite, how does the horse and stone ever make any progress? The difficulty is resolved by considering the horse and the stone separately and by dealing with *all* the forces involved.

The horse pushes backwards on the ground which pushes forwards on the horse. (This is a pair of Newton III forces.) If the ground's forward push is at least greater than the backward pull of the rope on the horse, then forward motion can occur. (*Note:* Even if the ground's push and the rope's pull were equal, they would *not* constitute a pair of Newton III forces because they both act on the horse and not on different bodies.)

Consider now the stone; it exerts a forward force on the ground and the ground exerts an equal backward force on the stone. (A pair of Newton III forces.) If this backward force is less than the forward pull of the rope, then forward motion occurs. It should be noted that the ground's frictional force and the rope's pull both act on the stone and do not form a pair of forces.

Newton III can also be stated in this form: "Action and reaction are equal and opposite, and act on different bodies." Often, when this form is quoted, the last five words are unfortunately omitted. This can lead to some confusing and fallacious statements when illustrations of Newton's third law are given. For instance, "If a book is resting on a man's hand, there are two equal and opposite forces acting; the weight of the book acting downwards and the reaction from the hand upwards." In itself this statement is correct but the forces mentioned are not examples of the "action and reaction" forces intended in Newton III. They are examples of forces that happen to be in equilibrium, and both are acting on the book (Fig. 6.6a). (Weight is the force acting *on* a mass placed in a gravitational field.) Apart from the fact that both forces act on the *same* body, Newton III does not apply when the hand is suddenly removed. If the upward hand force becomes zero, the weight certainly does not follow suit! The law has not failed here, it is simply being wrongly used.

To clarify this case, it must be realized that the Earth is pulling on the book with a force $\boldsymbol{W}$ and the book pulls on the Earth with a force $-\boldsymbol{W}$ (which is equal and opposite). These forces constitute a pair of Newton III forces (see Fig. 6.6b).

When the hand is supporting the book, it pushes upwards on the book with a force $-\boldsymbol{H}$ and the book pushes downwards on the hand with a force $\boldsymbol{H}$. These are also a "pair" of forces and both do become zero when the hand is removed. When this is done the pair of forces, $\boldsymbol{W}$ and $-\boldsymbol{W}$, still remain and cause the book and the Earth to fall together or towards one another (Fig. 6.6c). Because of its great mass the Earth is thought to move imperceptibly.

Chapter 7
THE EFFECTS OF FORCES: WORK AND ENERGY

7.1 Work
7.2 The unit of work
7.3 Distance
7.4 Power
7.5 Energy
7.6 Calculations of energy transfer (work done) when a mass is lifted in a gravitational field
7.7 The calculation of the energy transferred when the shape of an elastic body is changed
7.8 The calculation of the energy transferred when the speed of a mass is changed
7.9 The general case
7.10 The quantitative examination of some energy transfers
7.11 The conservation of energy
7.12 Conservative and dissipative forces
7.13 The energy stored in a rotating body
7.14 Energy "book-keeping" during collisions. The classification of collisions
7.15 The kinetic energy remaining after an inelastic collision
7.16 The transfer of kinetic energy resulting from elastic collisions
7.17 The condition for maximum energy transfer in an elastic collision
7.18 A scientific toy

Forces can perform a number of different tasks, for example:

a) Two or more forces can cause a system to be in a state of equilibrium.
Thus, in Fig. 6.6a, if the force exerted by the hand is equal but opposite to the weight of the book, then the book is in equilibrium. Without the hand's force, the book falls.

b) A mass can move at constant velocity if the forces acting on it are balanced.
A force can lift masses vertically against the Earth's gravitational attraction and can move a mass along a surface when frictional forces between mass and surface are acting. We shall see later that the first of these processes results in in the storage of energy and the second in the dissipation of energy.

c) Two or more forces can change the shape of an object.
They can cause a short spring to change into a long one. A force applied to the string of an archer's bow will cause the bow to bend into an arc of smaller radius, and so increase the forces of strain in the bow itself.

d) A force can cause a mass to undergo a change in velocity.
By means of their oars, a rowing crew can cause a force to be exerted on their boat which will move off from rest and gain a velocity.

e) A force can cause rotation.
Applying a force to the handle of a hand-drill causes the drill to rotate.

7.1 WORK

All of these tasks can be considered to be useful by those requiring them to be done. However, they do not all constitute work in the scientist's meaning of the word. In deciding whether or not a task constitutes work, the test we may make is to enquire if *fuel* has been used. In applying the test, it is necessary to replace any human participant by a mechanical device and then to see if such a device needs a continual supply of fuel to make it perform the task.

In (a) the hand can be replaced by a shelf—which does *not* require a continuous supply of fuel.

In (b) the mass could be the load lifted by a crane that is powered by an electric motor or diesel engine. The mass could be a lawnmower and this needs fuel whether the force causing it to move is supplied by a man or a petrol engine.

In (c) the tasks could all be performed by an electric motor or petrol engine—fuel is therefore needed.

The crew in example (d) could dispense with their oars and use an outboard motor instead—causing a change of velocity is a fuel-using task.

The drill of example (e) could be made to rotate by an electric motor supplied with electricity produced by burning a fuel.

Of all these examples, only (a) does not come into the category of work. As another example of a task that is not work in the scientific sense, consider Jack, whose job it is to hold pieces of wood while another man saws them. Jack can do this all day, get tired, and be paid well; but he has not done work, for he could be replaced by a clamp or vice that requires no continuous input of fuel. Of course, Jack is also likely to be required to lift and carry the wood, and this *is* work.

We can now say what work is. "Work is done when a fuel-using task is performed."

How is work to be measured, and on what factors does the amount of work done depend? Distance is one factor, because a feature that distinguishes task (a) from (b), (c), (d), and (e), is that when work is done, the force involved moves with the object upon which it is acting. To examine the relationship between work and distance, consider a crane that is to lift a load of bricks weighing 9000 N, from the ground to the top of a building, 20 metres high. While the load is moving steadily upwards, the crane is exerting a force of 9000 N in the direction the load is moving. (Of course, to accelerate the load initially to this steady velocity, the crane must exert a force greater than 9000 N.) Suppose the load is first lifted from the ground through only 10 m; a certain quantity of fuel, petrol for example, will have been used. If the load is then lifted from 10 m to 20 m, an equal quantity of petrol will have to be used. This quantity is half that required to lift the load through the whole 20 m in one go. Thus, the fuel used, or the work done, is proportional to the distance the force has moved along the direction in which it is acting.

Consider the effect of force (or load) on the work done. Suppose a double load of bricks has to be lifted and the crane is not capable of providing a force of 18000 N. There are two possibilities: either to separate the bricks into two equal loads of 9000 N each and to lift first one half through 20 m and then raise the other half, or, to employ two identical cranes and hook them both on to the 18000 N load so that each crane is effectively lifting 9000 N. In either case, the total fuel used, or the work done, to lift 18000 N through 20 m is twice that required to raise 9000 N through the same distance. Thus, work done is also proportional to the load lifted, or the force applied.

Distance and force both have a hand in determining the quantity of work that is done in a particular situation when a force moves its point of application along its line of action. This result is combined in the relationship:

$$\text{Work} \propto (\text{magnitude of force}) \times \begin{pmatrix}\text{Distance moved along} \\ \text{the direction of force}\end{pmatrix}$$

or in symbols:

$$W \propto Fs. \tag{7.1}$$

7.2 THE UNIT OF WORK

In discussing the unit of force we encountered an equation (Eq. 6.3) which was similar in form to this one. Using the same procedure as before this equation for work may be written

$$W = k'Fs.$$

The value of the constant k' will decide the number that describes the work done in a given situation. Again, k', can have any value provided we all agree to use it. The value chosen is $k' = 1$, and so the equation becomes:

$$W = Fs, \tag{7.2}$$

where, $W \equiv$ work done, $F \equiv$ magnitude of force

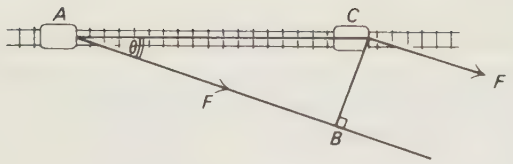


Fig. 7.1

acting, $s \equiv$ distance moved along the direction of the force.

This equation allows the unit of work to be defined. If the force is 1 N, and $s = 1$ m, then the work done is 1 newton metre to which the name 1 joule is given.

One joule is the work done when a force of 1 newton moves its point of application through 1 metre in the direction of the force.

Equation (7.2) can now be used for calculating work. For example, when bricks weighing 9000 N are lifted through 20 m, then 180 000 J of work have been done.

7.3 DISTANCE

Whenever distance has been mentioned in this section on work, it has always been "distance along the direction of the force," and there is another way of considering this point. Suppose a truck is on rails, and a man pulls it along using a rope inclined at an angle θ to the rails (see Fig. 7.1). The truck is shown to have moved a distance AC along the rails. However, in the direction of the force it has been displaced a distance AB. Consequently, the work done $= F\,AB = F\,AC \cos\theta$. This last expression can be rearranged as $AC\,F \cos\theta$. $F \cos\theta$ is the component of F in the direction of motion, and so an alternative way of defining the symbols of Eq. (7.2) is:

$W \equiv$ work done, $F \equiv$ component of force in the direction of movement, $s \equiv$ distance moved.

7.4 POWER

In all of the discussion so far, no mention has been made of the time needed to perform a piece of work. In fact, whether the crane lifts the bricks through 20 metre in 10 seconds or 10 hours, it still does the same 180 000 J of work. It is necessary to consider time if we are discussing power, or the *rate* at which work is done. In the faster case above, the power is 18 000 J s^{-1}. The J s^{-1}, the unit of power, is named the watt. In the (unlikely!) slower case above, the power is only

$$\frac{180\,000}{10 \times 60 \times 60} = 5 \text{ W}.$$

The watt is the rate of working of 1 joule per second.

7.5 ENERGY

We are all familiar with this word "energy"; so much so that we tend to overlook the fact that none of us have ever seen energy and consequently find it impossible to describe. We can easily recognize the existence or presence of energy, not by observing energy directly but by seeing what it can do. If a man has plenty of energy he can do the job of cutting a lawn by pushing a mower. A man with less energy, or perhaps a bigger lawn, will use the energy *stored* in petrol, to do the work. Energy is often associated with the performance of work because the more work that is done, the greater the quantity of energy needed. Here are some attempts to define energy:

"Energy—the go of things" (J. C. Maxwell).
"Energy is the capacity for doing work."
"Energy is that which is changed from one form to another when work is done."
"Energy is a measure of the price of mass in motion."

Where can energy be found? Since energy and work are associated, and work involves the use of fuel, we must first look to fuels—food, coal, oil, etc., for our store of energy. It may not be generally realized that food is as much a fuel for a living being as coal, which must be fed to a boiler, is for a steam turbine, or petrol is for a car engine. When horses were much used on farms for ploughing and similar heavy work, it was well known that they ate more oats during such periods of high energy output. You also feel hungry after climbing a mountain, or digging the garden.

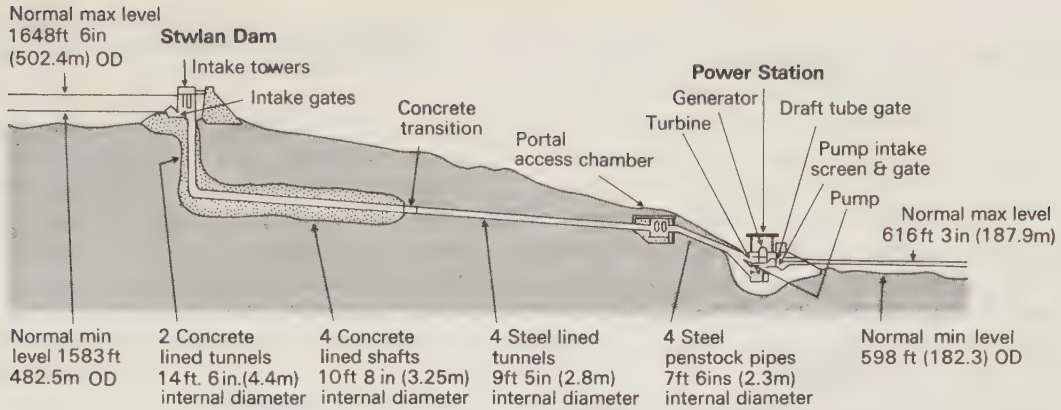


Fig. 7.2 The pumped storage system at the Ffestiniog power station. (Courtesy of Central Electricity Generating Board.)

Energy, once transformed from its fuel form by living beings or engines, can be stored in several ways, for example,

a) by changing the position of a mass in a gravitational field,
b) by changing the shape of an elastic body,
c) by changing the motion of a mass.

In considering examples to illustrate these stores of energy we will see that energy is being continually transformed from one form to another. The work which is done in these transformations of energy from one form to another form is a measure of the energy so transformed. There are also other methods of storing energy that involve electric or magnetic fields, or chemical changes, but only the mechanical examples listed above are under consideration in this section.

An early use of a mass as a store of energy is to be found in grandfather, or long-case, clocks. Here, energy from a man is used to raise a large mass inside the clock. The mass now has more energy than formerly and as it falls slowly (it usually reaches the bottom of the case in 8 days) the chain supporting it causes the gears and the pendulum to operate the clock hands.

Recently mass and the Earth's gravitational field have been used to store, indirectly, electrical energy. The problem is that electricity must be used as it is being produced; there is no way of storing large quantities of electrical energy. But there is a need for more than can be produced during certain periods of the day (peak periods) when many factories are operating, and there is a surplus of generating capacity during other periods and especially at night. At the Ffestiniog power station in Wales, water from one reservoir is pumped into a second reservoir 300 m higher up a mountain during the off-peak periods when the demand for electrical energy is low. During the peak periods of heavy demand for electrical energy, the stored water is allowed to flow through four tunnels about 3 m in diameter to the same power station from which it started. But this time the current of water is utilized to drive the four 90 Megawatt generator/motor units. So electrical energy is produced at a rate of 360 MW and transferred to the country's national power line network (see Fig. 7.2).

It should be noted that in this instance the work which is done by the pumps in lifting the mass of the water against the force of gravity of the earth can, to a large extent, be recovered. We may, with justice, say that energy has been stored in the system.

Many clocks and watches provide further examples of energy storage. In this case the energy is usually stored in a spiral spring whose shape is changed in the process of winding. As the spring slowly returns to its unwound shape, a larger spiral, the energy stored within it is used to operate the trains of gears, the escapement and the hands.

Again, an archer, drawing on his bowstring is transferring some of his own fuel energy to energy which is stored by bending the material of the bow into an arc of smaller radius. While the bow is in this shape that energy is potentially available; it is transferred to other forms when the bowstring is released.

As the word *potential* means "existing in possibility," the term potential energy is used to describe what is stored when masses are lifted in the earth's gravitational field and springs are stretched or bent.

Potential energy (P.E.) is the energy a body possesses by virtue of its position.

The word "position" in this definition implies something more than the place occupied by a mass. It refers also to the state of strain of an elastic body, to the energy situation in a dry battery and so on.

When the archer of the last example releases his bowstring the potential energy of the bow is transferred mostly to the arrow. The arrow now has motion and is said to possess kinetic energy.

Kinetic energy (K.E.) is the energy a mass possesses by virtue of its motion.

As we shall see in Unit 6, a spacecraft en route to the Moon is moving in the combined force field of the Earth and the Moon, and its contents exhibit apparent weightlessness for much of the journey. Between the Earth and the Moon there is a *neutral* point at which the gravitational forces acting on a mass due to the two bodies are equal and opposite. At this point a mass is truly weightless, but the position would be an unstable one. Such a mass has only to move infinitesimally towards the Earth and it will continue falling. It has only to move infinitesimally towards the Moon and it will continue falling. This point is at a distance of about 342 000 km from the Earth. So we see that a spacecraft returning from a lunar mission has fallen a great distance by the time it reaches the Earth's atmosphere. It therefore transfers a very large amount of potential energy to kinetic energy and reaches a very high velocity. This kinetic energy is largely transferred to heat on re-entry and results in a rise in the temperature of the heat shield. On the microscopic scale, extra energy is given to the molecules which make up the shield when its temperature is raised. This energy is, however, highly disorganized.

7.6 CALCULATION OF ENERGY TRANSFER (WORK DONE) WHEN A MASS IS LIFTED IN A GRAVITATIONAL FIELD

Consider a mass m which exists in a gravitational force field such as we find near the Earth or the Moon. Let the magnitude of the force exerted on it be g.

Then the total force acting on the mass due to the gravitational attraction is mg. If now that mass is lifted slowly through h in the direction of the force.

$$\text{energy transfer} = \text{force} \times \text{distance moved in direction of the force},$$
$$= mgh.$$

This is the potential energy transferred to the mass,

$$\text{potential energy} = mgh.$$

Problem 1 The Earth's gravitational field exerts a force of nearly 10 N kg^{-1} at the surface. How much potential energy is stored in the Ffestiniog pumped storage system when the upper reservoir contains 2×10^9 kg of available water? The head of water is 300 m. How long would this operate the four generators at their full output of 90 MW each?

The potential energy of the water $= mgh$

$= 2 \times 10^9 \times 10 \times 300$ J

$= 6 \times 10^{12}$ J.

$$\text{Power or rate of working} = \frac{\text{total energy output}}{\text{time}}$$

$$360 \times 10^6 = \frac{6 \times 10^{12}}{t}$$

$$t = 1.7 \times 10^4 \text{ s}.$$

Graphical treatment

A useful technique for calculating transfers of energy results from plotting a graph of force involved against

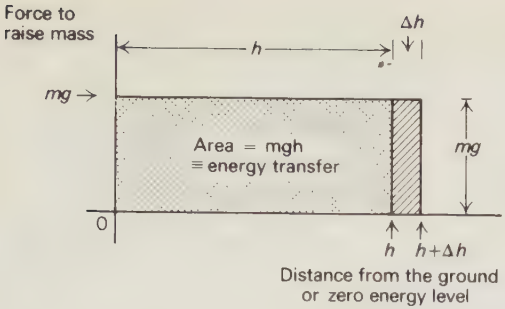


Fig. 7.3

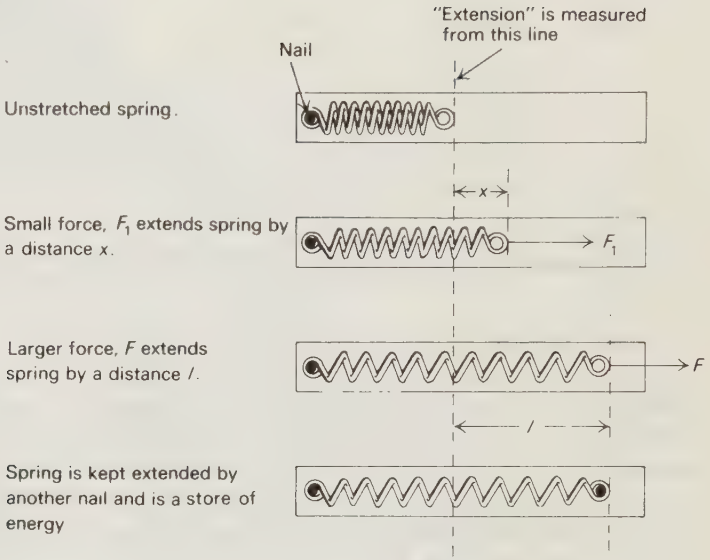


Fig. 7.4

the distance moved. If the force is constant, the graph is a line parallel to the distance axis.

The energy transferred can be calculated by finding the area under the line of a force-distance graph since one side of the rectangle is the force, mg, and the other side is the distance, h.

If the mass is raised a further distance (Δh), that is from a height of h to ($h + \Delta h$), then the extra energy transferred is $mg(\Delta h)$. This is shown on the graph (Fig. 7.3) as the area of a narrow strip. This graphical technique is used in the two examples that follow.

7.7 THE CALCULATION OF THE ENERGY TRANSFERRED WHEN THE SHAPE OF AN ELASTIC BODY IS CHANGED

The difference between this and the previous case is that the force required to do the work is not constant since it depends upon how much the shape is changed. The simplest case to consider is that of a helical spring because if it is used so that it obeys Hooke's law, then we know that the force is directly proportional to the extension of the spring beyond its unstretched length. Figure 7.4 shows such a spring, hooked on to a nail, being stretched along a strip of wood.

If Hooke's law is obeyed, a graph of force against extension will be as shown in Fig. 7.5(a) and 7.5(b).

The energy transferred to the spring is not Fl as the force is clearly not constant throughout the distance l. To calculate the energy transferred, and stored in the spring, a large number of small steps must be taken; during each one, the change of extension, Δx, is so small that it has negligible effect on the force acting which may then be considered

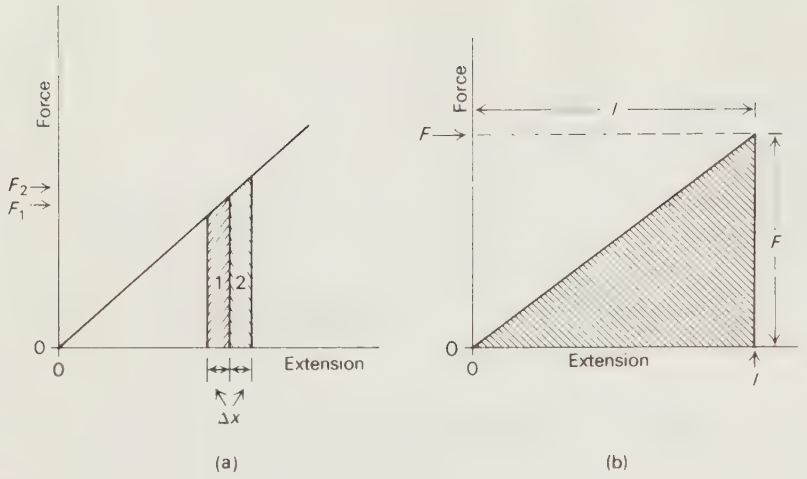


Fig. 7.5

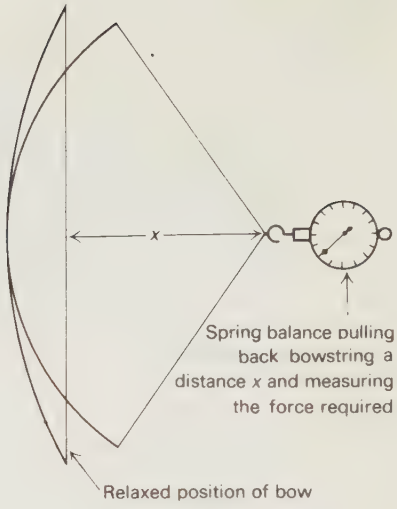


Fig. 7.6

constant. If the force is F_1 while the spring changes length by Δx, then the energy transferred is $F_1 \Delta x$, which is the area of strip 1 in Fig. 7.5(a). If a further extension of Δx is made during which the force is larger but significantly constant at F_2, then the extra energy transferred is $F_2 \Delta x$, which is the area of strip 2. The total energy transferred when the extension changes from zero to l is given by the total of the areas of all the strips involved, which is the same as the area under the line shown shaded in Fig. 7.5(b). This area is half that of the rectangle having sides of length F and l. Thus energy stored $= \frac{1}{2}Fl$.

The area method can be extended to the calculation of energy transferred when the force involved does not follow a simple rule like Hooke's law. An example of this occurs with the archer's bow. If the experiment indicated in Fig. 7.6 is performed, a graph can be plotted (Fig. 7.7) of the force required to pull the bowstring back a distance x. This is no longer a linear graph but the energy transferred to the bow can still be determined by dividing the area under the curve into strips and by finding the total area as shown in Fig. 7.7.

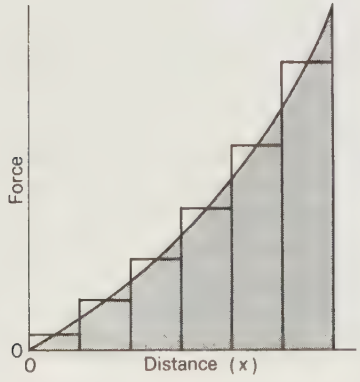


Fig. 7.7

7.8 THE CALCULATION OF THE ENERGY TRANSFERRED WHEN THE SPEED OF A MASS IS CHANGED

Before considering a general case, we will consider a simpler situation in which a mass increases its kinetic energy because it is acted upon by a constant force and therefore undergoes uniform acceleration. The general case will deal with the instance when a varying force acts.

In Fig. 7.8 the gain of kinetic energy comes from the work performed by the force while it is acting.

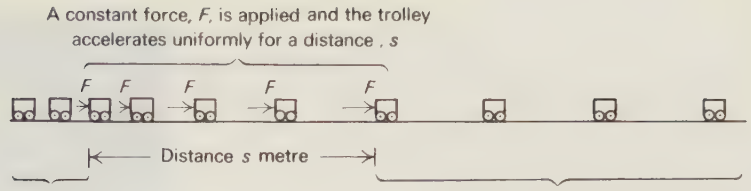


Fig. 7.8

Gain of energy

$= (\text{force}) \times (\text{distance in the direction of the force})$

$= Fs$

$= mas$ (because, $F = ma$)

$= \dfrac{m(v - u)s}{t}$ $\left(\text{because, } a = \dfrac{\Delta v}{\Delta t}\right)$

$= \dfrac{m(v - u)}{t}\dfrac{(v + u)t}{2}$ (because, distance = average velocity × time when acceleration is uniform)

$= \frac{1}{2}m(v^2 - u^2)$

$= \frac{1}{2}mv^2 - \frac{1}{2}mu^2,$ (7.3)

where m, u, v, and a are the mass, initial speed, final speed, and acceleration of the trolley and t is the time for which the force acts.

As this represents a gain or change of energy, $\frac{1}{2}mv^2$ is the final kinetic energy, and $\frac{1}{2}mu^2$ is the initial kinetic energy.

Note: We often say that the kinetic energy of a mass, m, travelling with speed, v, is $\frac{1}{2}mv^2$. This is because when the mass is brought to rest from a speed v, it transforms $\frac{1}{2}mv^2$ joules of energy from kinetic to other forms.

7.9 THE GENERAL CASE

A mass increases its kinetic energy when the force causing this to happen is not constant, as, for example, when a bow transfers its energy to an arrow.

Consider the trolley of the previous case again and let the force that acts vary with distance as the graph, Fig. 7.9 shows. Although the force is not constant, we can take a number of steps chosen to be so small that the force may be assumed to be constant during each step. Each of the steps or strips shown is Δs wide and for strip (1) the force is F_1, while for strip (2) the force is F_2, etc. While F_1 acts, the speed of the trolley changes from v_0 to v_1, and while F_2 acts, the speed of the trolley changes from v_1 to v_2, etc. While the last strip's force, F_n, acts the speed changes to the final or maximum speed, v_n, from the previous speed, v_{n-1}. Equation (7.3) can now be used to calculate the energy gained during each strip, and by addition, the total energy gained may be found, as shown in Table 7.1. If you inspect the last column you will see how, in addition, the terms cancel in pairs.

Table 7.1

For strip number:	The work done by the force =	The energy transferred to the mass =
(1)	$F_1\,\Delta s$	$\frac{1}{2}mv_1^2 - \frac{1}{2}mv_0^2$
(2)	$F_2\,\Delta s$	$\frac{1}{2}mv_2^2 - \frac{1}{2}mv_1^2$
(3)	$F_3\,\Delta s$	$\frac{1}{2}mv_3^2 - \frac{1}{2}mv_2^2$
etc.	etc.	etc.
$(n - 1)$	$F_{n-1}\,\Delta s$	$\frac{1}{2}mv_{n-1}^2 - \frac{1}{2}mv_{n-2}^2$
(n)	$F_n\,\Delta s$	$\frac{1}{2}mv_n^2 - \frac{1}{2}mv_{n-1}^2$

The total energy gained by the mass

= total of the final column

$= \frac{1}{2}mv_n^2 - \frac{1}{2}mv_0^2.$

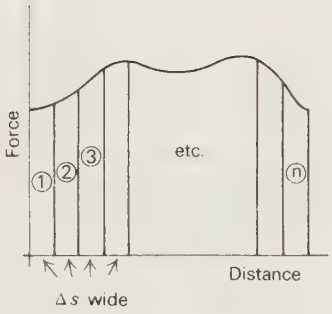


Fig. 7.9

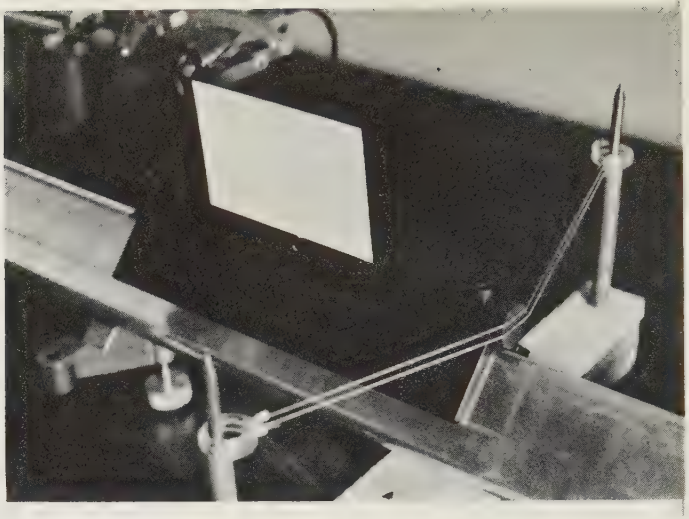


Fig. 7.10 Investigation of energy transfer to a linear air-track vehicle.

Consequently, whether a mass gains its kinetic energy by means of a constant force, or a varying one, the initial and final speeds and the mass moving, are the factors that determine the change in energy.

7.10 THE QUANTITATIVE EXAMINATION OF SOME ENERGY TRANSFERS

a) The energy stored in rubber, used as a catapult, can be used to test the equation K.E. $= \frac{1}{2}mv^2$. The arrangement shown in Fig. 7.10 is a linear air track with the usual arrangement of photocell and scaler to measure the speed of the vehicle once it has been catapulted along the track. Trolley elastic threads, stretched between two vertical posts, act as the catapult. Five elastic threads must be selected so that each one, used separately at the same stretch, will give the vehicle the same speed. By using one, or a number of elastic threads, the vehicle can be given 1, 2, 3, 4, or 5, "*unit doses*" of energy. Typical results are shown in Table 7.2. Inspection of these results, or a graph of energy against v^2 which is a straight line through the origin, show that energy transferred to kinetic energy is proportional to v^2.

Similar experiments performed with a half-length vehicle, and combinations of vehicles, show that kinetic energy is also proportional to the mass in motion.

Table 7.2

Number of elastic threads (Energy)	Transit time (s)	Speed v (m s^{-1})	v^2
1	0.189	0.53	0.28
2	0.133	0.75	0.56
3	0.108	0.93	0.86
4	0.095	1.05	1.10
5	0.085	1.18	1.40

It is instructive to measure the actual energy stored in the catapult and to see if it is all converted to kinetic energy in the vehicle. To do this a force-distance graph for an elastic thread must be plotted, and to obtain data for this, small masses are hung from its centre as shown in Fig. 7.11 and for each one, the displacement, d, of the centre is measured. A set of results obtained with one of the elastic threads used above is given in Table 7.3.

You should plot a graph from these results and find the energy stored when $d = 0.033$ m, using the area method described earlier.

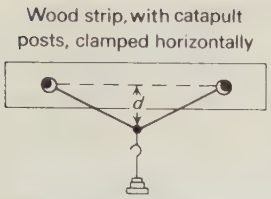


Fig. 7.11

Table 7.3

Force displacing middle of elastic thread (newton)	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Displacement, d (metre)	0.01	0.017	0.023	0.029	0.034	0.038	0.042

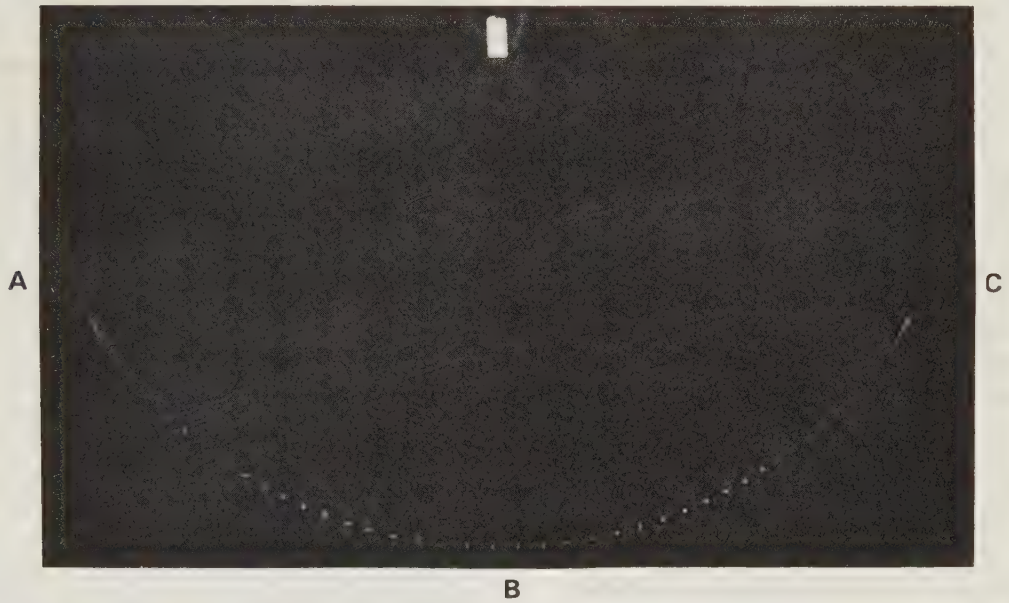


Fig. 7.12 Multiflash photograph of a swinging pendulum. The flashing rate was 100 Hz. The pendulum was 0.325 m long. The photograph is printed at one fifth of full scale. (Photograph by B.T.)

This energy should be compared with the kinetic energy of the vehicle which has a mass of 0.048 kg and speed 0.53 m s^{-1} when one elastic thread was used with a displacement, d, of 0.033 m.

b) We can also examine the energy transfers that occur when a pendulum swings. The strobe photograph, Fig. 7.12, is of a pendulum that swings from rest, at A, across to a rest position at C. As can be seen, it has a maximum speed at the lowest point B. When it is at A or C, the mass has more potential energy than when it is at B. In going from A to B, there is a transfer of energy from the potential form to the kinetic form. The mass of the pendulum was 0.07 kg, and the flashing rate of the strobe was 100 flashes per second. By taking measurements off the photograph, which is printed one fifth full scale, and so determining the distance h "fallen" and the maximum speed, the energy transferred from potential to kinetic and vice versa can be calculated. You should find close agreement between the two. The pendulum was 0.325 m long.

Consequently, energy is continually transferred from one form to another without loss apart from that caused by the air resistance. That will eventually

claim all the original potential energy and the pendulum will come to rest at *B*.

The pendulum bob was a shiny steel ball and the bright points of light recorded in the photograph are reflections of the xenon strobe lamp in the spherical surface.

c) Again, we can examine the energy transfers occurring when a brick falls. If a brick is allowed to fall through a known distance, its speed can be measured by allowing it to interrupt the light beam falling on to the photocell controlling a scaler-timer. The energy transferred from the potential form (mgh) to the kinetic form ($\frac{1}{2}mv^2$) can be compared.

7.11 THE CONSERVATION OF ENERGY

In each of these cases, and within the limits of experimental error, the prediction of Eq. (7.3) is confirmed. That equation was a direct consequence of Newton's second law and we have interpreted it in terms of the interchange of potential and kinetic energies during the experiments. It is tempting to speculate whether the balance is disturbed when other forms of energy are involved. In experiment (c) for example, what happens if the brick hits the floor? The kinetic energy disappears. Does any other form of energy take its place? Certainly sound is produced; so is heat since the temperatures of the brick fragments and the floor both rise slightly as a result of the collision.

As we shall see in Unit 3, during the nineteenth century this problem was studied intensively. J. P. Joule and many other experimenters performed many different experiments in which a given quantity of energy was completely converted into heat. In each case, the same quantity of heat was produced. The French engineer, G. A. Hirn, explored the same phenomenon but in the reverse direction. He supplied a steam engine with heat, accounting very carefully for losses, and measured the useful energy which resulted. All these researches showed that if heat is included in the book-keeping with the other identified forms of energy then a law of conservation of energy could be applied, i.e., within a system, the total energy remains constant as it changes from one form to another.

7.12 CONSERVATIVE AND DISSIPATIVE FORCES

A conservative force is one that permits the storage of P.E. whereas forces such as those resulting from friction, viscosity, etc., are known as dissipative forces.

Conservative forces can be provided by springs and also by such fields as gravitational, electric and magnetic. In these cases, a kinetic energy can be converted to a stored potential energy, e.g., a ball thrown upwards has a kinetic energy that decreases as the potential energy rises. The stored potential energy can be retrieved and converted back directly to its original form—as the ball falls back to the ground.

Work done against a dissipative force, such as friction, causes a temperature rise; this energy is not destroyed or lost for ever. However, it is not directly recoverable as is the energy stored in a spring. To retrieve the energy converted to heat by a dissipative force, it would be necessary to employ some form of engine that changes heat to mechanical energy.

The brakes on a car dissipate unwanted kinetic energy when it is brought to rest from speed. Brakes provide a frictional force against which work is done resulting in a temperature rise. Some kinetic energy is also acquired by the Earth itself but this is an infinitesimal part of the whole. Every time a car is stopped from a speed of 30 m s^{-1} enough heat is produced in the brakes, often in a few seconds, to boil a kilogramme of water. Unfortunately, this energy cannot be used to accelerate the car back to the original speed. The price of stopping the car is paid for in the petrol used when it accelerates again.

7.13 THE ENERGY STORED IN A ROTATING BODY

The flywheel inside a toy flywheel car possesses enough kinetic energy, when it is rotating, to make the car travel up a sloping board. (Such toys are named by manufacturers friction-powered cars. This is a classic example of a misused scientific term. How could friction possibly drive anything? It could certainly raise the temperature of a body.)

A flywheel rotates about an axis passing through its centre of mass. Consequently, although the wheel

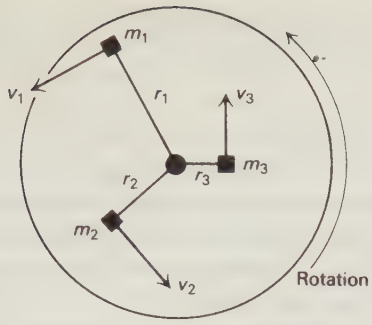


Fig. 7.13

may be rotating, the centre of mass does not have a speed. In what sense can the equation, K.E. $= \frac{1}{2}mv^2$ be applied here since v is zero?

The flywheel must be considered, not as a whole, but as being composed of a large number of masses, m_1, m_2, m_3, etc., distributed at various distances, r_1, r_2, r_3, etc., from the axis (Fig. 7.13). Each mass has a speed, v_1, v_2, v_3, etc., which depends upon its distance from the axis. The kinetic energy of each mass can therefore be determined. The total of the energies stored in all the "elemental masses" that make up the flywheel, gives the energy stored by the flywheel itself.

Let the flywheel rotate at n rev s^{-1}. Since there are 2π radians in 360° or one revolution, then the angular velocity of the flywheel, $\omega = 2\pi n$ radians s^{-1}. Because in general, speed, $v = r\omega$, the speeds of the masses composing the flywheel $= r_1\omega$, $r_2\omega$, $r_3\omega$, etc. The total kinetic energy of the flywheel

$$= \tfrac{1}{2}m_1v_1^2 + \tfrac{1}{2}m_2v_2^2 + \tfrac{1}{2}m_3v_3^2 + \text{etc.}$$

$$= \tfrac{1}{2}m_1r_1^2\omega^2 + \tfrac{1}{2}m_2r_2^2\omega^2 + \tfrac{1}{2}m_3r_3^2\omega^2 + \text{etc.}$$

$$= \sum \tfrac{1}{2}mr^2\omega^2 = \tfrac{1}{2}\omega^2 \sum mr^2.$$

$\sum mr^2$ is shorthand for the sum of all the terms like mr^2 and is called moment of inertia, for which the symbol is I. Thus the kinetic energy of a rotating body is $\frac{1}{2}I\omega^2$.

Since $\sum mr^2$ depends not only upon mass, but also on its distribution, it follows that a flywheel that has most of its mass concentrated in the rim will possess more kinetic energy than a flat disc having the same mass and radius, and rotating at the saem speed. In the case of a flat disc of mass M and radius r, it may be shown that the moment of inertia about an axis through the centre and perpendicular to the plane is given by: $I = \frac{1}{2}Mr^2$. Therefore, the kinetic energy stored in such a disc that is rotating at n rev s^{-1} $= \frac{1}{2}\frac{1}{2}Mr^2(2\pi n)^2 = Mr^2\pi^2n^2$.

The moment of inertia of a rotating body plays the same role in rotational motion as mass plays in linear motion. Where in linear motion we have mass, distance, velocity, force, Newton's second law in the form force $= \Delta(mv)/\Delta t$, momentum (mv) and momentum conservation, etc., so in rotational motion we have moment of inertia, angle, angular velocity, torque (which is the product [force × arm of force]), Newton's second law in the form torque = Δ(angular momentum)/Δt, angular momentum (i.e., momentum × arm) and the conservation of angular momentum, etc.

Rotation (both revolution as in the case of the moon rotating around the Earth and spin as in the case of the Earth spinning on its axis) is of fundamental importance in the fields of astro-physics and of atomic and nuclear physics. Perhaps the most important concept is that of the conservation of angular momentum—an isolated spinning body cannot change its angular momentum.

Stand on a freely moving rotating table holding a heavy book in each hand. Get someone to start you spinning. What happens to your angular velocity as you raise and lower your arms? What is happening to your moment of inertia as you do this?

Problem 2 A solid cylinder of mass M and radius R rolls from rest down an inclined plane without slipping (Fig. 7.14). Find its speed (linear) when it reaches the bottom.

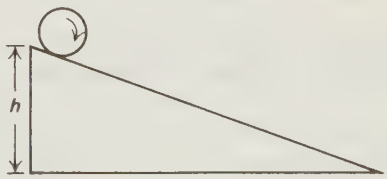


Fig. 7.14

As it rolls down the plane potential energy Mgh is converted to kinetic energy both linear and rotational.

$$Mgh = \tfrac{1}{2}I\omega^2 + \tfrac{1}{2}Mv^2,$$

where h is the height of the plane, v is the linear speed at the bottom and ω is the angular speed at the bottom.

Now

$$I = \tfrac{1}{2}Mr^2$$

and

$$\omega = \frac{v}{r}.$$

So

$$Mgh = \tfrac{1}{2}(\tfrac{1}{2}Mr^2)\frac{v^2}{r^2} + \tfrac{1}{2}Mv^2$$

$$= \tfrac{3}{4}Mv^2$$

and

$$v = \sqrt{\tfrac{4}{3}gh}.$$

What would the linear velocity be if this cylinder had slid, without rolling down the slope?

The velocity is given by

$$Mgh = \tfrac{1}{2}Mv^2$$

so that

$$v = \sqrt{2gh}.$$

The speed of the sliding cylinder is, therefore, greater than that of the rolling cylinder for the latter has its total energy distributed between rotational and linear kinetic energy.

7.14 ENERGY "BOOK-KEEPING" DURING COLLISIONS. THE CLASSIFICATION OF COLLISIONS

a) Inelastic collisions (or perfectly inelastic)

In these, there is no relative velocity after the interaction as the colliding masses stick together. An example of this type is experiment E. Another example is the case of a lump of putty that falls on to a table and does not bounce.

Kinetic energy is not conserved as dissipative forces act during the interaction.

b) Partly elastic collisions

After such an event, the participating masses do separate. The relative velocity after the interaction is reversed in direction but is smaller than the relative velocity initially. Examples are: experiment D, experiment F, and experiment I.

Consider experiment F: relative velocity before interaction $= 0.240 - (-0.150) = 0.390\ \text{m s}^{-1}$, relative velocity after interaction $= -0.120 - (0.210) = -0.330\ \text{m s}^{-1}$ (the $-$ sign indicates reversal of direction).

Another good example is provided by a rubber ball bouncing on the ground. The relative velocity between ball and ground is reversed during the interaction and is reduced in magnitude since the ball does not bounce back up to its original height.

Dissipative forces act during the interaction and kinetic energy is not conserved.

c) Elastic collisions (or perfectly elastic)

The relative velocity after the interaction is reversed in direction and is *equal* in magnitude to the relative velocity initially. Examples are: experiment B and experiment C. Consider experiment C: relative velocity before interaction $= 0.55 - 0.00 = 0.55\ \text{m s}^{-1}$, relative velocity after interaction $= 0.00 - 0.55 = -0.55\ \text{m s}^{-1}$. (Again the $-$ sign indicates reversal of direction.)

Another example that makes the reversal of relative velocity particularly clear is provided by an air track vehicle that hits a fixed rubber band buffer at one end of the track.

In an elastic interaction, conservative forces act and kinetic energy *is* conserved as far as the initial and final conditions are concerned.

d) Hyperelastic collisions

The relative velocity after this interaction is reversed in direction and is greater than the relative velocity initially. Kinetic energy is actually gained during the

interaction. Such a case can be contrived by arranging an explosion to occur during a collision. For example the spring plunger of a trolley could be released by contact with another trolley and the potential energy stored in the spring is then transferred to extra kinetic energy of the trolleys. However, as we shall see in Unit 4 such occurrences do take place in nature. Should, for example, two vibrating molecules collide they may show an increase of translational kinetic energy but at the expense of their vibrational energy.

7.15 THE KINETIC ENERGY REMAINING AFTER AN INELASTIC COLLISION

How much remains after such a collision? Does the kinetic energy dissipated depend on the masses involved? Can all of the original kinetic energy be lost?

Consider a mass m_1 moving with velocity u and interacting inelastically with mass m_2 which is at rest. Their final, common, velocity is v. The conservation of momentum rule applies, thus,

$$m_1 u = (m_1 + m_2)\, v$$

$$v = \frac{m_1 u}{(m_1 + m_2)}.$$

Kinetic energy is *not* conserved as dissipative forces act during the collision.

K.E. remaining

$$= \tfrac{1}{2}(m_1 + m_2)\, v^2 = \tfrac{1}{2}(m_1 + m_2)\, \frac{m_1^2 u^2}{(m_1 + m_2)^2}$$

$$= \tfrac{1}{2} m_1 u^2 \, \frac{m_1}{(m_1 + m_2)}$$

$$= \tfrac{1}{2} m_1 u^2 \left[\frac{1}{1 + (m_2/m_1)} \right]. \qquad (7.4)$$

In this equation $\frac{1}{2} m_1 u^2$ is the original K.E. of mass m_1 and so

$$\left[\frac{1}{1 + (m_2/m_1)} \right]$$

is the fraction of the original K.E. that remains.

Let us write

$$a = \frac{1}{(1 + m_2/m_1)}.$$

The kinetic energy remaining clearly depends upon m_2/m_1 and there are three cases of interest.

a) When

$$m_1 \ll m_2,$$

so that

$$m_2/m_1 \to \infty,$$

consequently, $a \approx 1/\infty = 0$, so the kinetic energy remaining is zero. In other words, all of the original kinetic energy is dissipated. This case applies when a lump of putty is dropped on to the ground. m_2 is effectively the mass of the Earth and $m_2/m_1 \approx \infty$. Thus the putty does not bounce as all of its original kinetic energy is dissipated. The result is an obvious one for the two masses were assumed to stick together but it at least confirms the mathematical analysis.

b) When

$$m_1 = m_2,$$

so that

$$m_2/m_1 = 1.$$

The fraction, a, is now equal to $\frac{1}{2}$, and so $\frac{1}{2}$ of the original kinetic energy remains and the other $\frac{1}{2}$ is dissipated. This fact can be verified by using the data from experiment E to calculate the initial and final kinetic energy.

c) When

$$m_1 \gg m_2$$

so that

$$m_2/m_1 \to 0.$$

This time, $a \approx 1$, and all of the initial kinetic energy remains; consequently none is dissipated. This is what would be expected if a large mass hit a particle of dust that was at rest.

7.16 THE TRANSFER OF KINETIC ENERGY RESULTING FROM ELASTIC COLLISIONS

Consider again a mass m_1 moving with velocity u and interacting elastically with mass m_2 which is at rest.

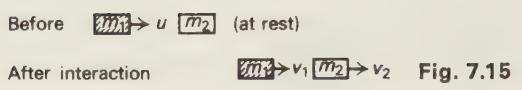


Fig. 7.15

This time, however, the masses do not stick together but have individual final velocities v_1 and v_2 in the same direction as u (Fig. 7.15).

The conservation of momentum rule applies, thus,

$$m_1\boldsymbol{u} = m_1\boldsymbol{v_1} + m_2\boldsymbol{v_2},$$

and

$$m_1 (u - v_1) = m_2 v_2. \tag{7.5}$$

From the original definition of an elastic collision,

$$\begin{pmatrix}\text{relative velocity}\\ \text{before the interaction}\end{pmatrix} = - \begin{pmatrix}\text{relative velocity}\\ \text{after the interaction}\end{pmatrix}$$

or,

$$(u - 0) = - (v_1 - v_2),$$

and

$$u + v_1 = v_2. \tag{7.6}$$

Combining this with Eq. (7.5) gives,

$$m_1 (u - v_1) (u + v_1) = m_2 v_2^2,$$

or,

$$m_1 u^2 - m_1 v_1^2 = m_2 v_2^2,$$

and dividing throughout by 2 gives,

$$\tfrac{1}{2}m_1 u^2 - \tfrac{1}{2}m_1 v_1^2 = \tfrac{1}{2}m_2 v_2^2.$$

This last equation shows that the kinetic energy lost by m_1 is gained by m_2. Thus kinetic energy is conserved and this requirement is often used to define an elastic collision although it was necessarily deduced from the original relative velocity definition.

We will now examine such questions concerning elastic collisons as: how v_2 is related to u; does the relationship depend upon the masses involved? Is there a maximum value for v_2? If so, does this imply a maximum kinetic energy transfer to m_2? Since kinetic energy is conserved, can it all be transferred from m_1 to m_2?

First, in order to relate v_2 and u, use Eq. (7.6) to eliminate v_1 in Eq. (7.5):

$$m_2 v_2 = m_1 (u + u - v_2)$$

$$v_2 (m_1 + m_2) = 2m_1 u$$

$$v_2 = \frac{2m_1 u}{(m_1 + m_2)}$$

$$v_2 = \frac{2u}{[1 + (m_2/m_1)]}. \tag{7.7}$$

Before discussing this equation, it is a small step to obtain a relationship between v_1 and u by combining Eqs. (7.7) and (7.6). Thus,

$$v_2 = u + v_1 = \frac{2u}{[1 + (m_2/m_1)]}.$$

It is left as an exercise in algebra to show that this leads to:

$$v_1 = u \left[\frac{1 - (m_2/m_1)}{1 + (m_2/m_1)}\right]. \tag{7.8}$$

Some of the questions posed earlier may now be discussed with the aid of Eqs. (7.7) and (7.8). These show that the velocities of the masses after the interaction are related to m_2/m_1 and three cases are considered.

a) If $m_1 \gg m_2$ so that $m_2/m_1 \to 0$, putting this value in the equations gives $v_1 = u$ and $v_2 = 2u$. Thus when a massive body collides elastically with a very light one at rest, the velocity of the massive one is little changed, and the velocity of the light one is twice that of the massive one.

b) If $m_1 = m_2$ then $m_2/m_1 = 1$ and the equations give

$$v_1 = 0 \qquad \text{and} \qquad v_2 = u.$$

This situation has been met before in experiments B and C.

c) If $m_1 \ll m_2$ then $m_2/m_1 \to \infty$ and Eq. (7.7) gives

$$v_2 = 2u/\infty = 0.$$

To avoid having ∞ in the numerator and denominator of Eq. (7.8), multiply top and bottom of this equation by m_1/m_2 so that it becomes

$$v_1 = u\frac{[(m_1/m_2) - 1]}{[(m_1/m_2) + 1]}.$$

Now, $m_1/m_2 \to 1/\infty = 0$ and when this value is used in the above, then $v_1 = -u$.

This case would apply to a perfectly elastic ball dropped on to the ground, for m_2 is effectively the mass of the Earth and in comparison, m_1 is negligible. v_2 is then zero (i.e., the Earth does not move) and the ball rebounds with a velocity equal and opposite to its original velocity.

These results are best summarized in Table 7.4.

Table 7.4

m_1	m_2	v_1	v_2
large	small	u	$2u$
equal	equal	0	u
small	large	$-u$	0

This raises some interesting points. Firstly, the maximum velocity, v_2, of the struck mass is only twice that of the striking mass. Consequently no matter how hard one kicks a ball, it can never travel at more than twice the velocity of one's foot.

The second point concerns the kinetic energy transferred to m_2. When does m_2 gain the maximum kinetic energy? An obvious answer would seem to be when v_2 has the maximum value, $2u$, but it must be remembered that when v_2 is a maximum, the mass of m_2 is very small. The answering of this question is the subject of the next section.

7.17 THE CONDITION FOR MAXIMUM ENERGY TRANSFER IN AN ELASTIC COLLISION

For m_2 to have the maximum energy, it is not sufficient for its velocity to be as large as possible; $\frac{1}{2}m_2v_2^2$ is the quantity that must have a maximum value.

Using Eq. (7.7):

$$\tfrac{1}{2}m_2v_2^2 = \tfrac{1}{2}m_2\frac{4u^2}{[1 + (m_2/m_1)]^2}$$

$$= \tfrac{1}{2}m_1u^2\frac{4(m_2/m_1)}{[1 + (m_2/m_1)]^2}$$

or

$$\tfrac{1}{2}m_2v_2^2 = \tfrac{1}{2}m_1u^2b,$$

where

$$b = \frac{4(m_2/m_1)}{[1 + (m_2/m_1)]^2}.$$

Now, $\frac{1}{2}m_1u^2$ is the original kinetic energy of m_1, and b is the fraction of this energy that is transferred to m_2. Therefore, for the maximum transfer of energy, b must have a maximum value.

The variation of b with different values of m_2/m_1 is best examined by making a table of values (Table 7.5) and by plotting a graph (Fig. 7.16). As can

Table 7.5

$\frac{m_2}{m_1}$	$\left[1 + \frac{m_2}{m_1}\right]^2$	$b = \frac{4\frac{m_2}{m_1}}{\left[1 + \frac{m_2}{m_1}\right]^2}$
0.01	1.02	0.039
0.1	1.21	0.33
0.25	1.56	0.64
0.5	2.25	0.89
1.0	4.00	1.00
1.5	6.25	0.96
2.0	9.00	0.89
4.0	25.00	0.64
10.0	121.00	0.33

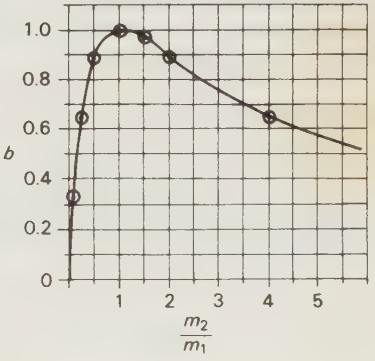


Fig. 7.16

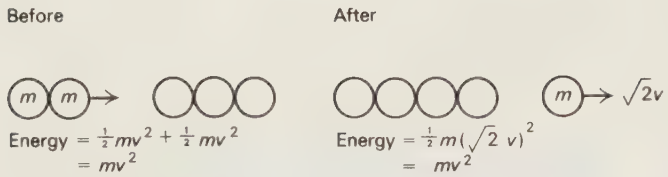


Fig. 7.18

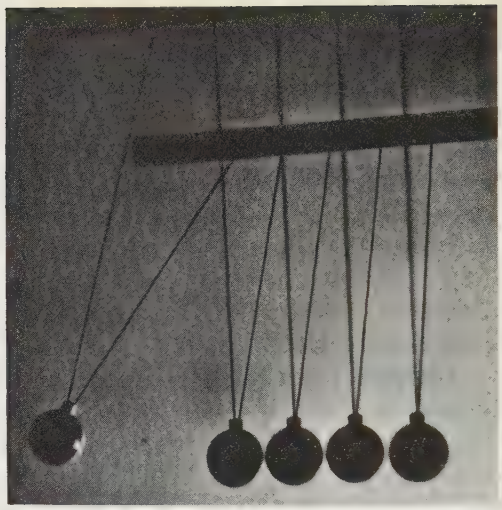


Fig. 7.17 Newton's cradle: a scientific toy. (Photograph by E.J.W.)

be seen from the graph, the maximum value of b is 1, and this occurs when $m_1 = m_2$. Thus it is possible for all of the initial energy to be transferred from one mass to another in an elastic collision. This is clearly the case in experiments B and C.

To sum up, when two masses collide elastically, the one struck gains the greatest energy when the two masses are equal, but the one struck has the greatest velocity when its mass is negligible with respect to the first mass. When this is so, their velocities are in the ratio 2:1.

As we shall see when considering electrical circuits (in Unit 6), an analogous situation arises when we try to transfer energy from a generator (e.g., an amplifier) to a load (e.g., a loudspeaker). The condition for the maximum transfer of energy requires the impedances of the source and the load to be equal, but the maximum current is associated with a very small value for the impedance of the load.

7.18 A SCIENTIFIC TOY

"Newton's cradle" (a name given to it by its makers) is shown in Fig. 7.17 and it consists of solid steel balls suspended from threads. It depends for its action on the conservation of energy and momentum laws, and also on the fact that when steel balls interact, the collision is very nearly elastic.

If one ball is allowed to swing down and collide with the rest, one will swing away from the far end. If two swing down together, then two swing away together at the far end, and so on for other numbers of balls. This effect can also be demonstrated with a row of steel balls in a horizontal groove, or even with a row of identical coins on a smooth surface.

This behaviour is consistent with both energy and momentum being conserved.

The question is sometimes raised that if two balls swing in at one end, can their combined energies all be given to one ball at the far end. This situation is illustrated in Fig. 7.18. If the original energy is given completely to the single ball then it will gain a velocity of $\sqrt{2}v$. However this cannot happen, for if this was the case, then momentum is not conserved.

$$\text{Initial momentum} = mv + mv = 2mv.$$

$$\text{Final momentum} = m\sqrt{2}v.$$

It is clear that these are not equal, and this example is completely fictitious.

Chapter 8
FRAMES OF REFERENCE

8.1 The conservation of energy in different frames of reference
8.2 Interactions in two dimensions
8.3 Inertial frames
8.4 Motion along a curved path
8.5 Motion in a circle
8.6 Circular motion and frames of reference
8.7 Spaceflight and weightlessness

All of the momentum experiments performed earlier were viewed from a frame of reference fixed to the Earth. In other words, velocity measurements were taken from a fixed point on Earth. In experiment F, for example, this point was the camera. The results of this experiment are shown in Fig. 8.1., which gives the speeds (m s^{-1}) and directions of travel of the air track vehicles. All these speeds are in m s^{-1} and are the speeds of the image across the photograph.

Such experiments and fixed frames of reference were used to deduce the law of conservation of momentum. But, suppose experiment F had been performed with the camera mounted on a small truck that could appear to move with a uniform velocity of 0.24 m s^{-1} from left to right.

The camera does not itself move at 0.24 m s^{-1}, but on the strobe photograph it produces, the Earth would move past at this speed. The speeds of vehicles and camera are proportional to "photograph" speeds.

The photograph taken from this moving frame of reference would show P to be initially stationary with Q moving at 0.39 m s^{-1} towards it. This, and the apparent velocities after interaction are as shown in Fig. 8.2.

We can see that momentum is again conserved even though the whole process was viewed from a moving frame of reference.

Next, suppose a fixed observer views the experiment performed in a moving frame of reference, e.g., the air track is on a table which is moving to the right

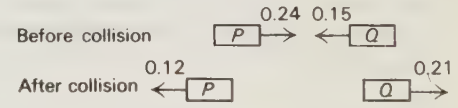


Fig. 8.1

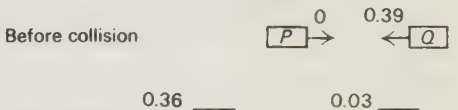


Fig. 8.2

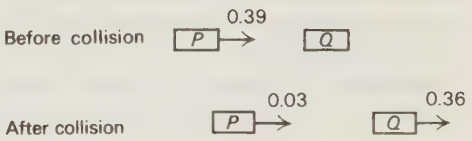


Fig. 8.3

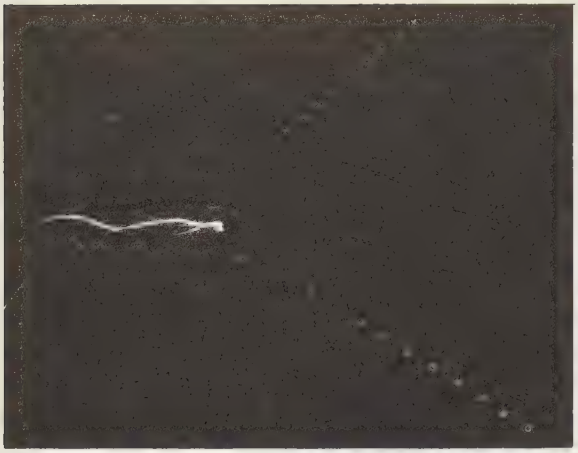


Fig. 8.4(a) Multiflash photograph of an explosion of two equal mass magnetic carbon dioxide pucks. The "explosion" is triggered by the breaking of a solder wire connection between the two pucks by melting the wire in a match flame. The continuous white line is due to this flame. The coupled pair are initially moving towards the right. The motion is shown in the laboratory frame of reference. (Courtesy of Education Development Center, Inc.)



Fig. 8.4(b) A multiflash photograph of the same explosion as shown in Fig. 8.4(a). This photograph was taken simultaneously with the other by a camera moving in the zero momentum (centre of mass) frame of reference. In this frame, the pucks are initially at rest and they move away from each other along a straight line after the explosion. (Courtesy of Education Development Center, Inc.)

at 0.15 m s^{-1}. Let the velocities relative to the table be the same as in Fig. 8.1. The observer now sees Q initially at rest and P is moving to the right at 0.39 m s^{-1}.

The complete experimental results are shown in Fig. 8.3, and again momentum is conserved.

If another observer sat on the table and moved along with it, he would see the situation shown in Fig. 8.1 and would also say that the law of conservation of momentum held.

8.1 THE CONSERVATION OF ENERGY IN DIFFERENT FRAMES OF REFERENCE

We have seen that momentum is conserved for any frame that is moving with constant velocity, but does the law of conservation of energy also hold? The masses of the vehicles used for Figs. 8.1, 8.2, and 8.3 are all equal so it is a simple matter to calculate the total energy before and after interaction in each of these figures. You will find that the loss of energy during interaction is the same for all three cases.

8.2 INTERACTIONS IN TWO DIMENSIONS

It is interesting to study, from different frames of reference, interactions in two dimensions. This can be done using two magnetic pucks that are initially linked and eventually allowed to "explode", and by photographing the process with two cameras, one stationary, and the other moving along with the same velocity as the linked pucks. The results of such experiments are shown in Figs. 8.4 and 8.5.

8.3 INERTIAL FRAMES

A frame of reference in which the conservation rules apply is called an inertial frame. As we have seen, a frame moving at constant velocity relative to an inertial frame is also itself an inertial frame. Any car, aircraft, or boat moving with constant velocity relative to the Earth is an example of an inertial frame. If an air puck is put on a horizontal surface in any of these moving "laboratories," it will be found that—a body remains at rest (relative to the laboratory) or



Fig. 8.5(a) Multiflash photograph of two carbon dioxide pucks of unequal mass. The coupled pair of pucks is initially moving towards the right in the photograph. The puck marked (O) has twice the mass of the puck marked (X). (Courtesy of Education Development Center, Inc.)

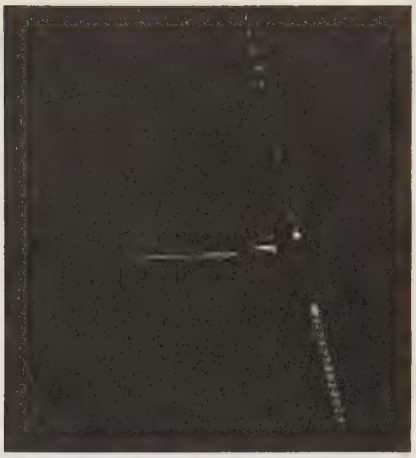


Fig. 8.5(b) A multiflash photograph of the same explosion as shown in Fig. 8.5(a). This photograph was taken simultaneously with the other by a camera moving in the zero-momentum frame of reference. In this frame, the pucks are initially at rest and they move away from each other after the explosion in a straight line. (Courtesy of Education Development Center, Inc.)

moves with constant velocity unless acted upon by an unbalanced force. Also, calculations of energy transfers could be carried out. In fact, these moving labs are indistinguishable from stationary ones and no experiment performed in them will give any clue about their uniform motion. Such motion can only be recognized by looking out of the window! An aircraft on the ground with its curtains drawn and engines running is indistinguishable from one flying with uniform velocity in smooth air conditions.

Consider an open car travelling in a straight line at 10 m s^{-1} carrying a horizontal surface on which a puck appears to be at rest to an observer, C, in the car. To an external observer, B, who is looking down from a balloon, the puck has a uniform velocity of 10 m s^{-1}. The car now accelerates because the engine and tyres cause a force to act on it, but this force is not transmitted to the puck since friction between it and the car is made negligible. Consequently, to B above, the puck still has a uniform velocity of 10 m s^{-1} but C thinks the puck is accelerating towards the rear of the car. Since, as far as C is concerned, the puck's acceleration apparently occurs without a force acting on it, Newton's laws no longer operate. This accelerating frame is not an inertial frame of reference.

A rotating frame of reference is another example of a non-inertial frame. An experimenter, T, is sitting on a turntable with a puck on a horizontal glass sheet (Fig. 8.6). An external observer, E, is above on a fixed platform near the ceiling. Initially the turntable is at rest and T pushes the puck. Both T and E observe its uniform motion in a straight line. Both are in an inertial frame. Next the turntable is made to rotate with uniform angular velocity and T again gives the puck a push. Since friction between the puck and glass is negligible, the motion of the turntable cannot affect the motion of the puck. Consequently E again observes its motion to be uniform and in a straight line in his fixed frame, although the glass is moving beneath it. However, T does not observe such a motion and he sees the puck move in a curved path which is not what Newton's laws predict! T's rotating frame is a non-inertial one.

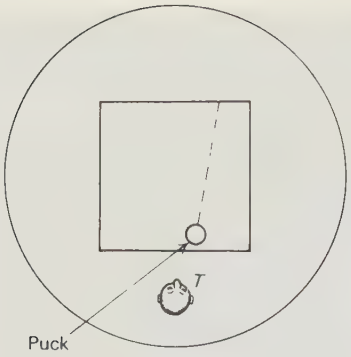


Fig. 8.6

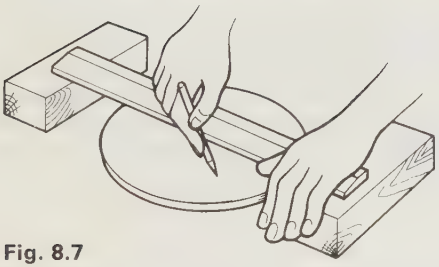


Fig. 8.7



Fig. 8.8 Bombs falling from a World War II Mitchell bomber of the R.A.F. (Courtesy of Imperial War Museum.)

Experiment

The type of curve that T observes can be easily drawn if a gramophone turntable is provided with a white cardboard disc. A ruler is arranged on blocks so that it straddles the turntable which is free to rotate beneath the ruler (see Fig. 8.7). If the disc rotates at 16 or 33 revolutions per minute while a pencil is moved along the ruler at a steady speed, a curved path will be drawn on the card.

The topics are beyond the scope of this book but it is worth mentioning that Einstein's theory of special relativity deals with inertial frames, while his theory of general relativity deals with accelerated frames.

8.4 MOTION ALONG A CURVED PATH

If an aircraft drops a bomb, a description of its trajectory will depend on the frame of reference of the observer. To an observer in an aircraft flying alongside with the same velocity as the bomber, the bomb will appear to fall in a straight vertical line (Fig. 8.8). It appears to fall in exactly the same way as a stone dropped from a stationary balloon when viewed from the ground.

To an observer in the bomber looking vertically downwards, the bomb, although it is getting smaller, appears to be travelling forwards with respect to the ground with a uniform velocity, i.e., that of the bomber.

To an observer on the ground, standing a long way to the side of the flight path, the bomb appears to move in a parabola. This trajectory is a combination of the two separate motions mentioned:

a) a horizontal uniform velocity,
b) a vertical uniform acceleration like that of a stone falling from rest.

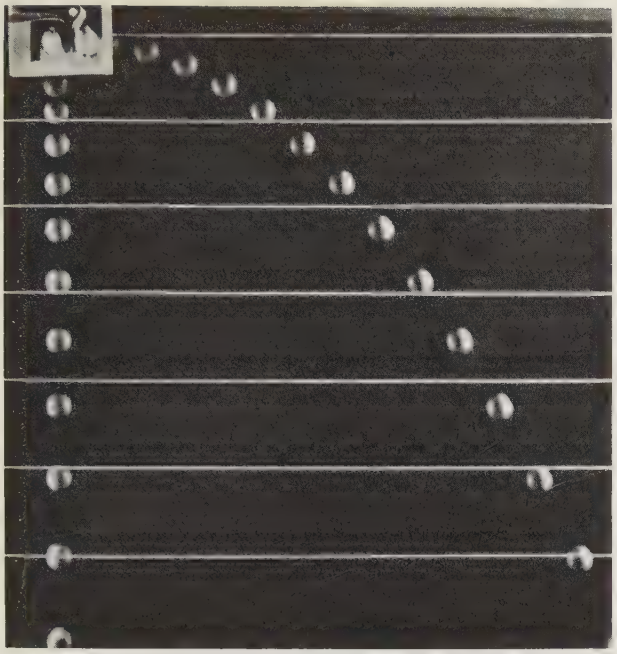


Fig. 8.9 Multiflash photograph of two golf balls released simultaneously from the mechanism shown. One of the balls fell freely. The other was projected horizontally with an initial velocity of 2.00 m s^{-1}. The flash rate was 30 Hz. The white lines in the figure are 15 cm apart. (From P.S.S.C. *Physics*, D. C. Heath, 1965, reproduced by permission of Education Development Center, Inc.)

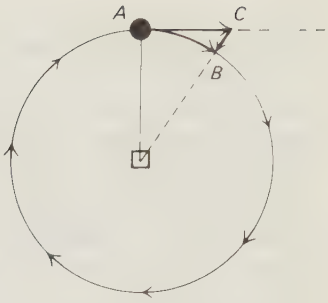


Fig. 8.10

This is illustrated by the strobe photograph, Fig. 8.9, which shows two golf balls, one falling freely from rest and, at the same instant, the other being projected horizontally. Both fall with the same vertical motion, and while doing that, the right hand ball moves equal horizontal distances in equal times.

8.5 MOTION IN A CIRCLE

The technique of considering a curved trajectory to be the result of two linear motions, can be used when discussing the force that must act on a mass if it is to move in a circle with uniform speed.

Consider a simple experiment involving an air puck and a horizontal glass surface. Put a nail into a small block of wood (about 2 × 2 × 1 cm) and tether the puck to it with about 20 cm of thread. The block can be held down on the glass plate with one finger. If the puck is pushed, once, tangentially, it will continue to move in a circular path with a constant speed (see Fig. 8.10). If, when the puck is at A, the finger holding down the block is removed, the puck will travel with the same constant speed but along a straight line AC which is a tangent to the circle.

A flexible thread in tension can only exert a force along its length, and so, when the tether is present it provides a force that acts on the puck towards the centre of the circle, i.e., a centripetal force acts *on* the puck. Now, a force acting on a mass gives it an acceleration, and so the circular motion of the puck can be considered as being made up of two *simultaneous* linear motions, (a) a uniform tangential velocity, along AC, and (b) an inward radial acceleration, along CB. Let us calculate the value of this acceleration that changes uniform linear motion into

circular motion, i.e., which makes the puck arrive at B on the circle rather than at C.

Suppose the puck moves around the circle, of radius r, with a constant speed v, and that the time taken to travel from A to B along the arc is t. Let the inward radial acceleration be a.

The journey, A to B, can be replaced by a combination of : uniform motion from A to C where $AC = vt$, and acceleration from C to B where $CB = \frac{1}{2}at^2$.

(Since the acceleration of the puck $= a$, in time t it gains a speed $= at$. It starts from rest at C and so its average speed $= \frac{1}{2}at$. Distance = average speed $\times$ time, $CB = \frac{1}{2}at^2$.)

These values have been shown on Fig. 8.11 which depicts part of the circular path.

If B is close to A, then (1) arc $AB \approx$ line $AB \approx AC = vt$ and (2) $A\hat{B}C \approx 90°$ and so the shaded triangles are similar.

Therefore:

$$CB/AC = AD/AO,$$

i.e.
$$\tfrac{1}{2}at^2/vt = \tfrac{1}{2}vt/r,$$

and
$$a = v^2/r.$$

If the puck has a mass m, then the centripetal force that must act on it if it moves in a circular path is mv^2/r. This force is provided by the tension in the thread; if the thread breaks, the force on the puck ceases to act and it will no longer move in a circle.

8.6 CIRCULAR MOTION AND FRAMES OF REFERENCE

It should be noted that this treatment of circular motion has been carried out as viewed from a stationary frame of reference. It is clear that when a mass is moving in a circle there must be exerted on it a force that acts towards the centre of the circle. It is interesting to look at the same problem from a rotating frame. Consider again the turntable used in Fig. 8.7 and suppose the experimenter T is holding the puck in the position shown. As far as he is concerned in his rotating frame the puck is stationary. Moreover, he would conclude that bodies at rest have forces acting on them for he can feel the puck trying to move his hand away from the centre of rotation. He could argue that a centrifugal force exists because he has to provide an equal centripetal force so that the resultant force on the puck is zero and the puck can be in a state of rest. In fact, T has invented a fictitious centrifugal force in order to be able to apply Newton's laws in his rotating frame which is an accelerating one and non-inertial. A stationary, external, observer realizes that the puck, in moving around the circular path, is continually accelerating towards the centre of rotation and that T is supplying the necessary force towards the centre.

The fictitious centrifugal force is indeed a mysterious one for it has no obvious source, no strings, no magnets, no springs. The mystic character of this centrifugal force is even more apparent when T releases his hold on the puck and observes what happens. If he applies his knowledge of Newton's laws he would expect the outward-acting force to produce a linear motion away from the centre of rotation. But no, he sees the puck move in a type of spiral called an involute—and moreover this motion occurs with no obvious application of a force!

You may like to draw the exact shape of an involute. This is easily done by taking a length of string, a cylindrical container about 4 cm diameter (a small tumbler will do), a pencil and paper. Tie the string to the pencil point as in Fig. 8.12 and wind the string around the cylinder until the pencil is close to the curved surface. With one hand hold the cylinder down on the paper and use your other hand to move the pencil so that the string remains taut and unwinds from the stationary cylinder. The path traced out is an involute. It is instructive to draw the curve in a slightly different way. Rotate the cylinder and paper as one (it may be necessary to glue them together) and keeping the string taut, allow the pencil to move in a constant direction in your fixed frame of reference. The curve produced is exactly the same as before but this time you are having the view of a fixed observer above the turntable when T releases his hold on the puck, namely, the puck moves in a straight line with constant speed as Newton's laws predict!

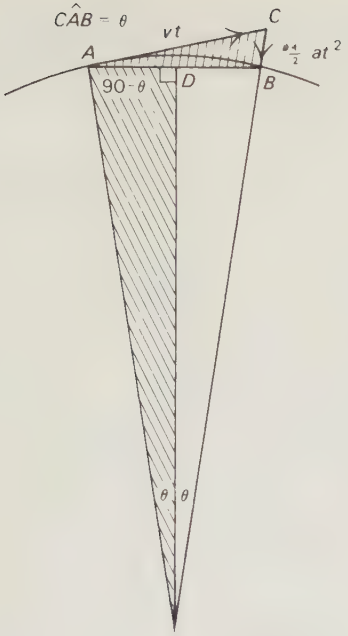


Fig. 8.11



Fig. 8.12

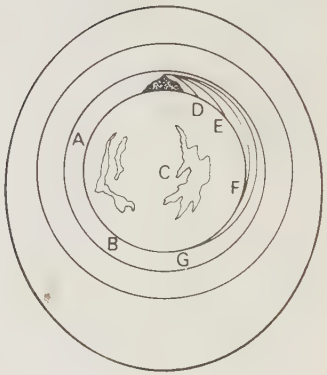


Fig. 8.13 Newton's presentation of the paths of projectiles launched horizontally from various heights above the Earth's surface.

8.7 SPACEFLIGHT AND WEIGHTLESSNESS

Experiment 1

If two large ball-bearings are allowed to fall from rest they will move with the same acceleration. If they start at the same instant the two may be expected to fall together even though their masses may be different, as the "guinea and feather experiment" shows. Suppose that, when they are released they are just touching, then they will continue to fall in contact.

Experiment 2

The two masses can together be thrown horizontally. This means that as well as having the same vertical acceleration they also have the same velocity in a horizontal direction. As Fig. 8.9 illustrates, such a combination of linear motions results in a parabolic trajectory and so the two masses will follow two such identical paths. If they start in contact they will remain in contact.

The term "in contact" has been used in the same sense that they are just touching but not exerting a force on one another, for, if they did, they would move apart.

Experiment 3

If the masses are given a higher common velocity they will again travel in contact but will strike the ground further away. If they are given a sufficiently high horizontal velocity they will go into Earth orbit. (The

Earth's gravitational attraction provides the centripetal force necessary for circular motion.)

Isaac Newton performed such a "thought experiment" with a cannon on a mountain and Fig. 8.13, which is from his *System of the World* (dated 1729), illustrates the point well.

Suppose that one of the ball-bearings is now put inside a metal box and the three previous experiments are repeated, i.e., the box is dropped, thrown, put into orbit. As before, the box and the ball are each in the same gravitational field and have the same motion; consequently they are travelling together and are exerting no forces on one another.

This is the situation an astronaut finds himself in when he is in a space capsule which is in free flight (i.e., the rocket engines are not burning). Since the capsule is exerting no force on him, he describes himself as *weightless*. This is perhaps not the best word to use because he is in the gravitational field of a planet (or planets) and must therefore have weight. It might be better to describe an astronaut's condition as *reactionless* rather than weightless.

On the other hand, like all Earth-bound creatures, the astronaut is accustomed to a gravitational field acting throughout the volume of his body while the surroundings exert a supporting force on a small area of his surface. This supporting force, or reaction, is recognized as being the result of his weight. When he is in orbit no supporting force exists and so he describes his state as weightless. This approach is especially excusable if he has tried to find his weight by connecting himself to his surroundings, i.e., the capsule, with a spring balance, for this will read zero!

PROBLEMS: Unit Two

2.1 A man descending by parachute, and having nothing to do for a minute or two, passes the time by thinking about Newton's third law of motion. He reasons that his speed is constant because his weight is equal but opposite to the force the parachute exerts on him and so he experiences no resultant force and hence he has zero acceleration.

a) Is the reasoning contained in the last sentence correct?
b) Is the man correct in applying Newton III to the forces mentioned?
c) Name four other forces involved in this situation. (Consider the air, the parachute, the man and the Earth.)
d) Take all six forces involved and group them into three pairs of forces that obey Newton III.

2.2 Explain why Newton III applies when you press your hand against a brick wall, but not when you are holding a brick and equating the weight of the brick to the upward force of your hand.

2.3 A super tanker has a mass of 409 600 tonnes (1 tonne = 10^3 kg), its engines have been stopped but there is still a slight forward motion of only 1 m in 2 s. What kinetic energy does the ship have?

In order to stop completely, the anchor is dropped and once it starts to hold, the ship comes to rest in 20 m. Estimate the average tension in the anchor chain. If the tensile strength of the chain's steel is 8×10^8 N m^{-2}, what is the minimum area of cross-section of one side of a link? How would your answer be modified if the ship had been moving at a speed of 1 m s^{-1} initially? (Assume that the stopping distance is still 20 m.)

2.4 A car of mass 1000 kg is travelling at 15 m s^{-1} along a straight road. How much kinetic energy does it have? If the application of the brakes causes an average retarding force of 2000 N to act on the car, how far does the car travel before coming to rest? Also calculate the stopping distance if the car had been travelling at 30 m s^{-1}. What moral can be drawn from your answers?

2.5 The following table is printed in the Highway Code.

M.p.h.	Thinking distance (ft)	Braking distance (ft)	Overall stopping distance (ft)
20	20	20	40
30	30	45	75
40	40	80	120
50	50	125	175
60	60	180	240
70	70	245	315

a) Explain why thinking distance is directly proportional to speed.
b) Use your knowledge of physics to explain the figures in the braking distance column.

2.6 A car has an engine and transmission that delivers 50 b.h.p. to the wheels; (b.h.p. means brake horse power, i.e., actual measured power rather than a calculated value. 1 b.h.p. = 746 W). If it is travelling along a level road at 144 km/h, calculate the total wind and road resistance to motion.

2.7 Use the information in Chapter 5 to calculate the approximate dimensions of the prototype kilogramme.

2.8

a) The equation $F = kma$ is simplified by measuring F in newtons and making $k = 1$. What units does 1 have?
b) The choice of k is arbitrary. Make $k = 2$, name the resulting unit of force after yourself and write out its definition.
c) Is your unit of force bigger or smaller than the newton?

2.9 Repeat Problem 2.8 using the equation $W = KFs$ where W = work done, F = force, s = distance in the direction of the force.

2.10 An archer's bow is drawn back using a spring balance. Readings of force (F) and displacement (d) are listed in the table below.

F in N	0	50	100	160	250	400
d in m	0	0.2	0.3	0.4	0.5	0.6

Plot a graph of these values and use it to determine the energy transferred to the bow when it is fully drawn. An arrow of mass 400 g is fired with this bow. Calculate (a) the initial acceleration of the arrow, and (b) the maximum velocity of the arrow.

2.11 A car of mass 1000 kg climbs a hill, 300 m high, along a road 2.1 km in length. The total resistance to motion due to the friction with the road and wind is, on average, 600 N. (Assume that g = 9.8 N kg^{-1}.)

a) Calculate the energy used by the car climbing the hill.
b) Calculate the component, parallel to the road, of the force exerted by the tyres while the car is climbing.
c) If the car stops on the hill, what is the force now exerted parallel to the road that prevents it running backwards?
d) What have you assumed about the hill in parts (b) and (c)?

2.12 The gravitational field of the Earth is 9.8 N kg^{-1}.

a) What is the force of attraction exerted on 2 kg?
b) What is the force of attraction exerted on x kg?
c) What is the acceleration of the 2 kg mass if it falls?
d) What is the acceleration of the x kg mass if it falls?
e) What can you conclude about freely falling masses near to the Earth's surface?

2.13 A trolley of mass 40 kg has free-running wheels and is initially at rest on a horizontal surface (Fig. P2.1) It is connected by a light rope, which passes over a well-oiled pulley, to a mass of 10 kg. Assume that g = 10 N kg^{-1}.

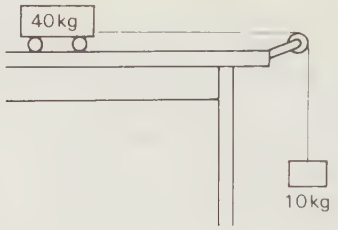


Fig. P2.1

a) What is the horizontal force acting on the trolley?
b) What is the total mass in motion?
c) What is the acceleration of the masses?

If the trolley moves a distance of 2 m from rest,

d) What is the potential energy lost by the system?
e) What is the kinetic energy gained by the trolley?
f) What is the speed of the trolley?

2.14 Some small toy cars use for propulsion the energy stored in a flywheel; these toys are popularly known as friction powered. Think of a better name for these cars, and write a letter to a manufacturer in which you argue the case for adopting your description.

2.15 When driving his car, Jock McTaylor bemoans the fact that on applying the brakes, the kinetic energy of the car is mostly converted into heat and wastefully discarded. Jock invents a mechanical system with which he can connect the car's wheels through a complicated system of gears, to a flywheel, whenever he wants to stop. As the flywheel accelerates it stores the kinetic energy that the car loses on slowing down. The energy in the rotating flywheel is then available to set the car in motion once again, just as in some toy cars. Supposing the mechanical difficulties can be overcome, is this a viable idea? To help you decide, calculate the mass of a suitable flywheel assuming it to be a disc of radius 0.5 m that must not exceed a speed of 3000 revolutions per minute. Take the mass of the car to be 1000 kg and its initial speed as 20 m s^{-1}. Calculate also the thickness of the disc if the density of its material is 8000 kg m^{-3}.

2.16 You may consider that the flywheel in Problem 2.15 is inconveniently large. Repeat your calculations of mass and disc thickness if the radius is reduced to 0.2 m and the

peak speed is 2400 revolutions per minute. Is this a more convenient size of flywheel?

2.17 A rifle of mass 5 kg fires a bullet of mass 0.15 kg with a muzzle velocity of 600 m s^{-1}. Calculate the velocity with which the rifle starts to recoil. If the rifle recoils through a distance of 9 cm, what is the average force exerted on the shoulder?

2.18 In radioactive decay, a Radon nucleus, of mass 3.52×10^{-25} kg emits an alpha particle of mass 6.4×10^{-27} kg with an energy of 6.3 MeV. If 1 eV = 1.6×10^{-19} J, calculate,

a) the speed of the alpha particle,
b) the speed of recoil of the remains of the parent nucleus.

2.19 During the process of nuclear fission, a uranium nucleus at rest splits into two parts of mass numbers 96 and 140 respectively. The total energy of these parts is about 200 MeV. Find the energies and the velocities of the two fragments. (The mass of a proton = 1.6×10^{-27} kg; 1 eV = 1.6×10^{-19} J.)

2.20 An astronaut is engaged on activity some distance from his orbiting spacecraft. How could he return to the capsule without pulling on his life-line?

2.21 A truck, *A*, of mass 3000 kg is shunted along a railway line at 8 m s^{-1} when it collides, and links with, truck *B* of mass 5000 kg that is travelling in the same direction at 4 m s^{-1}. What is their final common velocity? Also, calculate the energy lost in the process.

2.22 Repeat Problem 2.21, but reverse the direction of *B* so that it has a head-on collision with *A*. If, as before, they link together calculate their final common velocity, stating carefully its direction of course. Also, what is the energy lost this time?

2.23 An air track vehicle of mass 0.2 kg is moving at 5 m s^{-1} when it collides elastically with a stationary vehicle of mass 0.5 kg. Calculate the velocities of the vehicles after the interaction. Calculate the fraction of the original energy that is transferred to the second vehicle, and use the graph in Fig. 7.16 to check your answer.

2.24 An air track vehicle, *A*, of mass 0.3 kg is travelling at 2 m s^{-1} towards another vehicle, *B*, of mass 0.2 kg, which is moving at 5 m s^{-1}. As a result of the collision the velocity of *A* is reversed. What is the final velocity of *B* and calculate the loss of energy.

2.25 Discuss the popular belief that in order to provide propulsion, the exhaust gases of a jet engine or rocket must push against the air in the atmosphere.

2.26 An alpha particle of mass number 4 travelling at 1.8×10^{7} m s^{-1} has a head-on elastic collision with a gas atom. The gas atom moves forwards at 0.8×10^{7} m s^{-1} and the alpha particle rebounds at 1.0×10^{7} m s^{-1}. From this data calculate the mass number of the gas atom.

2.27 A rocket has an initial mass of *M* and its exhaust gases pass through a jet of effective area *A* with a uniform velocity *v*. If the density of the gases is *d*, calculate the minimum value of *v* which will enable the rocket to take off vertically. Calculate also the acceleration with which the rocket starts to rise, and discuss qualitatively how you would expect the acceleration to vary during the ascent of the rocket.

2.28 Calculate the minimum rate, in kg s^{-1} of fuel consumption of a rocket if it takes off vertically. At the moment of lift-off, its total mass is 3×10^{6} kg and the exhaust gases have a velocity of 1500 m s^{-1} relative to the rocket.

2.29 A method that can be used for determining the speed of a bullet is to fire it into a suspended box of sand (Fig. P2.2). The bullet becomes embedded in the sand, and the box then swings, and rises through a vertical distance *h* as shown. Calculate the speed of the bullet if its mass is 0.2 kg, the mass of the box of sand is 29.8 kg, and $h = 0.8$ m. Why is it incorrect to equate the kinetic energy of the bullet to the rise of potential energy of the bullet and box?

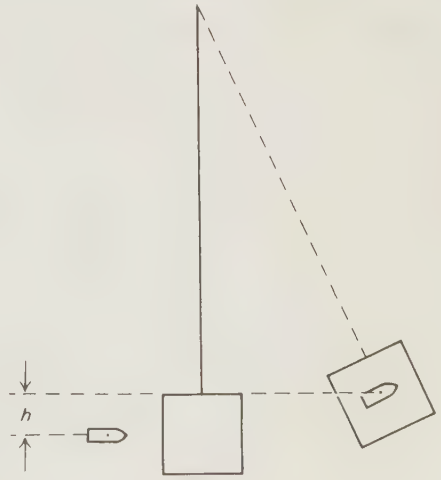


Fig. P2.2

2.30 A bullet of mass B and velocity v passes through a pendulum bob of mass M and emerges with velocity $\frac{1}{2}v$. The pendulum bob is at the end of a string of length l. What is the minimum value of v such that the bob will swing through a complete circle?

2.31 Two equal lumps of putty are suspended as shown in Fig. P2.3. One is drawn aside so that its centre of mass rises a vertical distance h. It is released and it collides inelastically with the other lump. Calculate the vertical distance risen by the centre of mass of the combination.



Fig. P2.3

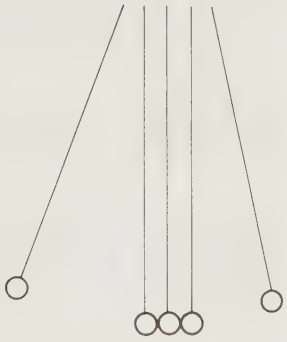


Fig. P2.4

2.32 Apply the principle of conservation of momentum to Newton's Cradle and explain what happens when the two outer balls are released simultaneously from different heights as shown in Fig. P2.4.

2.33 The Earth is rotating and our experiments on Earth show that Newton's laws apply to a frame of reference fixed to it, yet a rotating frame is not inertial. Discuss.

2.34 A control-line model aircraft has two thin wires connecting it to a handle held by the "pilot". If when the wires are taut the aircraft flies in a horizontal circle of diameter 28 m at a speed of 72 km h^{-1}, calculate the tension in each wire. The mass of the model is 0.35 kg.

The model is then flown so that it climbs in a vertical circle, passes overhead and dives down almost to ground level before resuming horizontal flight. Calculate the total wire tension when it is directly overhead if its speed then is 57.6 km h^{-1} (take $g = 10$ N kg^{-1}). Calculate also the minimum speed the aircraft must have if the control wires are to remain in tension during the above manoeuvre.

2.35 Taking the radius of the Earth to be 6400 km, estimate the speed of a satellite travelling close to the Earth's surface. (Assume that $g = 10$ N kg^{-1}.)

How long does it take to complete one orbit?

2.36 A synchronous communications satellite moves in a circular orbit in the equatorial plane and remains over the same point on the Earth at all times. If the radius of the orbit is 42 400 km, calculate the gravitational field strength at this distance.

Is your answer consistent with g being proportional to $1/r^2$, where r = distance from the centre of mass of the Earth? Take the radius of the Earth to be 6400 km and g at its surface as 9.8 N kg^{-1}.

2.37 Explain carefully why it is easier to launch a synchronous satellite (see Problem 2.36) from America, than from Russia.

BIBLIOGRAPHY: Unit Two

Arons, A. B. (1965). *Development of Concepts of Physics*, Addison-Wesley.

Cohen, B. (1961). *The Birth of a New Physics*, Heinemann.

A brief historical account of man's changing view of motion.

Gamow, G. (1962). *Gravity*, Heinemann.

Harvard Project Physics. (1968). Text 1 *Concepts of Motion*, Reader 1 *Concepts of Motion*; Text 2 *Motion in the Heavens*, Reader 2 *Motion in the Heavens*, Holt, Rinehart and Winston.

Well illustrated and interesting reading about the subject. The Readers are especially valuable source material.

Holton, G. and Roller, D. H. D. (1958) *Foundations of Modern Physical Science*, Addison-Wesley.

Jardine, J. (1970). *Mass in Motion*, Longman.

Includes a valuable collection of illustrations.

Nuffield Physics *Teachers' Guide* IV. (1967). Longman/Penguin.

P.S.N.S. (1969). *An Approach to Physical Science*, Wiley.

P.S.S.C. (1965). *Physics*, Heath.

Rogers, E. M. (1960). *Physics for the Inquiring Mind*, Oxford.

The first part "Matter, Motion and Force" is very relevant.

Stong, C. L. (1964). *The Dynamics of a Golf-ball*, Scientific American.

Wenham, E. J. (1969). *Planetary Astronomy*, Longman.

UNIT THREE
Thermal Physics

9 Temperature
10 Temperature, Heat, and Work
11 Heat Capacities
12 A Second Law

Chapter 9
TEMPERATURE

Matter, including of course gases, has a number of properties which, like pressure, are temperature dependent. Among these are the length of a metal bar (for we know that metals expand on heating), the pressure of a fixed volume of a gas, the volume of a mass of gas whose pressure is kept constant, the resistance of a metallic conductor, the electro-motive force generated at a junction of two metals, the colour of a hot furnace. Devices can be constructed using any of these properties and used to indicate the thermal state of a body.

Imagine that such a device is placed in each of two metal cans containing water. One of these cans, *A*, has been in good thermal contact with the flame of a bunsen burner: the other, *B*, has been standing on the bench. Now they are brought together and placed in good thermal contact with one another. Experience tells us that, before long, the two similar devices will record that an equilibrium state exists. If, however, the two cans have been separated by an insulating wall, the devices will continue to vary widely from one another.

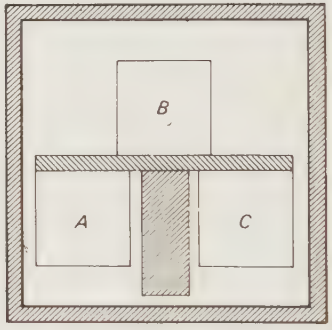


Fig. 9.1

Now imagine that we have three such cans, *A* and *B* as before but the third, *C*, has been kept in a refrigerator for some time. The three are placed in an insulated chamber with an insulating wall separating *A* and *C*, and a conducting wall between *A* and *B*, *B* and *C* (Fig. 9.1).

Both *A* and *C* are in good thermal contact with *B* but not with each other. Given the lapse of sufficient

time, A, B, and C will come into thermal equilibrium. This state of affairs would not be upset if we were to withdraw the insulating wall from between A and C. This experimental result is often described as the zeroth law of thermodynamics and can be stated formally as "Two systems which are in thermal equilibrium with a third system must be in thermal equilibrium with each other." Under these conditions we say that all three systems are at the same temperature. This gives a very clear meaning to the word temperature. It is the property of a system which determines whether or not the system is in thermal equilibrium with other systems.

Problem Does the law analogous to the zeroth law of thermodynamics apply to two pieces of iron and a magnet, and to three men, Smith, Jones, and Robinson?

a) Piece of iron A attracts—or is attracted by—the magnet. So is piece B. Does A attract B? It might or it might not. There is no analogous law.
b) Smith knows Jones. Jones knows Robinson. But does Smith know Robinson? He may do; but then again, he may not. There is no analogous law.

Properties of matter which are temperature dependent are all capable of use in a thermometer to measure temperature. To devise an effective method of providing a temperature scale, some fixed reference point must be agreed. One such is the triple point of water. Water can exist as solid, liquid, and gas in the same vessel at one temperature and one pressure alone. The pressure is 4.58 mm of mercury; the temperature can be assigned arbitrarily. By international agreement, the arbitrary number 273.16 K is assigned to the triple point of water, the K indicating a temperature in degrees Kelvin (formerly written °K or even °A). Naturally, there were very good reasons for selecting such a number. Not least among these was that it preserves the difference of 100 degrees between the temperature at which water boils and that at which ice melts.

Having made this arbitrary decision, we can define a temperature in terms of the property which is being used in our thermometer. For a mercury-in-glass thermometer, we choose to write

$$\frac{\text{temperature}_X}{273.16} = \frac{\text{length of mercury column at } X}{\text{length of the mercury column at the triple point of water}} \text{ K}$$

that is

$$T_X = 273.16 \frac{l_X}{l_{273.16}} \text{ K}.$$

Had we been using the pressure of a fixed volume of a gas we should have

$$T_X = 273.16 \frac{\text{pressure of a gas at } X}{\text{pressure of a gas at the triple point of water}} \text{ K}.$$

Whilst for a resistance thermometer we should have

$$T_X = 273.16 \frac{\text{resistance of the thermometer at } X}{\text{resistance of the thermometer at the triple point of water}} \text{ K}.$$

With this approach to the measurement of temperature, we should be surprised if two thermometers using totally different physical properties gave identical readings when put together into an enclosure at an unknown temperature—unless, of course, that temperature was that of the triple point of water.

What is perhaps even more surprising is that there is a group of thermometers which approach the same reading. They are all constant volume gas thermometers operating at very low pressures. Whether the gas be air, oxygen, hydrogen, or helium, at very low pressures these instruments agree as to the temperature (see Fig. 9.2). This agreement sets a standard—a constant volume gas thermometer at zero pressure. In practice, helium is found to be the best material and it can be used as low as 1 K.

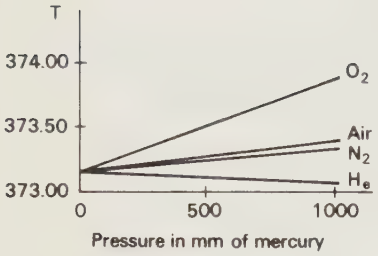


Fig. 9.2 Readings of a constant volume gas thermometer at the temperature of condensing steam, when different gases are used at various pressures. (F. W. Sears and M. W. Zemansky, *University Physics*, Addison-Wesley, 1970.)

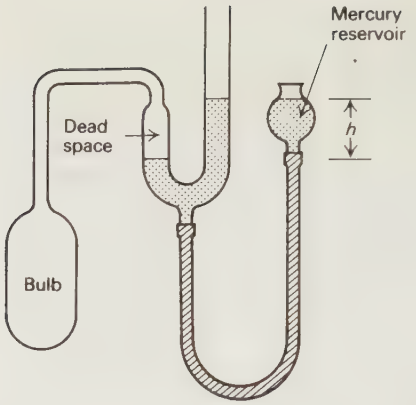


Fig. 9.3 A constant volume air thermometer. (F. W. Sears and M. W. Zemansky, *University Physics*, Addison-Wesley, 1970.)

A constant volume gas thermometer consists essentially of a bulb containing the chosen gas, which is connected through a capillary tube to a mercury column (Fig. 9.3). This is itself connected to a second tube of mercury which can be raised or lowered to adjust the pressure in the bulb. The difference in height of the two mercury levels is measured when the gas is contained in its fixed volume at the temperature of the triple point of water and at the temperature which is to be measured. Then the equation

$$T_x = 273.16 \frac{p_x}{p_{273.16}} \text{ K}$$

gives the temperature.

When very low values of gas pressure are used, this measure of temperature depends but little on the properties of any particular substance but on the properties of gases in general. It is taken to be the ideal gas temperature.

Chapter 10
TEMPERATURE, HEAT, AND WORK

In the study of energy the term *work* has acquired a specific meaning as a measure of energy transformed from one form to another. For example, the application of a vertical force of 50 newtons to a mass which is moved at a slow constant speed through a distance of 2 m vertically involves the transfer of 50×2 joules of energy.

$$\text{Work} = \text{force} \times \text{distance moved in the direction of the force.}$$

In this example energy has been transferred from whatever energy store was contained in that which exerted the forces to a form which is called gravitational potential energy.

The mass which has been lifted through these 2 m may, at some time in the future, return the potential energy stored as a result, by falling back to its starting point. In so doing it might drive a paddle wheel contained in a carefully insulated vessel of water as in the experiments of J. P. Joule. Those experiments revealed that the temperature of the water would rise as a result of this process of energy transformation; and, moreover, that the expenditure of the same amount of work W always produced the same temperature rise in the same mass of water.

The implications of this discovery are far-reaching. Consider for a moment a simple "thought experiment." You have two similar insulated systems each containing a vessel of water. Under one of these vessels is a flame; the other contains the paddle wheel and falling mass mechanism. Each also contains a thermometer and, whilst you can see the reading of that thermometer, you cannot see inside the insulated containers which surround the two systems.

Each system is then set in operation. You observe that the temperature of the water rises in each vessel. You have no method of differentiating between the two systems. You know that in one case energy is being transferred to the water from the falling mass and that you can measure this as work. One consequence is that the temperature of the water rises. If you accept that energy is conserved, you must assume that the water is increasing some internal store of energy as this happens and that the temperature rise

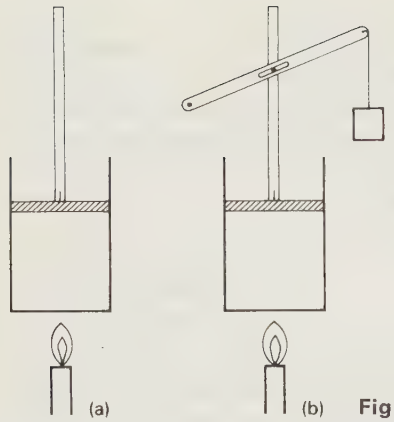


Fig. 10.1

indicates this. There is no reason why you should believe anything different is happening to the other system. But, commonly, it is said that the flame is transferring heat to that system. Since we take work to be in Maxwell's phrase, "the transference of energy from one system to another," it would seem sensible to take heat, too, as energy in transit. There must be some mechanism which enables the water to store this energy internally just as there must be some mechanism which permits the storage of what is called gravitational potential energy in the lifted mass. That mechanism will be considered later.

The classic series of experiments performed by Joule in the nineteenth century was not limited to the simple case of energy transformations which occur when gravitational potential energy is transferred to water through a paddle wheel mechanism. Joule also arranged for the falling masses to drive a dynamo which was short-circuited and itself immersed in the water; on another occasion the output current of the dynamo was allowed to pass through a resistor which was immersed in the water. Joule forced water to flow through fine tubes and observed the temperature change which resulted from this fluid friction; he compressed air in a cylinder which was immersed in water and observed the temperature change; he replaced the water by other fluids; he rubbed metal plates together. Always he found that, in a given system, the temperature rise was the same whatever form of energy transformation was responsible. These experiments were extended by other physicists but always with the same result. In consequence, scientists accept that energy is conserved in all such changes. The law of conservation of energy rests firmly on a mass of experience. If we can find some model for matter which can accommodate the internal energy store which we have associated with the temperature rise, then we shall believe even more strongly in the validity of this law.

These experiments revealed that the transfer of a quantity of energy ΔE to a system—whether by mechanical, electrical, or other means—changed the internal energy of that system by an equivalent amount ΔU,

$$\Delta E = \Delta U.$$

Since ΔE is normally measured directly or indirectly as a certain quantity of work, W, it follows that

$$\Delta U = W.$$

A process such as this which involves the performance of work, changes in the internal energy and temperature, is known as an *adiabatic* process.

Now consider the case of the water which was receiving its additional internal energy from the hot flame of the burner. This was an example of a flow of heat without the involvement of work. The only indication we have of change is that the temperature of this water rose. If the rise in temperature in each of the two similar systems in our thought experiment was the same then the internal energy of the two masses of water increased by the same quantity of energy. We may say that an amount of heat (ΔQ) entered the system and that

$$\Delta Q = \Delta U.$$

These two processes are very simple ones and are rarely met with in practice. A much more usual situation involves a working substance such as a gas which is contained within a cylinder and enclosed by a piston so that it can expand and contract. Consider such a system and imagine that heat can be supplied to it (see Fig. 10.1a).

As heat energy ΔQ enters the gas, the internal energy rises, the temperature increases and the pressure also increases. The piston therefore moves outwards to allow the gas to expand. In so doing it exerts a force and therefore does work externally. It may, for example, store up some gravitational potential energy (Fig. 10.1b). Since this mechanical energy is being extracted from the system it must be written $-\Delta W$. Now, if the law of conservation is to hold, it follows that

heat energy entering the system	−	mechanical work extracted from the system	=	change in internal energy of the system.

$$\Delta Q - \Delta W = \Delta U,$$

or

$$\Delta Q = \Delta U + \Delta W. \qquad (10.1)$$

This statement is a formulation of the law of conservation of energy and is generally known as the *first law of thermodynamics*.

Chapter 11
HEAT CAPACITIES

Consider two isolated systems A and B (Fig. 11.1). System A has internal energy U_1 and is at temperature T_1. System B has internal energy U_2 and is at temperature T_2.

Whilst remaining isolated from their surroundings they are placed in good thermal contact. Energy is transferred as thermal equilibrium is attained and then system A has internal energy U_3 at temperature T whilst system B has internal energy U_4 at the same temperature T.

From the law of conservation of energy

$$U_1 + U_2 = U_3 + U_4,$$

and so

$$U_1 - U_3 = U_4 - U_2,$$

which may be written

$$(\Delta U)_A = -(\Delta U)_B. \qquad (11.1)$$

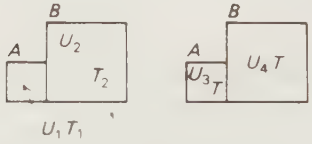


Fig. 11.1

Since the changes are adiabatic (i.e., no heat energy is transferred to or from the systems from any external system), we may use the first law of thermodynamics to express Eq. (11.1) as

$$-(\Delta Q)_A = (\Delta Q)_B.$$

That is, the heat lost by A is equal to the heat gained by B in this process.

During the transfer of heat from A to B, the temperature of A falls and that of B rises. Now experiment shows that, for a particular range of temperatures, the heat transfer from A is proportional to the temperature fall in A and similarly, the heat transfer to B is proportional to the temperature rise in B,

$$(\Delta Q)_A \propto (\Delta T)_A.$$

So the ratio $\Delta Q/\Delta T$ is a constant for this range of temperatures. It is called the *heat capacity* of the body (C) and is measured in joules per K.

Table 11.1

Substance	Temperature range		Specific heat capacity (c_p) at constant pressure J kg^{-1} K^{-1}	Molar heat capacity (C_p) at constant pressure J mol^{-1} K^{-1}
	K	°C		
Ice	263	−10	2100	37.7
Water	283	10	4190	
	288	15	4186	75.5
	303	30	4177	
	373	100	4210	
Water vapour	373	100	2030	36.5
Aluminium	290-373	17-100	908	24.5
Iron	290-373	17-100	473	26.4
Copper	20		12.6	
	60		121	
	80		197	
	200		368	
	300	27	389	24.7
Silver	288-373	15-100	235	25.3
Lead	293-373	20-100	130	26.9
Sodium chloride	283-303	10-30	879	50.2

Experiment also shows that the heat capacity of the body is proportional to its mass. It is therefore convenient to introduce a *specific heat capacity* (i.e., the heat capacity per unit mass). This may be referred either to a mass of one kilogramme of the substance or to the mass of a mole of the substance. The mole is the amount of substance of a system which contains as many elementary units as there are carbon atoms in 0.012 kilogrammes of carbon–12.

$$\text{Specific heat capacity } (c) = \frac{1}{m}\frac{\Delta Q}{\Delta T}. \qquad (11.2)$$

Some typical values are given in Table (11.1).

This table illustrates clearly that the specific heat capacities of water and copper are a function of temperature. Presented graphically as in Fig. 11.2, the values for water show a minimum at a temperature of about 35°C whilst those for copper (Fig. 11.3) show how the specific heat capacity rises to a value of about 25 J mol^{-1} K^{-1} and changes but little as the temperature is further increased. This is a typical behaviour for the solid elements (see the discussion of the law of Dulong and Petit in Unit 8).

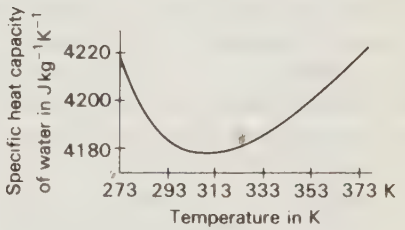


Fig. 11.2 The specific heat capacity of water as a function of temperature.

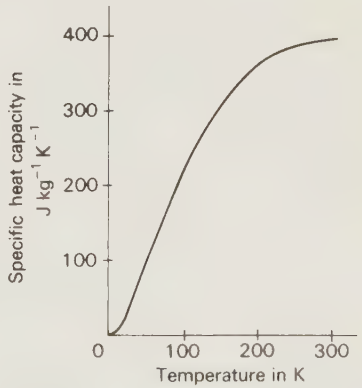


Fig. 11.3 The specific heat capacity of copper as a function of temperature.

For energy transfers of the type discussed in this section it follows that

$$\Delta Q = cm\Delta T. \qquad (11.3)$$

This provides a very convenient approach to the measurement of the heat transferred. The energy transfer (heat) when a body of mass m and specific heat capacity c has its temperature changed by ΔT is $cm\Delta T$. This is the foundation of the science of calorimetry, a foundation which was laid by Joseph Black in the eighteenth century. Black's investigations into the heat exchanges between hot and cold bodies when mixed together, (the method of mixtures) and into the change of phase which occurs when ice melts to water, were based upon a different model of heat from that which we have used. Essentially, heat was then regarded as a fluid (called caloric) which could flow from a hot body into a cold body. This caloric model was satisfactory for the changes to which it was applied. The first substantiated doubts about the model, which eventually led to its supersession by an energy theory, originated in the experiments of Rumford in 1798. Count Rumford (born Benjamin Thompson in Massachusetts in 1753) showed that the amount of heat generated when a cannon was bored out with a blunt drill was proportional to the mechanical energy transferred. As we have seen, this is the style of experiment which was perfected by J. P. Joule later in the nineteenth century and which led to the complete replacement of the caloric model for heat by the energy model.

Problem 1 An aluminium kettle weighing 1 kg, containing 1 kg of water at a temperature of 20°C, is placed on a gas ring and the gas is lit. The water is observed to boil after 230 s. How much energy has been transferred to the kettle and its contents?

If the specific heat capacity of water in this temperature range is taken to be 4200 J kg^{-1} K^{-1} and of aluminium 910 J kg^{-1} K^{-1} the energy which must be supplied to raise the temperature of the water from 20°C to 100°C is $4200 \times 1 \times (373 - 293)$ joules or

336 000 joules.

Similarly the energy required to raise the temperature of the kettle itself to 100°C is $910 \times 1 \times (373 - 293)$ joules or

72 800 joules.

The total energy supply is 408 800 joules.

Problem 2 What is the rating in kilowatts of an electric kettle of equivalent performance?

The gas was supplying energy at a rate of 408 800/230 joules per second. This is 1774 J s^{-1} or, approximating, 1.7 kilowatts.

Have a look at the plate on an electric kettle and see if the rating is anything like this.

Chapter 12
A SECOND LAW?

12.1 The second law
12.2 The Carnot cycle
12.3 The efficiency of a Carnot engine
12.4 The thermodynamic temperature scale
12.5 The thermodynamic scale and the ideal gas scale
12.6 The place of thermodynamics

A petrol engine may have an efficiency (i.e., output of work/input of energy) of about 25%; a diesel engine about 35%; a steam engine might reach 15%; a steam turbine about 40%. All four are devices for the transformation of energy from one form to another more useful form; all four use energy in transit (heat), develop an output of work, and emit hot exhaust gases; all four are heat engines. Is this low efficiency a consequence of incompetence; or is it due to some natural limitation? The first law of thermodynamics placed no such limitation on the possibility of transforming energy from any one form to any other form. Nothing in that law suggests that the total supply of energy cannot be converted to work. It is experience which suggests this.

This theoretical problem about the heat engine was first approached by Carnot, a French engineer, in 1824. His work and that of such men as Clausius and Kelvin led to the realization that there did exist a further natural law and that this law imposed a natural limitation on these energy transformations.

Before examining this limitation it will be necessary to look at some of the simple processes which the working substance within an engine might undergo. We have already used the term *adiabatic* to describe a process in which no transfer of energy as heat occurs. A further term *isothermal* is used to describe a process which occurs at a constant temperature. They are contrasted in Table 12.1.

Table 12.1

	Heat transfer ΔQ	Internal energy change ΔU	Work W	Temperature change ΔT
Adiabatic process	0	ΔU	W	ΔT
Isothermal process	ΔQ	ΔU	W	0

It is useful to indicate changes such as these on a graph. Where the working substance is a fluid, a graph of pressure against volume (Fig. 12.1) is suitable

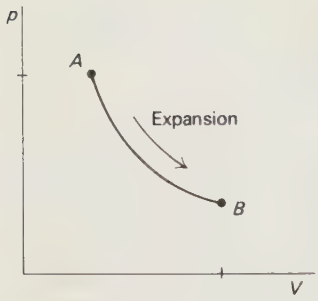


Fig. 12.1 Isothermal expansion.

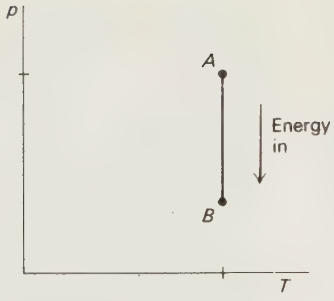


Fig. 12.2 Isothermal expansion.

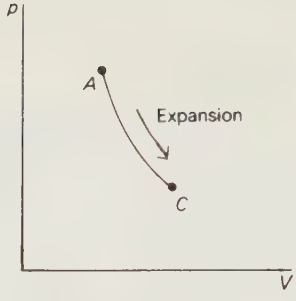


Fig. 12.3 Adiabatic expansion.

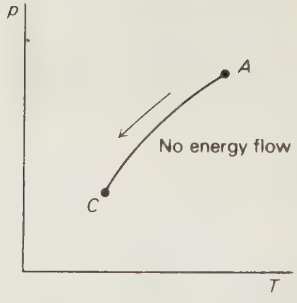


Fig. 12.4 Adiabatic expansion.

and one of pressure against temperature (Fig. 12.2) is also informative. Suppose that the working fluid has a pressure p_A and volume V_A at a temperature T_2.

This state is indicated by the point A. If now the fluid expands isothermally, the pressure diminishes and the state of the fluid may follow a line such as AB to the new state p_B, V_B, T_2. During this process energy must flow into the system from the surroundings for otherwise the temperature will fall and the process will not be isothermal.

If, however, the fluid expands adiabatically, (Figs. 12.3 and 12.4) work will be done at the expense of the internal energy and the temperature must fall. In this change of state from A to C, the curve AC is a steeper one than the curve AB.

Under certain conditions such processes may be reversible: that is, they may be reversed so that the system and its surroundings are restored to their original states without causing any changes elsewhere. Such a reversible change is a rare thing indeed. Consider its implications. The system in its initial state is in equilibrium. In its final state it is equilibrium. At all intermediate states it is in equilibrium. How then can a change be possible? Only by a sequence of changes which occur infinitely slowly.

No real processes can attain this perfection: they must be considered irreversible. Nevertheless, the reversible process is a far easier one to manipulate in the mind than are such very real irreversible processes as the diffusion of gases, chemical reactions, the solution of a salt in water, the radiation of heat, and so on. The reversible process is yet another model for reality and a convenient one to use in our discussion of the inefficiency of the heat engine.

12.1 THE SECOND LAW

If work is done against friction, say, then an equivalent amount of heat is made available. The work results in the temperature of the bodies rising, that is to say, their internal energy increases. Now we may prevent this temperature rise by immersing the rubbing bodies in a current of air or water. The energy is transferred to the fluid and taken away. If enough fluid is used the resulting temperature rise is not detectable and we have a complete transformation from work to heat with no change in the state of the system. All experience indicates that it is not possible to reverse this process and to convert that heat back into work at that temperature. There is nothing in the first law of thermodynamics to prevent us from recovering that energy; but, in fact we cannot. In the world as it exists, it is not possible to extract energy from the vast stores of internal energy existing in the air, the earth or the seas and to make use of it. If we could do this, we could use friction to re-transform this useful energy to the internal energy of some part of the air, or the earth or the seas and so raise the temperature of that part. That this is not possible is a statement of the second law of thermodynamics. Following Kelvin we may state this formally:

> "A transformation whose only final result is to transform into work heat from a source which is at a single temperature is impossible."

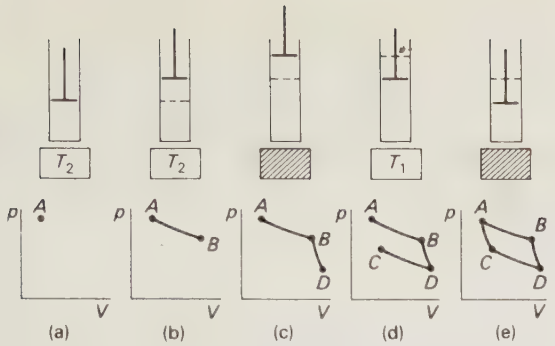


Fig. 12.5

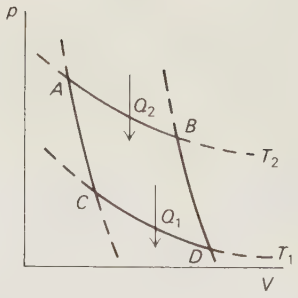


Fig. 12.6 p–V diagram for a Carnot cycle.

So we see that not all the changes which might satisfy the energy accounting system of the first law are actually possible.

12.2 THE CARNOT CYCLE

Let us imagine an engine which consists of a piston moving freely in a cylinder. The piston and the wall of the cylinder are non-conducting, but heat may be transferred through the base of the cylinder. The cylinder contains a suitable working substance—a fluid. This cylinder has been standing on a source of heat at a temperature T_2 and thermal equilibrium has been reached. The pressure and volume of the fluid are p_A and V_B and the temperature is T_2 (A in Fig. 12.5a).

Let us increase the volume of the fluid to V_B by raising the piston in such a way that this change is a reversible, isothermal one (B in Fig. 12.5b). Energy of amount Q_2 flows into the system and the fluid adopts the state p_B, V_B and T_2.

Let us place the cylinder upon a stand of insulating material (D in Fig. 12.5c). This ensures that the next change is an abiabatic one. Let us increase the volume reversibly by V_D. As we do so the temperature falls to T_1 but energy neither enters nor leaves. The new state will be p_D, V_D, and T_1.

Next let us transfer the cylinder to a second source of heat at a temperature of T_1, which is lower than T_2. There we will compress the fluid isothermally and reversibly until its volume has decreased to V_C. Energy of amount Q_1 will be transferred to the source at temperature T_1 (Fig. 12.5d).

Finally, let us replace the cylinder on the insulating stand and compress the fluid abiabatically and reversibly until its temperature reaches T_2 again. The fluid has now been restored to its initial state (Figs. 12.5a and e).

During the cycle, energy has been transferred to the system which has absorbed Q_2 from the hotter source and surrendered Q_1 to the cooler source. Moreover, work will have been done and from the first law

$$W = Q_2 - Q_1.$$

Part of the energy has been transformed into work and part is surrendered to the cooler source.

The p–V diagram for the complete cycle is shown in Fig. 12.6. It has been drawn for a gas; but similar diagrams may be drawn for other Carnot cycles. These may include a liquid-vapour mixture and a reversible electric cell such as an accumulator.

The efficiency of the simple ideal reversible engine is found by comparing the work output with the energy input.

$$\text{Efficiency} = \frac{W}{Q_2} = \frac{Q_2 - Q_1}{Q_2}$$

$$= 1 - \frac{Q_1}{Q_2}. \qquad (12.1)$$

The cycle is reversible, for all the processes involved are themselves reversible. Acting in the

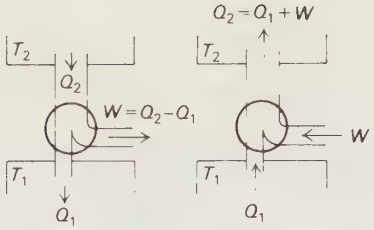


Fig. 12.7 Carnot engine Carnot refrigerator

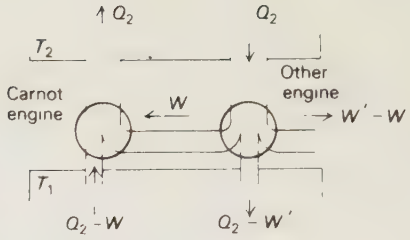


Fig. 12.8 Two coupled engines.

reverse direction, the cycle is a refrigerating one; mechanical energy is absorbed, heat Q_1 is taken in at the lower temperature T_1 and heat Q_2 transferred at the higher temperature T_2. We may represent the engine and the refrigerator diagrammatically (Fig. 12.7).

12.3 THE EFFICIENCY OF A CARNOT ENGINE

It was from a consideration of the efficiency of the Carnot engine that Lord Kelvin was led to formulate the thermodynamic scale of temperature.

Suppose we have two engines—one a Carnot engine, the other different, but constructed to work between the same two temperatures as the Carnot engine. The two cycles are compared in Table 12.2.

Table 12.2

Carnot engine	Other engine
Takes in Q_2 at temperature T_2	Takes in Q_2 at temperature T_2
Does work W	Does work W'
Surrenders $Q_2 - W$ at temperature T_1	Surrenders $Q_2 - W'$ at temperature T_1

Let us imagine that W is less than W'.

Then the work W' provided by the second engine could be used to drive the Carnot engine in reverse and also provide $(W' - W)$ for other purposes. So driven, the Carnot engine will absorb heat from the source at T_1 and deliver Q_2 back to the heat source at T_2. The original condition is restored. Figure 12.8 represents this double cycle.

The net effect is that the heat source T_2 has the same amount of energy after the cycle as it had before. Heat source T_1 has gained

$$(Q_2 - W') - (Q_2 - W),$$

or

$$-(W' - W).$$

Now W' has been assumed to be larger than W; so energy equivalent to $(W' - W)$ has been extracted from the heat source T_1. In other words, useful work has been obtained from the system at the expense of a heat source at a single temperature—and this is contrary to the second law of thermodynamics. It follows that no engine which absorbs a given amount of energy from a source at a high temperature and delivers it to another at a lower temperature can do more work than a reversible engine working between the same conditions. This argument assumes reversibility which, as we have seen, is our model for reality in the case of heat engines. We may now state Carnot's theorem; "No engine operating between two heat sources can be more efficient than a Carnot engine operating between the same two heat sources."

Any two Carnot engines operating under these conditions must have the same efficiencies for by the same argument, since the engines are reversible, not only is W' not greater than W but, also, W is not greater than W'. So

$$W = W',$$

and the efficiencies are also equal. This is true irrespective of engine design or working substance chosen.

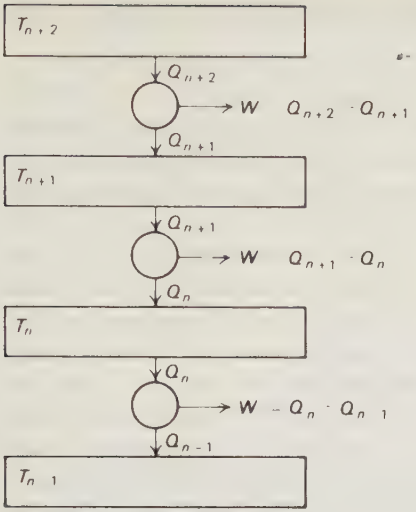


Fig. 12.9 A series of heat engines.

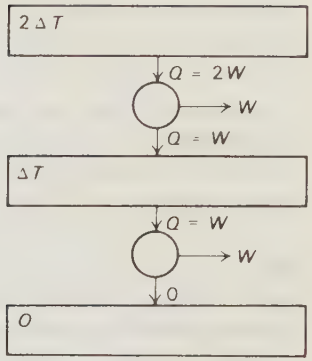


Fig. 12.10 Heat engines near the absolute zero of temperature.

It was this revelation that the efficiencies of all Carnot engines working between the same temperatures depended only on the two temperatures irrespective of the working substance, which led Kelvin to propose a temperature scale which was independent of the properties of any particular substance.

12.4 THE THERMODYNAMIC TEMPERATURE SCALE

We have seen that the efficiency of an engine is given by

$$1 - \frac{Q_1}{Q_2}.$$

If the efficiencies of all Carnot engines working between the same two temperatures are the same, the ratio Q_1/Q_2 must depend in some way on the temperatures of the two heat sources. Kelvin suggested that a temperature scale could be defined so that

$$\frac{Q_1}{Q_2} = \frac{T_1}{T_2}.$$

Completely to specify this scale let us assign (as before) the value 273.16 to the temperature of the triple point of water. Then for a Carnot engine operating between a source at a temperature T and a source at a temperature of 273.16 K (i.e., the triple point of water) we shall have

$$\frac{Q_1}{Q_2} = \frac{T}{273.16},$$

and so

$$T = 273.16 \frac{Q_1}{Q_2} \text{ K}. \qquad (12.2)$$

This defines a temperature scale which is independent of the characteristics of any working substance used in the engine. In fact, it can be shown that this thermodynamic temperature scale and the ideal gas scale are identical.

This scale may be interpreted in physical terms by considering a whole series of ideal, reversible heat engines, connected together so that the heat surrendered by one is accepted reversibly by its neighbour (Fig. 12.9).

If each of these engines provides the same amount of work, the differences of temperature between

adjacent heat sources ($T_{n+2} - T_{n+1}$), ($T_{n+1} - T_n$), ($T_n - T_{n-1}$) etc., are equal.

At some temperature (known as the absolute zero of temperature) no waste heat remains to be surrendered. The final heat engine in the series which provides the work W must itself receive energy W for no energy is surrendered. The efficiency of this last engine is unity (Fig. 12.10).

This approach to a scale of temperature which is not dependent upon the properties of any particular substance gives a unique meaning to the term *absolute zero of temperature.* At this temperature, an ideal engine can discharge no more heat.

This thermodynamic scale tells us how hot an object is in terms of the heat absorbed by a reversible engine working between the temperature of the object and this zero of temperature.

Temperatures on this scale are expressed in Kelvins (K). The Kelvin is defined as the fraction 1/273.16 of the thermodynamic temperature of the triple point of water.

12.5 THE THERMODYNAMIC SCALE AND THE IDEAL GAS SCALE

Numerically the thermodynamic temperature and the ideal gas temperature of the same body are equal, although it is beyond the scope of this book to demonstrate this. Consequently it is permissible to measure thermodynamic temperatures with a gas thermometer, although corrections have to be made to the indicated readings, because the filling pressure is finite.

12.6 THE PLACE OF THERMODYNAMICS

In this unit we have used some of the fundamental ideas of thermodynamics to show how an absolute scale of temperature (that is, a scale which is independent of the properties of any particular substance) may be arrived at. The approach we have adopted depends on the work done in the nineteenth century on the theory of heat engines. It is no coincidence that the century which saw the development of the steam engine (invented in the eighteenth century) and the invention of the hot air and the internal combustion engines should also see the development of thermodynamics. From that development of a principle first detected in connection with the crude steam engines of Newcomen and Watt, there emerged a clear understanding of the first and second laws of thermodynamics.

Later in that century a parallel development occurred in the field of statistical mechanics. We shall see in Unit 4 how the kinetic theory of gases led to an understanding of the behaviour of matter in the macroscopic scale through an analysis of its behaviour on the microscopic scale. This study led to an extension of the science of thermodynamics to all branches of science, including the life sciences and cosmology. It raised fundamental questions about man's employment and deployment of his energy reserves: it led to rapid advances in chemical technology and chemical theory: it inspired developments in high and low temperature research: it gave rise to the quantum theory: it may be thought of as pervading all science in a unique manner.

PROBLEMS: Unit Three

3.1 A resistance thermometer has a resistance of 10 ohms at the triple point of water. What is its resistance at an indicated temperature of 373.16 K?

3.2 In Joule's experiments with a paddle wheel immersed in water and driven by a pair of falling masses, each of about 14 kg, twenty successive falls of about 2 m were arranged between each reading of the thermometer in the water. The total mass of water was about 7 kg. Estimate the temperature rise. (Specific heat capacity of water is 4200 J kg^{-1} K^{-1}.)

3.3 An air-gun pellet moving at 100 m s^{-1} becomes embedded in a massive wooden block. Estimate the rise in temperature of the pellet noting all the assumptions you maké.

3.4 Estimate the minimum velocity of the air-gun pellet of question 3.3, if it is to melt on becoming embedded in the block. (Melting point of lead is 328° C.)

3.5 Hot coffee is sealed up in a vacuum flask and then shaken vigorously. Assuming that the coffee is an isolated system, would you expect its temperature to rise? Explain.

3.6 In changing from the liquid phase to the solid phase at a temperature of 0°C, water loses 3.4×10^5 J kg^{-1}. If a tank containing 6000 kg of water at 0°C is losing energy at 300 J s^{-1} how long will it be before all the water is frozen?

3.7 An electric kettle working at a rate of 1.5 kW is used to raise the temperature of 2 kg of water from 293 K to the boiling point. The mass of the kettle is 0.5 kg and its mean specific heat capacity 900 J kg^{-1} K^{-1}. How long will this process take? You may assume that 2% of the energy is lost.

3.8 Devise an experiment to investigate how the temperature change in a body to which energy is transferred by a heater is related to the mass of the body and to the material of which the body is made.

3.9 What effect does leaving open the door of a refrigerator have on the temperature of the room in which it is kept?

3.10 If work is done on a gas by compressing it into a cylinder, the temperature of the gas rises. After a time, the temperature falls to that of the surroundings and yet the compressed gas can still transfer energy by expanding. Where does this energy come from?

3.11 Show that the statement "the efficiency of an ideal heat engine is

$$1 - \frac{Q_1}{Q_2},$$

is equivalent to an efficiency of

$$\frac{T_2 - T_1}{T_2}.$$

A simple engine operates between temperatures of 200° C and 100° C. What is its efficiency? What steps could be taken to double this figure?

What are the conditions under which such an engine could operate with a thermal efficiency of 100%? Comment.

3.12 A Carnot engine may have its efficiency improved by increasing the temperature of the hotter source whilst keeping that of the cooler source constant or by decreasing the temperature of the cooler source whilst keeping that of the hotter source constant. Which technique would be the most effective?

3.13 Sketch, using the same axes, pV diagrams for an ideal gas which is initially at a pressure p and a volume V and which then undergoes (a) an isobaric expansion (i.e., an expansion at constant pressure), (b) an isothermal expansion, (c) an adiabatic expansion, (d) an isochoric expansion (i.e., an expansion at constant volume). In each case, write down a statement applying the first law of thermodynamics. Suggest possible examples for each process.

3.14 How far is the term "heat pump" applicable to a refrigerator?

3.15 What is the efficiency of the next three engines in the sequence shown in Fig. 12.10? State a rule for the efficiency of the engines in the series.

BIBLIOGRAPHY: Unit Three

An informal and yet rigorous treatment of the basic principles of thermodynamics as seen from the view-point of the chemist is contained in

Bent, H. A. (1965). *The Second Law*, Oxford University Press.

An introductory text relating the historical development of thermodynamics in theory and practice is

Sandfort, J. F. (1964). *Heat Engines*, Heinemann.

Our views on the meaning and use of the term "heat" are changing rapidly and few elementary sources reflect the usage of the thermodynamicist. Indeed this may prove to be impossible. A good general reference is

Sears, F. W. and Zemansky, M. W. (1970). *University Physics*, Addison-Wesley. 4th edition.

For a useful discussion of the concept of temperature at this level see

Zemansky, M. W. (1964). *Temperatures Very Low and Very High*, Van Nostrand.

A thorough-going treatment and a standard reference at the undergraduate level is

Zemansky, M. W. (1968). *Heat and Thermodynamics*, McGraw-Hill.

UNIT FOUR
A Model for Matter

13 A Model for a Gas: A Second Look
14 A Model for a Solid
15 Crystalline Structure
16 Applying the Model for Matter

Chapter 13
A MODEL FOR A GAS: A SECOND LOOK

13.1 A simple mathematical analysis of the model
13.2 An equation of state
13.3 Some implications of the model
13.4 Molecular speeds
13.5 Temperature and kinetic theory
13.6 Avogadro's number
13.7 The atomic model

Now that we have a working knowledge of dynamics, it becomes possible to quantify the model for a gas proposed in Unit 1 and, in particular, to examine its implications mathematically. It will be recalled that the model was based on a system of randomly moving, widely spaced particles. It was assumed that the collisions between these massive point particles were elastic and that the interparticle forces were only observable at extremely short ranges. The only structure we demand for these particles is that they should not readily absorb energy.

13.1 A SIMPLE MATHEMATICAL ANALYSIS OF THE MODEL

So far we have only supposed that the model behaves like a gas in a container, exerting pressure by virtue of the impact of the particles on the wall. We cannot suppose this model behaves like a gas unless we can show that it does so in a quantitative way as well, for example, that pressure and volume in the model are related in the same way as they are in a real gas.

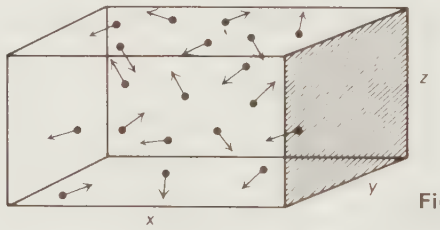


Fig. 13.1

To do this, we must calculate the pressure these flying particles will exert. To do this rigorously is quite difficult, but some of the difficulties can be avoided by taking a few reasonable short-cuts.

Imagine the "gas" to be confined in a rectangular box whose sides are respectively x, y, and z metres long, and that it has been there long enough to be in a steady state (Fig. 13.1). At any instant, the particles are moving in any one of a variety of directions as the arrow on each particle is intended to show. Let us concentrate attention on the impact of the particles on the shaded end of the box. The particles are approaching this end of the box from all possible directions. All bounce off the end, exerting a certain force on the

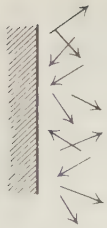


Fig. 13.2

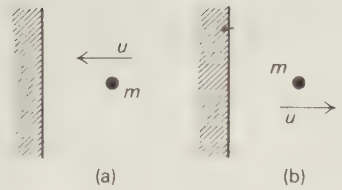


Fig. 13.3

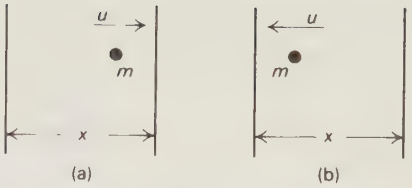


Fig. 13.4

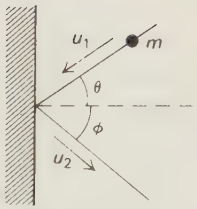


Fig. 13.5

face of the box as they do so (Fig. 13.2). How can we set about finding this force?

Consider first of all a single particle of mass m and travelling with velocity u along a line perpendicular to the wall (Fig. 13.3a).

Let us assume that in this case, it bounces off in the opposite direction. As the particle approaches the wall, it slows up as the short range intermolecular force of repulsion comes into play, and its momentum is reduced to zero. At this instant, the kinetic energy of the particle has been entirely converted to potential energy. The particle is accelerated until it leaves the wall with a velocity u exactly reversed from its approach velocity (Fig. 13.3b). Its final kinetic energy is the same as its initial kinetic energy but its momentum is now mu in the opposite direction to its initial momentum. It has suffered a momentum change of

$$mu - (-mu) \qquad \text{or} \qquad 2mu.$$

To bring about this change, the wall must have provided an impulse—a force acting for a short time. We cannot say how big this force is for a single particle hitting the wall. It all depends on the time during which the intermolecular forces between the wall and the particle act. This is a simple application of Newton's second law of motion which states that the force acting on a body is proportional to its rate of change of momentum.

If this particle were the only one in the box, it would eventually collide with the opposite wall and could bounce back again towards the face of the box under consideration (Fig. 13.4).

The next collison on the end of the box to the left will take place $2x/u$ seconds later. The momentum of the particle is thus changed by $2mu$ at the face of the box on the left every $2x/u$ seconds, if it always bounces off in the opposite direction from its approach. The rate of change of the momentum at the face is

$$\frac{2mu}{\frac{2x}{u}} = \frac{mu^2}{x} \text{ momentum units per unit time.}$$

By Newton's second law of motion, this is the average force acting on the face—that is mu^2/x newtons—if we imagine the forces resulting from this series of impacts to be "smeared out" to a constant value.

So far we have considered the impact of one particular particle rattling back and forth between two opposite faces of the box. What happens when the box contains N particles all moving with different speeds and in different directions?

Consider again the elastic impact of a single particle on the face of the box. But this time suppose it moves at an angle θ to the perpendicular direction (Fig. 13.5). If energy and momentum are both conserved in the collision:

$$mu_1 \sin\theta = mu_2 \sin\phi,$$
$$mu_1 \cos\theta = mu_2 \cos\phi,$$
$$\tfrac{1}{2}mu_1^2 = \tfrac{1}{2}mu_2^2.$$

From this it follows that

$$\theta = \phi \qquad \text{and} \qquad u_1 = u_2.$$

The only velocity of interest, as far as the impact with the wall is concerned, is the perpendicular component $u\cos\theta$. Let us suppose that this direction is the x-direction of a set of rectangular axes. We will call

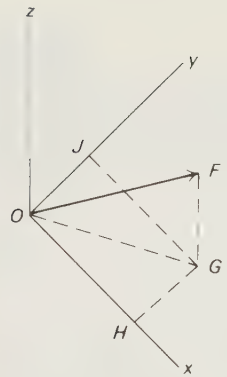


Fig. 13.6

this the x-component of the velocity in the x-direction.

We can resolve the velocity of any particle into three components, u_x, u_y, and u_z. These are related to each other and to the velocity u of the particle by an extension of Pythagoras' theorem.

Ox, Oy, and Oz are three mutually perpendicular axes (Fig. 13.6). OF represents a velocity u. FG is a perpendicular from F to the plane Oxy so that OGF is a right angle and, by Pythagoras' theorem

$$OF^2 = OG^2 + GF^2,$$

GH and GJ are perpendicular to the Oxz and Oyz planes respectively and

$$OG^2 = GJ^2 + GH^2,$$

whence

$$OF^2 = GJ^2 + GH^2 + GF^2.$$

But GJ, GH, and GF are the components of the velocity u in the x, y, and z directions. So

$$u^2 = u_x^2 + u_y^2 + u_z^2.$$

For each of N particles in the box, all of which are moving with a variety of velocities, we may write

$$u_1^2 = u_{x_1}^2 + u_{y_1}^2 + u_{z_1}^2$$
$$u_2^2 = u_{x_2}^2 + u_{y_2}^2 + u_{z_2}^2$$
$$\cdots\cdots\cdots$$
$$u_n^2 = u_{x_n}^2 + u_{y_n}^2 + u_{z_n}^2.$$

Add:

$$(u_1^2 + u_2^2 + \cdots u_n^2) = (u_{x_1}^2 + u_{x_2}^2 + \cdots u_{x_n}^2) + (u_{y_1}^2 + u_{y_2}^2 + \cdots u_{y_n}^2) + (u_{z_1}^2 + u_{z_2}^2 + \cdots u_{z_n}^2).$$

Divide throughout by N:

$$\frac{(u_1^2 + u_2^2 + \cdots u_n^2)}{N} = \frac{(u_{x_1}^2 + u_{x_2}^2 + \cdots u_{x_n}^2)}{N} + \frac{(u_{y_1}^2 + u_{y_2}^2 + \cdots u_{y_n}^2)}{N} + \frac{(u_{z_1}^2 + u_{z_2}^2 + \cdots u_{z_n}^2)}{N}.$$

This may be written:

$$\overline{u^2} = \overline{u_x^2} + \overline{u_y^2} + \overline{u_z^2}.$$

The term $\overline{u^2}$ is the average of the square of the speed and is usually called the *mean square speed.* It is a meaningful average to use in this case.

Now collisions among the particles within the box will serve to maintain such an even distribution of the particles that

$$\overline{u_x^2} = \overline{u_y^2} = \overline{u_z^2}. \tag{13.1}$$

(This is our reasonable short-cut. The last few lines form an essential part of our analysis, but the appeal for their correctness is to common-sense. It seems reasonable that a large assemblage of particles should behave in this way—it is mathematically complex rigorously to show that they will: we must guard against the specious argument that "of course" the mean value of all the three components must be the same because we know that in a gas the pressure is the same on all the faces of the box. That is true in a gas. We are trying to show that our model behaves like a gas—we must not assume that it does in the process of setting up a proof.)

From Eq. (13.1), it follows that

$$\overline{u_x^2} = \overline{u_y^2} = \overline{u_z^2} = \tfrac{1}{3}\overline{u^2}. \tag{13.2}$$

For one particle, the force on the left-hand face of the box was mu^2/x. For N particles, the force is

$$\frac{N(mu_x^2)}{x} = \frac{Nm}{x}\,\frac{1}{3}\overline{u^2}. \tag{13.3}$$

The area of the face is yz. So the pressure, p, on the face is

$$\frac{\text{force}}{\text{area}} = \frac{Nm}{x}\,\frac{1}{3}\overline{u^2}\,\frac{1}{yz}. \tag{13.4}$$

Now, $xyz = V$, the volume of the box, so

$$pV = \tfrac{1}{3}Nm\overline{u^2}. \tag{13.5}$$

For a given mass of gas, N and m are unchanging and Nm = total mass of gas. So pV is a constant, provided $\overline{u^2}$ remains constant.

The product, $Nm\overline{u^2}$, which we may write as $2(\frac{1}{2}Nm\overline{u^2})$ reminds us of the form of the equation for kinetic energy (K.E. = $\frac{1}{2}mu^2$). If our particle model has any validity, the product $\frac{1}{2}Nm\overline{u^2}$ may represent the internal energy store of a gas. If so, it will remain constant unless energy is supplied either as a consequence of a temperature rise or by the application of work. This is encouraging. This model of a gas will obey Boyle's law (pV = constant at constant temperature) provided that we can, in some way, relate the product $(\frac{1}{2}Nm\overline{u^2})$ to temperature.

13.2 AN EQUATION OF STATE

Experimental evidence long ago established two further empirical laws descriptive of the behaviour of real gases. The first stems from the work of Charles (1787) and of Gay-Lussac (1802) and relates the volume change of a fixed mass of a gas, maintained at a constant pressure, to the temperature change recorded, say, in °C on the mercury-in-glass scale of temperature.

$$V = V_0(1 + \alpha\Delta t), \tag{13.6}$$

where V and V_0 are the final and the initial volumes of the gas, Δt the temperature change in °C and α is a constant. Interestingly enough, this constant has very nearly the same value for all gases at low pressure and is $0.003660°C^{-1}$.

The second law concerns the way in which the pressure of a fixed mass of a real gas, maintained at a constant volume, changes as the temperature changes. Using the same temperature scale as above, we find that

$$p = p_0(1 + \beta\Delta t), \tag{13.7}$$

where p and p_0 are the final and initial pressures, Δt the common temperature change in °C and β is a constant. This constant β proves to be virtually the same as α and neither differs by much from the fraction 1/273. This is hardly an accident, as we shall see.

We have already observed (in Unit 3) that gases at low pressures are by far the best thermometric substances and we defined temperatures in terms of a constant volume gas thermometer.

$$T_x = 273.16\,\frac{p_x}{p_{273.16}}\ \text{K}.$$

$p_{273.16}$ is the pressure at the triple point of water.

Experiments on gases at low pressures show that the product pV is related to the pressure by an equation of the form

$$pV = A(1 + B'P + C'P^2 + \cdots),$$

where A, B', C', etc., are known as the virial coefficients. Their values depend on the temperature and also on the nature of the gas used. At very low pressures, the product pV approaches the same value for all gases at the same temperature. Figure 13.7 shows some typical graphs for hydrogen, nitrogen, and oxygen at different temperatures. We can see that, as the pressure p tends to zero, the equation tends to the form

$$pV = A,$$

and A, the first virial coefficient, is seen to be independent of the nature of the gas and to be a function of the temperature alone. So

$$A = RT,$$

where R is a constant which can be determined experimentally. R is known as the universal gas constant and has the value 8.314 J mol^{-1} K^{-1}.

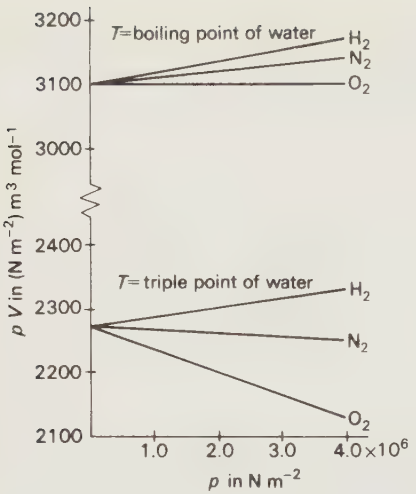


Fig. 13.7 Graphs of pV–p for different gases at the triple point and at the boiling point of water.

The equation of state for all gases under these conditions can be written

$$pV = RT. \tag{13.8}$$

This suggests that it would be valuable to imagine an "ideal gas" to which this equation of state would apply under all circumstances. For such a gas we should have

a) $p \propto 1/V$ at constant temperature (Boyle's law),
b) $p \propto T$ at constant volume,
c) $V \propto T$ at constant pressure.

Under ordinary conditions—i.e., for relatively small changes in pressure, volume and temperature—the behaviour of real gases differs but little from that of this ideal gas and the simple equation of state of Eq. (13.8) can often be applied.

13.3 SOME IMPLICATIONS OF THE MODEL

We have already developed an equation of state (13.5) for the particle model in the form

$$pV = \tfrac{1}{3}Nm\overline{u^2}.$$

Suppose we were to take a mole of the particles, i.e., 6×10^{23} particles, a number which we shall write N_A (Avogadro's number). Each of these particles had mass m and so, for this system

$$pV = \tfrac{1}{3}N_A m\overline{u^2}.$$

This equation may usefully be written in terms of density (ρ) which is relatively easy to measure. Rewriting the equation

$$p = \frac{\frac{1}{3}N_A m}{V}\overline{u^2}$$

we can see that the ratio $N_A m/V$ is total mass/volume or density and

$$p = \tfrac{1}{3}\rho\overline{u^2}. \tag{13.9}$$

If now we assume that the model may be applied to the air, say, for which ρ is 1.3 kg m^{-3} at ordinary atmospheric pressure (1.03×10^5 Nm^{-2}) it follows that

$$\overline{u^2} = \frac{3p}{\rho} = \frac{3 \times 1.03 \times 10^5 \ \text{Nm}^{-2}}{1.3 \ \text{kg m}^{-3}}$$
$$= 2.38 \times 10^5 \ \text{m}^2\,\text{s}^{-2}$$

and, taking the square root

$$\sqrt{\overline{u^2}} = 490 \ \text{m s}^{-1}.$$

This is the root of the mean square velocity and suggests that, if air can be regarded as made up of the sort of particles the model requires, these are moving at the remarkably high speed of about 500 m s^{-1}. How far does this deduction meet the facts?

Problem 1 The speed of sound in air is 330 m s^{-1}. How far does this confirm the correctness of the above arguments?

Even though the calculated speed is high, some of the particles must be travelling even faster than this for, in the process of colliding with other particles, the exchange of energy will be accompanied by changes in velocity. The model envisages that the particles

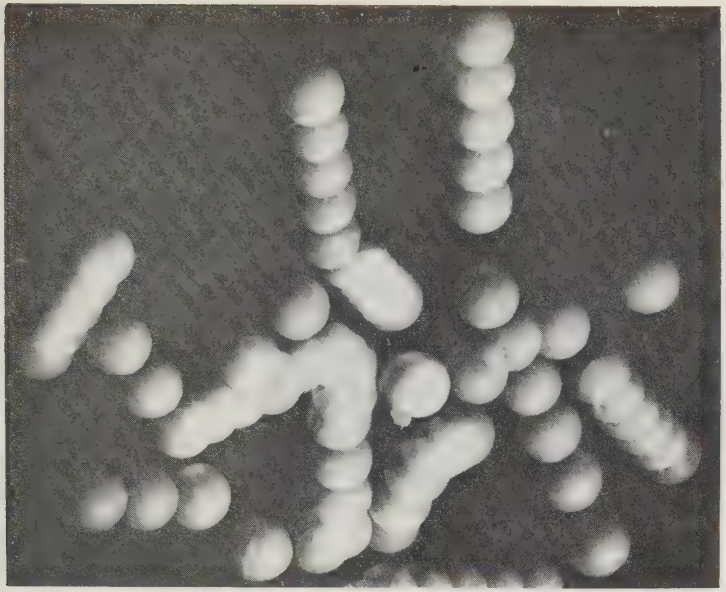


Fig. 13.8 Multiflash photograph of a group of balls on a vibrating table.

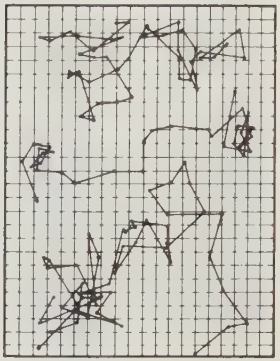


Fig. 13.9 Brownian movement. The position of three grains, suspended in water, observed at 30 s intervals. (After J. Perrin.)

travel at random, with a high average speed even at ordinary temperatures and with the actual velocities distributed in some way about this average value.

An old objection to this consequence of the model was concerned with the speed of diffusion of a gas into the air. Everyone knows that the smell of, say, town gas, takes time to spread through a room into which it is released. It is a reasonable guess, too, that convection is more effective in conveying the gas from the source to the nose than diffusion is. And yet the model suggests that the molecular velocities are very high.

This is an objection with which the model can deal quite easily and with only a small change in its construction. Consider the movement of a number of balls confined on a vibrating table (see Fig. 13.8).

This photograph was taken with the camera shutter open for a relatively long exposure time, whilst the balls were illuminated with a flashing strobe light. One can see at once that no ball travels very far in any one direction before colliding with another and suffering a change of velocity. The path of a single ball through the mass of balls must be a series of "dog-leg" sections.

Now the balls differ from point masses in one important respect—they collide rather more frequently because they have a finite size and, in consequence, "diffuse" through the available space quite slowly even though the average velocity is high. This provides the clue to the modification which must be made to the point mass model. We have merely to suppose that the particles of the gas have a small but finite size so that inter-particle collisions occur frequently.

A particle such as this will not travel very far from its starting point for the changes in direction which will occur with each collision are just as likely to move it one way as another. Indeed, it is perhaps surprising that it ever does move away from the starting point. The actual path followed is quite unpredictable but must be something like those pictured on the vibrating table and to the sort of observation which was recorded by Perrin in his study of the Brownian motion (Fig. 13.9).

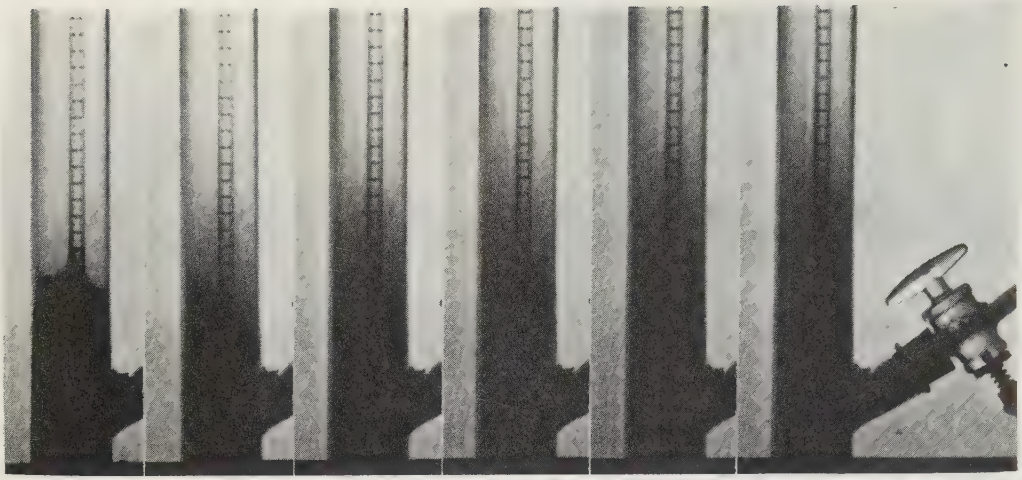


Fig. 13.10 Bromine diffusing into air at atmospheric pressure. The photographs were taken at intervals of 100 s. (Photographs by E.J.W.)

This motion is highly erratic—a typical random walk. As presented in Fig. 13.9, it is a two dimensional walk. But in the gas this must be imagined as a three dimensional random walk. In an experiment on the diffusion of the dense coloured gas, bromine, into air at ordinary temperature and pressure, the series of photographs in Fig. 13.10 reveals the slow spread of the advancing bromine front as it moves into the colourless air. Of course, the advancing front of the bromine reveals a gradual increase in the concentration of the bromine and the density of the colour, but estimates of a "half-dark" position suggest that the front is advancing at a rate of about 0.1 m in 500 s. In this time an average bromine molecule can be expected to travel a total distance of 500 u, where u is the magnitude of the *average velocity*, or *average speed*.

If the magnitude of the average velocity u is 200 m s^{-1}, this means that on average a path length of 10^5 metres has been folded up by the constant collisions into a distance of 0.1 m. The walk is indeed a random one.

Figure 13.11 represents Perrin's drawing of one typical walk.

Consider the first few steps of this walk. Let the molecule start out from O and end up at F.

It has travelled a distance OF by making a series of n steps, each of which can be resolved into components in the x and y directions as shown in Fig. 13.12. The final x component of OF is given by

$$x = x_1 + x_2 + x_3 + \cdots + x_n,$$

due allowance being made for signs, of course. The final y component of OF is given by

$$y = y_1 + y_2 + y_3 + \cdots + y_n.$$

The distance OF is given by

$$OF^2 = x^2 + y^2.$$

Now

$$\begin{aligned} x^2 &= (x_1 + x_2 + x_3 + x_4 + \cdots + x_n)^2 \\ &= x_1^2 + (x_1x_2 + x_1x_3 + \cdots + x_1x_n) \\ &\quad + x_2^2 + (x_2x_1 + x_2x_3 + \cdots + x_2x_n) \\ &\quad + x_3^2 + (x_3x_1 + x_3x_2 + \cdots + x_3x_n), \quad \text{etc.} \end{aligned}$$

The terms x_1x_2 are just as likely to be negative as positive. So, when very large numbers of steps are considered, the total sum of these terms is likely to be zero. Then

$$x^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 + \cdots + x_n^2$$

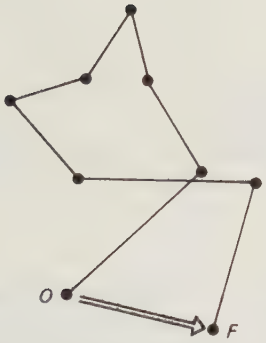


Fig. 13.11

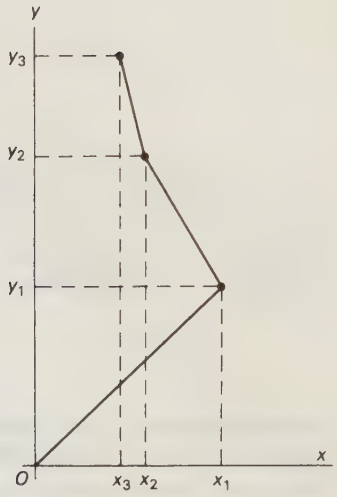


Fig. 13.12

and also

$$y^2 = y_1^2 + y_2^2 + y_3^2 + y_4^2 + \cdots + y_n^2$$

whence

$$OF^2 = (x_1^2 + y_1^2) + (x_2^2 + y_2^2) + \cdots + (x_n^2 + y_n^2).$$

If each individual step length is represented by λ_1, λ_2, etc., the theorem of Pythagoras gives

$$\lambda_1^2 = x_1^2 + y_1^2, \quad \text{etc.}$$

and

$$OF^2 = \lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \cdots + \lambda_n^2.$$

Taking λ^2 as the average of the squares of the value of all n individual steps

$$\lambda^2 = \frac{\lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \cdots + \lambda_n^2}{n}$$

$$= \frac{OF^2}{n}.$$

So $OF^2 = n\lambda^2$ in which λ is the square root of the average of the squares of the step lengths (a root mean square length). Hence

$$OF = \sqrt{n}.\lambda. \qquad (13.10)$$

This is a well-known statistical result and it can, in fact, be applied to the three-dimensional case as well as to the two-dimensional case considered above. It can be applied directly to the diffusion of bromine experiment.

The average step-length (mean free path) between collisions is also to be found by dividing the total distance $\bar{u}\Delta t$ travelled in time Δt by the number of steps n. So

$$\lambda = \frac{\bar{u}\ \Delta t}{n}.$$

From Eq. (13.10) we have

$$\lambda = \frac{\text{distance made good } (OF)}{\sqrt{n}}.$$

In the experiment, then

$$\lambda = \frac{200 \times 500}{n} = \frac{0.1}{\sqrt{n}}$$

whence

$$n = 10^{12}.$$

That is to say the bromine molecule which had travelled a total distance up the glass vessel of only 0.1 m had, en route, made 10^{12} collisions. This took

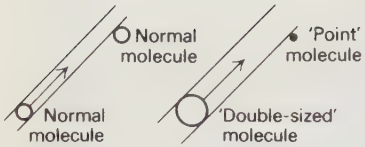


Fig. 13.13

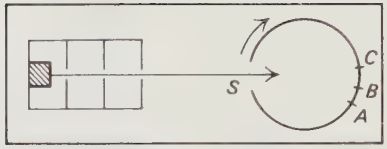


Fig. 13.14 Zartmann and Ko's experiment on the distribution of molecular speeds.

500 seconds and so the number of collisions per second must be 2×10^9. Having determined n to be 10^{12}, it remains to substitute in either of the equations for the mean free path λ to find that

$$\lambda = \frac{200 \times 500}{10^{12}} = 10^{-7} \text{ m}.$$

Problem 2 How long would it take an average molecule of a gas with a mean speed of 500 m s^{-1} to travel across a room which was 5 m wide?

This concept of a mean free path was first suggested by Clausius in 1858 and was associated with a finite size for the gas molecule. Clausius' picture of a collision between two gas molecules invoked both "not very" long range attractive forces and "very" short range forces of repulsion. He wrote:

> "Let us imagine two molecules moving in directions such that, if they preserved them unchanged, they would not strike one another but pass by at some distance. Two cases may here occur. If the distance is very small, the molecules which are drawn towards one another, even from some distance, by the force of attraction approach so closely that the repulsive force comes into play, and a rebounding of the molecules results. If the distance be somewhat greater, the paths of the molecules only suffer a certain change of direction through the attractive force, without the repulsive force being able to act. Finally, at still greater distances, the effect of the molecules upon one another may be altogether neglected.
> How great the distances must be in order that the one or other case might occur, could not be determined universally, even if we possessed exact knowledge of molecular forces; for the velocity of the molecules and the reciprocal inclinations of their paths are of influence. Nevertheless, mean values of these distances can be obtained."
> From Rudolf Clausius, "On the mean lengths of the paths described by the separate molecules of gaseous bodies", translated by F. Guthrie (1859).

Consider now a molecule in motion among its fellows following the series of dog-leg steps we have come to associate with any random walk. When it collides with another molecule the centres of mass of the two are, on average, one "molecular diameter", d, apart. We can envisage the process more clearly if we imagine the collision to be between one double sized molecule and a point molecule (Fig. 13.13).

The volume swept out by the "double-sized" molecule before it makes the collision is $\pi d^2 \lambda$.

In this volume, it finds, on average, just one point molecule. In other words, a typical molecule occupies a volume of

$$\pi d^2 \lambda.$$

Now liquid nitrogen (which is a safer material than liquid air) has a density of about 800 kg m^{-3} and, when it boils, its volume increases by a factor of 750. It would be sensible to assume that the molecules of the liquid nitrogen now occupy 750 times the space

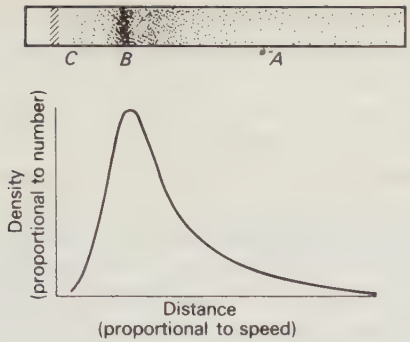


Fig. 13.15

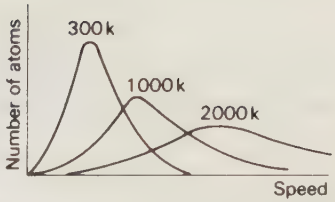


Fig. 13.16 Temperature and the distribution of molecular speed.

they occupied when in the liquid form. If a molecule of diameter d occupies a cubical box of side d (as though the molecules in the liquid are close packed like apples in a box) each molecule occupies $750d^3$ when in the gaseous state. So

$$750d^3 = \pi d^2 \lambda,$$

and

$$750d = \pi\lambda.$$

Taking $\lambda = 10^{-7}$ m, d is seen to be about 4×10^{-10} m or 4×10^{-1} nm.

13.4 MOLECULAR SPEEDS

Up to this point, the term *an average speed* of gas atoms or molecules has been used without careful specification of just what is to be understood by the term. The figure cannot be an arithmetical average for, if the motion is truly random, this must be zero. It is, as indicated above, the square root of the mean of the squares of the speed—this cannot be zero—and is meaningful in that it does tell us something about the way in which the speeds are distributed. Evidently, in a gas in which molecules are exchanging kinetic energies in a frequent series of collisions, individual molecular speeds must range from zero on the one hand to very high speeds on the other. Theoretical investigations into the distribution of speed among the molecules were carried out by James Clerk Maxwell in 1860 but experimental techniques for checking Maxwell's description of the distribution were not available until the 1920s when Stern devised a suitable technique which has been improved subsequently. In one method, used by Zartmann in 1931, atoms of bismuth were produced in a vacuum by heating in a furnace. A thin beam of moving atoms was obtained by allowing the main beam to pass through a series of slits. This method of obtaining a rectilinear beam of small dimensions is called *collimation*. The beam of vapour atoms then reaches a drum to which entry can only be gained through a slit S. This drum rotates at some 6000 rev min^{-1} and is lined with a sensitive film. As the slit S passes through the narrow beam a pulse of atoms enters the drum; and the speeds of the atoms will be distributed amongst a wide range. As the drum rotates the atoms move across its diameter, some taking much longer than others. The fastest atoms will strike the sensitive film at A, slower ones at B and the slowest at C (Fig. 13.14). The density of the film at any point will be a measure of the number of atoms striking that point.

The agreement between the observed results and the results expected from theoretical considerations was found to be remarkably good. This distribution of speeds, as shown in Fig. 13.15 is a familiar one. It is a form of the law of normal distribution (Gaussian distribution) which applies to so many things where chance is in control. For example, to the height of the men in a population, to the frequency distribution of a series of measurements and so on. The exact shape of the curve depends on the temperature and shifts as shown in Fig. 13.16 as temperature increases.

We have already noted that, for a mole of the particles,

$$pV = \tfrac{1}{3}N_A m\overline{u^2}, \qquad (13.11)$$

is an equation of state for the particle model of a gas and that a similar experimental equation of state is

$$pV = RT, \qquad (13.8)$$

where R is the molar gas constant. We may write (13.11) as

$$pV = \tfrac{2}{3}N_A \tfrac{1}{2}m\overline{u^2}.$$

If we assume that the kinetic energy of a particle is dependent upon temperature, and write

$$\tfrac{1}{2}m\overline{u^2} = \tfrac{3}{2}kT, \qquad (13.12)$$

where k is a constant then

$$\begin{aligned} pV &= \tfrac{2}{3}N_A \tfrac{3}{2}kT, \\ &= N_A kT. \end{aligned} \qquad (13.13)$$

This is of the same form as the equation of state derived experimentally. To make the equations identical we must write

$$R = N_A k$$

or

$$k = \frac{R}{N_A} \qquad (13.14)$$

k is known as Boltzmann's constant and is 1.38×10^{-23} J K^{-1}.

The kinetic energy of the particles in this ideal model gas in which no inter-particle forces exist is the sole store of internal energy. It is $\frac{1}{2}m\overline{u^2}$ per particle at temperature T. So the total internal energy for a mole of these particles is

$$\begin{aligned} &\tfrac{1}{2}N_A m\overline{u^2}, \\ &= \tfrac{3}{2}N_A kT, \\ &\text{from Eq. (13.12)} \\ &= \tfrac{3}{2}RT. \end{aligned} \qquad (13.15)$$

This relationship applies only to the model and to gases which approximate to the model. It tells us little about the internal energies in real cases: in particular, we must not apply it to the case of a gas approaching a temperature of 0 K, or to gases at high pressure.

In a real gas the total internal energy will include this translational kinetic energy and other forms as well. Since inter-molecular forces exist there will be a potential energy contribution and, in addition for gases other than monatomic ones, such other kinetic energies as rotational and vibrational.

Problem 3 *Diffusion.* Show that the ratio of the r.m.s. velocities of the molecules of two gases is given by the equation

$$\frac{u_1}{u_2} = \sqrt{\frac{m_2}{m_1}}$$

and consider the implication for the diffusion of gases through a porous wall.

From Eq. (13.12) for the first gas

$$\tfrac{1}{2}m_1\overline{u_1^2} = \tfrac{3}{2}kT,$$

and for the second

$$\tfrac{1}{2}m_2\overline{u_2^2} = \tfrac{3}{2}kT.$$

Since the temperature is the same for each gas we have

$$\tfrac{1}{2}m\overline{u_1^2} = \tfrac{1}{2}m_2\overline{u_2^2}$$

whence

$$\frac{\overline{u_1}}{\overline{u_2}} = \sqrt{\frac{m_2}{m_1}}.$$

It follows that the lighter gas molecules will escape through a porous wall more quickly than the heavier molecules. The gas which has passed through the barrier will therefore be richer in the lighter constituent than the original mixture. This process is the basis of the diffusion process for the separation of the two isotopes of uranium ^{235}U and ^{238}U. The gaseous compound uranium hexafluoride is used.

The molecular masses of the compound are 349 and 352 so that the "enrichment factor" $\sqrt{m_2/m_1}$ is 1.0043.

13.6 AVOGADRO'S NUMBER N_A

We have referred frequently to Avogadro's number which, when discussing a gas, is the number of particles per mole. The mole itself is the amount of substance which contains as many entities as there are atoms in 12 g of carbon 12. The entities may be atoms, or molecules, or ions, or electrons, etc., and the number per mole is, as already quoted, nearly 6×10^{23}.

The direct experimental determination of this number has been a very difficult undertaking and one method of interest is that of J. Perrin (1912). Perrin considered the distribution of particles in a column of fluid.

Consider the slice of thickness Δz of the gas column of area A (Fig. 13.17). The pressure p on the top surface of the slice is less than that on the lower surface because the column is in a gravitational field. So

$$A\,\Delta p = A\,\Delta z\, nmg, \tag{13.16}$$

where n = number of particles per unit volume,
m = mass of each particle,
Δp = pressure difference.

Now

$$pV = N_A kT$$

(from Eq. 13.13) whence

$$p = \frac{N_A}{V} kT$$

$$= nkT$$

and

$$\Delta p = kT\,\Delta n.$$

Substituting in Eq. (13.16)

$$AkT\,\Delta n = A\,\Delta z\, nmg$$

$$\frac{\Delta n}{n} = \frac{mg}{kT}\Delta z. \tag{13.17}$$

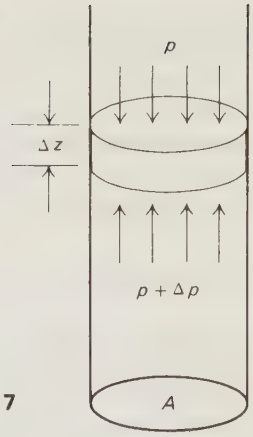


Fig. 13.17

In differential form

$$\frac{dn}{n} = \frac{mg}{kT} dz.$$

Integrating

$$\int_{n_1}^{n_2} \frac{dn}{n} = \frac{mg}{kT} \int_{z_1}^{z_2} dz,$$

where n_1 is the number of particles per unit volume at height z_1 and n_2 is the number at height z_2. Hence

$$\ln \frac{n_1}{n_2} = \frac{mg}{kT}(z_1 - z_2).$$

Replacing k by R/N_A from Eq. (13.14)

$$\ln \frac{n_1}{n_2} = \frac{mg\,N_A}{RT}(z_1 - z_2). \tag{13.18}$$

This equation applies to the particles of a gas for which the product mN_A, which is the molar mass, is known. However, Perrin showed that the equation also applies to a suspension of colloidal particles in a fluid. He was able to count the relative numbers of these particles at different heights and so to measure the ratio n_1/n_2 for particles whose mass m could also be determined. Substitution in Eq. (13.18) then gave a value for Avogadro's number N_A.

It will be appreciated that this step can be taken because N_A enters the equation as a scaling factor,

scaling the microscopic kinetic energies of individual particles to macroscopically measurable pressures.

Perrin obtained a value for N_A of 7.05×10^{23} particles per mole. The experiment, which is a difficult one, was repeated in 1915 by Westgren who obtained a value of 6.04×10^{23} which is much closer to the currently accepted value of 6.022×10^{23}.

13.7 THE ATOMIC MODEL

There is a good deal of evidence to suggest that the basic units of material structure (atoms, ions, or molecules) are only 10^{-10} m in diameter. We have noted already that the laws of chemical combination led Dalton to a particulate model for matter. In order to account also for the simple ratios in which the volumes of gases combined, it was necessary to adopt a further hypothesis, due to Avogadro, that equal volumes of gases at the same pressure and temperature contain the same number of particles. This hypothesis is not susceptible to direct proof. Its justification lies in the coherent pictures of atoms, singly and in combination, which developed from it. The kinetic theory of gases proposed a model which is also consistent with this hypothesis.

Suppose two different gases exert pressures p_1 and p_2. From Eq. (13.5)

$$p_1 V_1 = \tfrac{1}{3} N_1 m_1 \overline{u_1^2}$$

for one gas, or

$$p_1 = \tfrac{1}{3} \frac{N_1}{V_1} m_1 \overline{u_1^2}.$$

Writing n_1 in place of N_1/V_1, the number of particles per unit volume

$$p_1 = \tfrac{1}{3} n_1 m_1 \overline{u_1^2}.$$

Similarly

$$p_2 = \tfrac{1}{3} n_2 m_2 \overline{u_2^2}.$$

If $p_1 = p_2$,

$$n_1 m_1 \overline{u_1^2} = n_2 m_2 \overline{u_2^2}.$$

If $T_1 = T_2$,

$$m_1 \overline{u_1^2} = m_2 \overline{u_2^2}$$

from Eq. (13.12) whence

$$n_1 = n_2.$$

Since n_1 and n_2 are respectively the number of particles per unit volume, Avogadro's hypothesis follows.

On the basis of this hypothesis, and the laws of chemical combination, a detailed model of the particulate state emerges. The ultimate particle (as far as chemical interaction is concerned) is the *atom*. There is a different atom for each of the hundred-odd elements and their masses can be arranged in a table of *atomic masses* (referred originally to the mass of a hydrogen atom as 1, but today to the carbon atom taken as 12 atomic mass units). In the free, elemental state, these atoms may be grouped into particles consisting of two or more identical atoms. These particles are called *molecules*. In chemical combinations with other atoms, *compound molecules* are formed, consisting of two or more different atoms.

The relative masses combining in a chemical reaction are not the same as either the relative molecular or atomic masses. Referred to carbon = 12, the atomic mass of hydrogen is 1.008, while that of oxygen is 16.000. Their respective molecular masses are 2.016 and 32.000. However, in a chemical reaction to form water, each 1.008 g of hydrogen combines with only 8.000 g of oxygen. This leads to the idea that different atoms have different combining powers. On comparing, either directly or indirectly, the combining powers of each element with hydrogen, we find that each atom of oxygen will combine with two atoms of hydrogen. On the other hand, two atoms of aluminium will combine with three of oxygen, and consequently, aluminium has a combining power of three. This combining power is called the atom's *valency*.

The phenomenon of electrolysis shows that many atoms are charged in solution. For the passage of a given quantity of electricity, elements are deposited at anode or cathode in proportion to the ratio: atomic mass/valency. This is Faraday's second law of electrolysis.

This strongly suggests that the valency of an atom is related to the degree to which it becomes charged in solution. On this basis, hydrogen atoms

carry only one unit of charge, while oxygen carries two, and aluminium, three. No atom has a smaller valency than hydrogen.

Millikan's experiments on the charge carried by oil drops show that the smallest charge which can occur is 1.602×10^{-19} coulomb. This is the charge we eventually associate with an electron. It is tempting to assume that the charge associated with a valency of 1 is also 1.602×10^{-19} coulomb. Let us make this assumption.

Careful electrolysis experiments show that it takes 96 500 coulombs to release 1.008 g of hydrogen. If each hydrogen atom had carried 1.602×10^{-19} coulomb, there must be

$$\frac{96\,500}{1.602 \times 10^{-19}} \text{ atoms in 1.008 g of hydrogen}$$

$$= 6.02 \times 10^{23} \text{ atoms in 1.008 g of hydrogen.}$$

Similar calculations for other elements lead to the conclusion that there are 6.02×10^{23} atoms in every gramme-atom (or mole) of an element. This is the number to which we gave the name of Avogadro's number (N_A). This method of arriving at it depends upon two unproved assumptions:

a) Avogadro's hypothesis,
b) a valency of 1 is related to a charge in solution of 1.602×10^{-19} coulomb.

Perrin's method relied on a set of entirely different assumptions based on the kinetic model of a gas and its extension to a suspension. The convergence of the experimentally determined values is part of the consistent picture which has been developed of the atomic state and lends powerful support to our acceptance of the whole theoretical structure.

Problem 4 *Some particle arithmetic.*

a) Avogadro's number is 6.022×10^{23} particles per mole. Calculate the mass of an atom of a substance one mole of which has a mass of 1.000×10^{-3} kg.
The mass of one atom of this substance is

$$\frac{1.000 \times 10^{-3}}{6.022 \times 10^{23}} \text{ kg} = 1.66 \times 10^{-27} \text{ kg.}$$

b) The mass of one mole of atomic hydrogen is 1.008×10^{-3} kg. What is the mass of a single atom of hydrogen?
The mass of one atom of hydrogen m_H is

$$\frac{1.008 \times 10^{-3}}{6.022 \times 10^{23}} \text{ kg} = 1.67 \times 10^{-27} \text{ kg.}$$

c) The mass of one mole of molecular nitrogen is 28.0×10^{-3} kg. What is the mass of a single molecule of nitrogen?
The mass of a single molecule of nitrogen is

$$\frac{28.0 \times 10^{-3}}{6.022 \times 10^{23}} \text{ kg} = 46.5 \times 10^{-27} \text{ kg.}$$

d) What is the average kinetic energy of a molecule of an ideal gas at a temperature of 300 K?
From Eq. (13.12) the average kinetic energy is

$$\tfrac{3}{2}kT = \tfrac{3}{2} \times 1.38 \times 10^{-23} \times 300 \text{ J},$$
$$= 6.20 \times 10^{-21} \text{ J.}$$

e) What is the total random kinetic energy of the particles of 1 mole of such a gas at this temperature?
From Eq. (13.15) the total internal energy is

$$\tfrac{3}{2}RT = \tfrac{3}{2} \times 8.3 \times 300 \text{ J},$$
$$= 3735 \text{ J.}$$

f) What is the r.m.s. velocity of a molecule of bromine at 300 K? The mass of the bromine molecule is $160 \times 1.66 \times 10^{-27}$ kg. From Eq. (13.12)

$$\sqrt{\overline{u^2}} = \sqrt{\frac{3kT}{m}}$$
$$= \sqrt{\frac{3 \times 1.38 \times 10^{-23} \times 300}{160 \times 1.66 \times 10^{-27}}} \text{ m s}^{-1}$$
$$= 210 \text{ m s}^{-1}.$$

Chapter 14
A MODEL FOR A SOLID

14.1 A possible relationship between a force and the separation of the particles

14.2 The physical properties of solids

14.3 Further investigation of a model for the solid state

The success of this kinetic model of a gas in which an assemblage of particles has to be endowed with very few properties in order to produce something which behaves much as a real gas behaves is so encouraging that the extension to the liquid and the solid phases is an obvious step to take.

Now we know that all matter can exist in three phases—solid, liquid, and gas. Which phase or state a particular material happens to be in depends on its temperature and also on the external pressure. While gases take up the whole volume of their container and are highly compressible, liquids and solids have a definite volume of their own and, compared with gases, are virtually incompressible. If we are to extend our present model to incorporate liquids and solids, it would appear reasonable to assume that in these two phases, the particles are packed together as closely as possible. Certainly the change in volume of a material, in the liquid phase, is very small as it is cooled down further, compared with the great change in volume which occurs when it is condensed to a liquid from a gas.

Such a condition will be associated with very low values of kinetic energy of translation within the model, with no collisions between the highly ordered particles. Very low values of internal kinetic energy are in turn associated with very low temperatures. And yet we can see that the temperature of a block of copper resting on the bench is unlikely to be any different from the temperature of the gas (air) in the room. If heat is not being transferred from block to air or air to block, the internal energies of the two must be the same.

Moreover, the copper possesses one other important property not shown by air: the property we commonly call strength—the ability to resist not only compressive and extensive forces, but also shearing and twisting forces. It is this strength and rigidity which gives solids their important constructional properties.

None of these features can be explained by the simple model which matched the properties of a gas unless it is assumed that some force acts between the particles, holding them together, whilst some other force acts to keep them apart and prevents collapse.

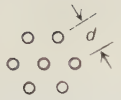


Fig. 14.1



Fig. 14.2

14.1 A POSSIBLE RELATIONSHIP BETWEEN A FORCE AND THE SEPARATION OF THE PARTICLES

This force must be of very short range. Two pieces of metal show no sign of attracting each other when brought close together other than gravitationally—and that force is almost negligible. Other evidence suggests that the force probably extends over a range of no more than a few particle diameters. (We assume that the particles are spherical in shape for this is the simplest assumption to make.)

Suppose then we start with a collection of particles, quite close to each other as shown in Fig. 14.1. (d = separation between the particles.) Now these particles are pulled apart (Fig. 14.2). The force on each particle might change as in Fig. 14.3. In order to separate them, energy must be expended from some outside source. When the particles are many diameters apart, their mutual potential energy will be zero. So a graph of potential energy against separation, in which potential energy is shown as negative to indicate that work has to be done on the system to secure the separation, would look something like that shown in Fig. 14.4. Can we extrapolate these graphs towards zero separation of the centres of the particles? By no means: this is another case where extrapolation would be very dangerous. Energy must be expended not only in separating the particles but also in compressing them. Once the particles are very close, a force of repulsion must be dominant (Fig. 14.5). Adding the graphs in Figs. 14.3 and 14.5 we obtain that shown in Fig. 14.6. The separation of the

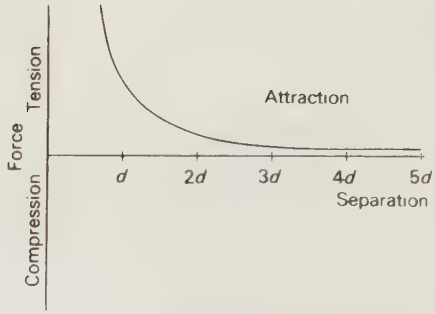


Fig. 14.3

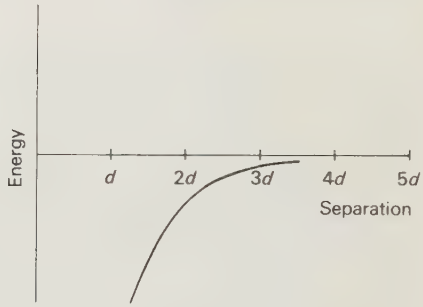


Fig. 14.4

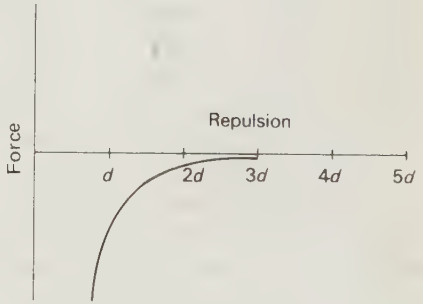


Fig. 14.5

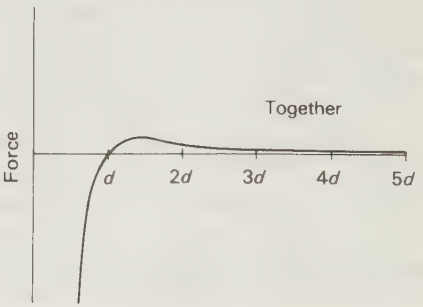


Fig. 14.6

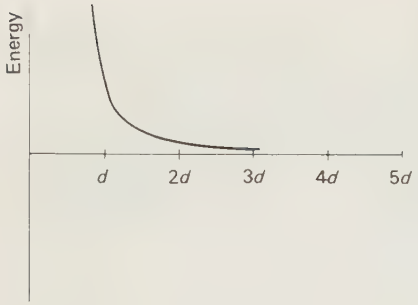


Fig. 14.7

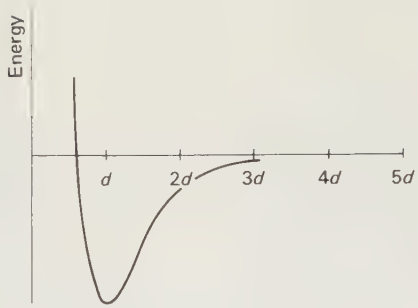


Fig. 14.8

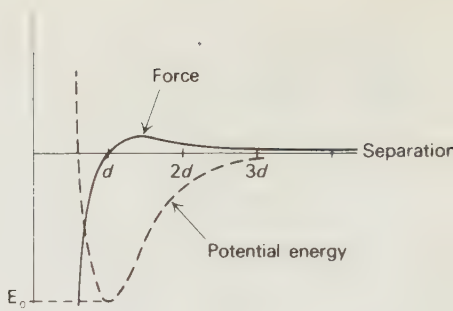


Fig. 14.9

centres corresponding to zero resultant force between the particles will be the *equilibrium separation* of the particles. We will return to this graph in a moment. Now let us look at the corresponding energy graph.

Energy must be transformed to push the particles closer together than their equilibrium separation (Fig. 14.7). Adding the two energy graphs (Figs. 14.4 and 14.7) we have Fig. 14.8. Equilibrium separation corresponds to the point where the potential energy of the assemblage of particles is a minimum. Figure 14.9 shows the graph obtained by superimposing the two graphs (Figs. 14.6 and 14.8). (It is worth noting in passing that a position of zero net force does not necessarily imply a position of zero potential energy; it will however imply a maximum or a minimum: why is this?)

Now let us reintroduce the internal energy. Suppose the minimum potential energy is $-E_0$ and that the internal energy is U_T (written with this suffix to show that it is a function of temperature). If we also assume that E_0 is a function simply of the material and its particles (atoms or molecules) then if $U_T \gg E_0$, the internal energy of the particles will always be much greater than the potential energy tending to bind the particles together in the collection. So the material will at all times be a gas.

If $E_0 \gg U_T$ the particles will always be bound together. Extra energy must be put into the material to pull the particles apart. The material will be a solid or a liquid. And in between? Well, if E_0 is only slightly greater than U_T, the particles will be loosely bound together. It is tempting to think of the material as a liquid, but even today, a good model of a liquid has not been obtained. Suffice it to say that the liquid state is something between the two extremes of solid and gas.

If $U_T \gg E_0$ we may ignore the inter-particle force, and our old model will do (although some of the properties of a gas associated with its departure from the ideal behaviour do depend on E).

Let us now see if this addition of an inter-particle force can help explain the properties of a solid, when $E_0 \gg U_T$.

14.2 THE PHYSICAL PROPERTIES OF SOLIDS

The inter-particle force and energy curves have been constructed on the basis of the observed existence of solids and their known ability to resist deformation. It would be unwise now to examine this force curve and show that it apparently predicted the existence of solids and their properties of resisting deformation.

This may seem a rather obvious statement. Nevertheless, you may read accounts of the properties of materials which follow that pattern. Such accounts state as a first principle that the forces between the particles (atoms, molecules, or ions) assumed to make up the material do vary according to Fig. 14.6. This graph is then used to explain the existence of the solid state with its properties of strength and elasticity. It would be possible to justify such an approach to the model, but one would then have to derive the inter-particle force curve from other considerations: say from some model of the structure of the particles. If

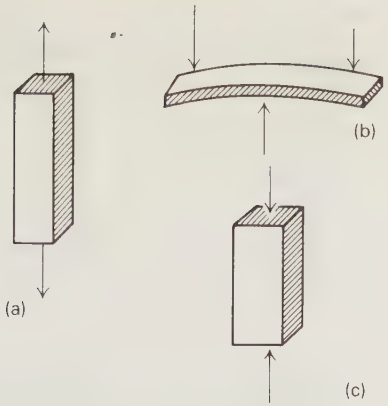


Fig. 14.10

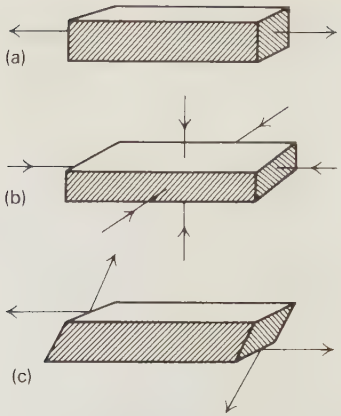


Fig. 14.11

we digress for a moment to consider such sophisticated models, it would appear that the force tending to bind solids together is not the same in all cases. Consider the three solids: sodium chloride, copper, and diamond. Evidence from the chemical behaviour of these materials and models of their atomic structure suggest that the interparticle forces arise in

a) sodium chloride, from the electrostatic attraction of the oppositely charged ions of sodium (positive) and chlorine (negative),
b) copper, from the interaction of positively charged ions (copper) with a uniformly distributed cloud of electrons,
c) diamond, from the sharing of electrons with neighbouring atoms (covalent bonds).

That the interparticle force curves for all these different bonding forces should have similar features is remarkable. Perhaps it is a recognition that these three substances share the common property of very strong resistance to deformation from outside forces.

We raise this point now just to point out how careful one must be in seeking to justify a model in terms of its relationships with real properties. One must be clear about which features have been used to set up the model and which features may be predicted *by* the model.

However, it would be naive to assume that all the properties of solids were summed up in their ability to resist deformations. Whereas gases are remarkable for the independence of their characteristic behaviour from their chemical nature, the characteristic behaviours of solids are remarkable for their wide disparity—as the well-known saying "As different as chalk from cheese" implies. Let us see under what particular headings we normally characterize the behaviour of solids.

a) Strength

By the strength of a material we shall mean explicitly the ability of the material to withstand a force without breaking. The converse of a strong material is a weak material. But there is more to strength than the breaking of a material. We need also to say how the material is going to be broken (see Figs. 14.10a, b, c). The ability of a material to resist breakage in (a) is called its *tensile* strength. Breakage in (b) is closely related to this. Breakage in (c) depends on its *compressive* strength, but if it is thin, it may deform as in (b). Tensile strength is the easiest to measure experimentally. To say that a material is either weak or strong is not enough. Some materials, like brick, are weak in tension but strong under compression.

b) Stiffness

Solids will deform under the application of any force no matter how small. Figure 14.6 shows that as soon

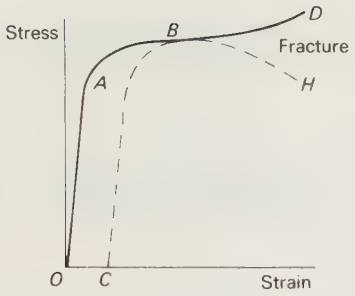


Fig. 14.12

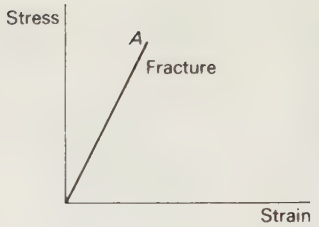


Fig. 14.13

as the particles of the solid are either pulled further apart or pushed together, a restoring force appears. Solids must deform in order to produce a reaction to the applied force.

There are three basic ways in which a material can be deformed (see Fig. 14.11a, b, c). It can be stretched or compressed in one dimension (14.11a). It can be uniformly compressed (14.11b). It can be sheared (14.11c). Of these, perhaps the most important deformation is that in Fig. 14.11a as it is closely related to the *bending* of materials under applied forces (see Fig. 14.10b). The tendency of a body to resist the deformation characterized by Fig. 14.11a is a measure of what is called its *stiffness*.

The extension produced in most solids by a tensile force is proportional to that force, at least for small extensions. This is the most general form of a law attributed to Robert Hooke, who first stated it for springs.

The extension of any piece of material depends not only on the applied force but also on its length and its cross-sectional area. It is therefore usual to talk of the applied *stress* (equal to the force/unit cross sectional area) and the consequent *strain* (equal to the extension per unit length). The ratio

$$\textit{stress}/\textit{strain}$$

is a quantity characteristic of the stiffness of the material. It is called Young's modulus (E).

This ratio, which is measured in newtons per square metre, tells us how stiff or how flexible a material is. This is quite a different property from the *strength* of the material. Compare, for example, stiff and strong steel with stiff and weak chalk or flexible and strong nylon with flexible and weak jelly. Table 14.1 will enable you to make other comparisons.

Table 14.1

Material	Young's modulus (E) N m^{-2}	Tensile strength N m^{-2}
Aluminium	70×10^9	70×10^6
Mild steel	200×10^9	400×10^6
Cast iron	20×10^9	$30 - 140 \times 10^6$
Nylon	$0.5 - 3 \times 10^9$	70×10^6
Wood	13×10^9	100×10^6 along the grain 4×10^6 across the grain
Glass	70×10^9	$3.5 - 150 \times 10^6$
Carbon fibres	900×10^9	1000×10^6
Concrete	—	4×10^6

c) Stress-strain behaviour

The behaviour of materials under increasing stress up to breaking point is usually characterized by stress-strain curves. Many metals behave typically as in Fig. 14.12. The stress-strain curve is linear up to point A (called the *elastic limit*). Thereafter, the material yields *plastically*, and then on removal of the force, the material is found to have a permanent deformation (OC).

Ultimately the material snaps after a plastic deformation many times greater than its initial elastic deformation, which is typically about 1% for engineering materials. Whether the curve follows BD or BH depends whether the average stress along the material

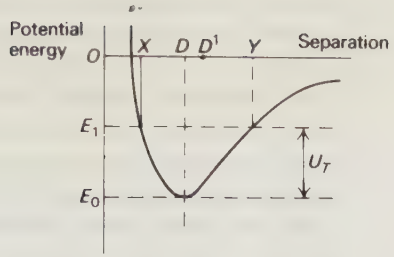


Fig. 14.14

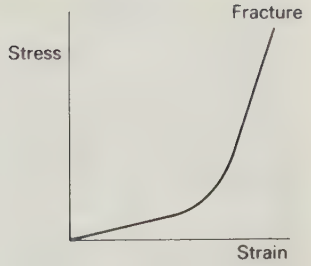


Fig. 14.15

is plotted (curve BH), or the real stress at the point of snapping (curve BD).

Solids which yield plastically in this way are referred to as *ductile*.

Some other materials such as glass behave as in Fig. 14.13. Fracture occurs at A and there is no plastic yielding. Typically the breaking stress is low compared with that for ductile materials. Materials which behave in this way are referred to as *brittle*.

d) The effect of temperature

Solids expand on heating. We can interpret this expansion in our model using the graph in Fig. 14.8. The equilibrium separation of the particles, centre to centre, is OD.

At temperature T, the particles will have additional energy U_T. Since the particles still have insufficient energy to escape from each other, the form of this energy will be vibrational. As a body vibrates, it constantly re-distributes its total energy between kinetic and potential forms. At X and Y, in Fig. 14.14, when it is momentarily at rest, its energy is entirely potential. In this case, that potential energy $(-E_1)$ must be $-(E - U_T)$. (Remember the total potential energy must remain negative.) The shape of the potential energy curve will control the amplitude of the vibration. The average position of the vibrating particle will be half-way between these two extremes—at D'. OD' is greater than OD because the potential energy curve is not symmetrical about D.

So the material expands on heating, not solely because the "thermal wobble" increases but because of the asymmetrical nature of the energy curve. We might expect the curve to be asymmetrical in this way because it is clearly impossible to squash the particles so close that they overlap completely (separation zero), but there is no limit to how far apart they may be pulled. Expansion with rise in temperature is confirmation of this guess.

Apart from expansion effects, temperature also has considerable effects on strength and stiffness. In general, a solid's strength and stiffness decreases with rising temperature.

The brittle characteristics of a solid can be affected by temperature changes. By suitable heating and rapid cooling, a ductile piece of steel can be changed into a brittle piece. Re-heating and slow cooling can reverse the effect.

This is also an appropriate point to note the behaviour of polymeric solids which chemical evidence has suggested are built up atoms arranged in long chains. At room temperature, rubber shows the typical stress-strain relationship given in Fig. 14.15. However, if the rubber is cooled down a sufficient amount, the stress-strain relationship radically alters to that exhibited in Fig. 14.13, which was typical of the behaviour of glass at room temperature. This change in behaviour and properties takes place at a characteristic temperature called the *glass-transition temperature*.

e) The effect of time

Time occurs as an effective variable in two important behaviours of materials.

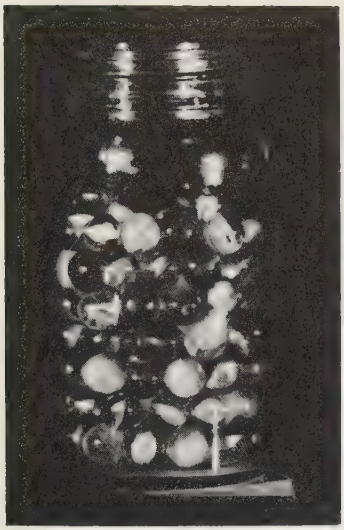


Fig. 14.16 Marbles in a glass jar.

1. Creep. Under the continuing application of a stress, some materials (e.g., timber and many plastics) progressively deform.

In considering earlier stress-strain relationships the observed strain was assumed to take place instantaneously in response to the applied stress. When creep takes place, a progressively increasing extension or other deformation takes place as long as the load is applied. This extension cannot be accounted for in terms of our present model.

2. Fatigue. The repeated application and removal of a stress well below the stress required for fracture can eventually fracture the material. This is a particular feature of metallic behaviour and is the mechanism by which a wire can be broken by bending it backwards and forwards in one's hands.

f) Metals and non-metals

So far we have only considered the behaviour of solids under stress. Further characteristics are to be found in the common classification of solids into metals and non-metals. These two classes of material are partially differentiated by their behaviour under stress. Metals are usually (but not always) distorted by a stress, i.e., are ductile, while non-metals are almost always brittle (although plastics, which are non-metals, may well not be). A more certain distinction between metals and non-metals is to be found in their electrical and thermal properties under normal conditions. All metals are good conductors of electricity and heat. Non-metals are poor conductors of both electricity and heat. Another distinguishing feature lies in the ability of metals to take a high polish.

The distinction between the two classes is never complete, and there have always been materials, such as carbon in its graphite form, which are difficult to classify. In this case, it appears electrically to behave like a metal, but its stress-strain properties are more to be associated with non-metals. In recent years, a class of material called semi-conductors has gained importance. Germanium, which is difficult to classify either as metal or non-metal, is a typical representative.

In the last six sections we have summarized the major physical properties of solids. These are not so simple as the properties of gases which we have also sought to interpret. All the properties described can, with simple apparatus, be observed in the laboratory.

14.3 FURTHER INVESTIGATION OF A MODEL FOR THE SOLID STATE

While the interparticle force and energy curves can be used to interpret the inherent strength and elasticity of solids, they assist but little in the interpretation of the wide qualitative differences in behaviour.

When we considered the gas model, the particles were assumed to be distributed with a random variation in velocity. This provides a reason why all gases behave similarly. There is only one sort of randomness for the particles.

What about the distribution of the particles as the gas is condensed to the liquid and then to the solid phase? We have assumed that they come as close to one another as they can.

If marbles are shaken down into a jar so that they make the most compact arrangement possible, their

distribution has an ordered look about it (see Fig. 14.16). A completely disordered arrangement takes up more space.

Similarly, we can blow bubbles of a fixed diameter in a soap solution, allowing them to pack close together over the surface. This arrangement is clearly an ordered one. Perhaps the particles in the liquid and solid phases pack themselves into ordered arrangements. And since these particles may not be simple spheres, there may be different arrangements for different materials.

Our model is now becoming more complex. We are having to assume that the constituent particles of matter have certain properties (unnamed) which distinguish them from each other. It is no longer good enough to ignore this fact.

Chapter 15
CRYSTALLINE STRUCTURE

15.1 Three-dimensional close packing
15.2 Orderly arrangements
15.3 Crystal structure and the properties of solids
15.4 Elastic deformation of an ordered structure
15.5 Solids as ordered structures
15.6 Interference patterns and crystals
15.7 Investigating the structure of real solids

A solid is referred to as crystalline when it exhibits a regularity of form recognized by plane surfaces meeting each other at characteristic angles. By building up ordered structures using large spheres (Fig. 15.1) it is possible to reproduce the macroscopic regularity of form which crystals exhibit. Maybe crystals too have such an underlying regularity of arrangement of the basic particles (atoms, ions, or molecules). An elementary account of this can be found in *Crystals*, by M. M. Hurst, 1969.

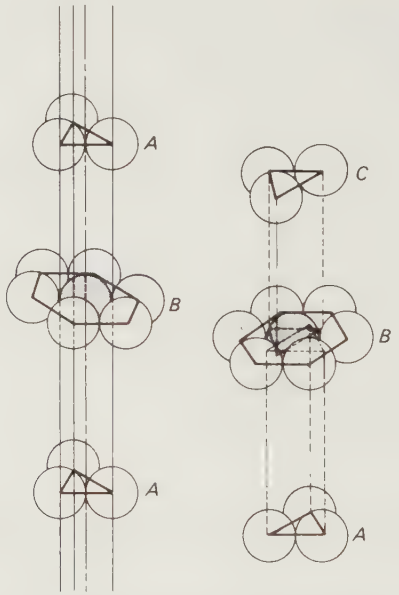


Fig. 15.1 Stacking sequences. (a) ABA sequence. (b) ABC sequence. (From L. H. Van Vlack, *Elements of Materials Science*, Addison-Wesley, 1964.)

The fact that many such materials are also possessed of a certain beauty of appearance is not to say that many less notable solids are not also crystalline in their *micro*-structure. Careful examination with a microscope reveals that any metallic specimen is made up of a conglomerate of minute crystals (Fig. 15.2). The outer form of a material may not necessarily reveal an inner regularity of structure.

Before pursuing the question of possible outcomes from a regular or a random arrangement of the



Fig. 15.2 Photomicrograph of hot-worked deoxidized copper. (Reproduced by permission of the Copper Development Association.)

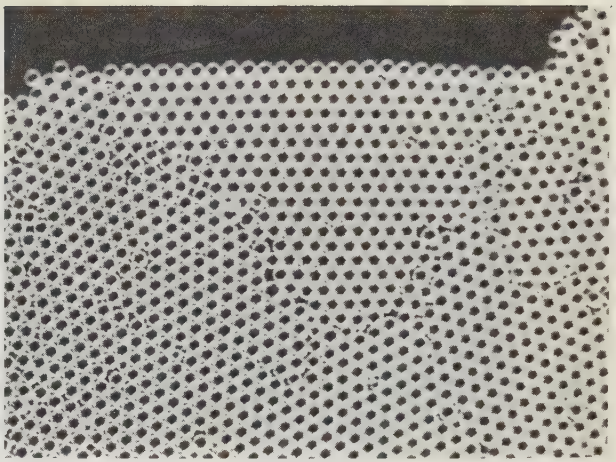


Fig. 15.3 A raft of bubbles showing close packing. (Photograph by E.J.W.)

particles of a solid, we ought first to see what sort of ordered arrangements we may expect to find. Again, we shall keep our model as simple as possible consistent with the new circumstances. We will assume the simplest shape—a sphere—for the particles and deal only with solids in which a single type of particle is known to be present—namely, elements such as copper and carbon.

In order to help the discussion, we may wish to build large-scale regular structures with large scale spheres (5 cm polystyrene spheres were used in the photographs, Fig. 5.14). These structures are often called "models", but this is a different use of the term "model" from that intended when we talk of, say, the "kinetic model" of a gas. It is much closer to the scale-model builders' use of the word. This double meaning which can be attributed to the word "model" when used by physicists can be confusing, but the confusion is often cleared up once one realizes the possibility of confusion. Remember, the term *model* can mean:

a) theory—a very important use of the word, unique to scientists;
b) piece of hardware to aid thinking—more nearly equivalent to every-day use of the word.

15.1 THREE-DIMENSIONAL CLOSE PACKING

A two-dimensional arrangement of soap bubbles, all the same size and packed together as closely as possible, would look like Fig. 15.3. This is the only close packed arrangement possible. No other arrangement, whether orderly or disorderly, will pack the bubbles so closely together.

A solid is three-dimensional. Close packing can be achieved by stacking two-dimensional rafts, one on top of the other. To see the sort of thing that can be achieved, we may make structure models from spheres.

Figure 15.1 shows a lower raft, or sheet (*A*) of close packed spheres with another one (*B*) placed above it. Each sphere in the sheet *B* fits into the spaces between every three spheres in the bottom sheet. If a third sheet, *C*, is placed upon the sheet *B*, we find that there are two choices for its position. This will give two different spatial arrangements of the spheres —both close-packed.

If the relative position of the spheres in the first sheet is identified by the letter *A*, and those in the second sheet by the letter *B*, then a three-dimensional structure which continuously repeats this form of

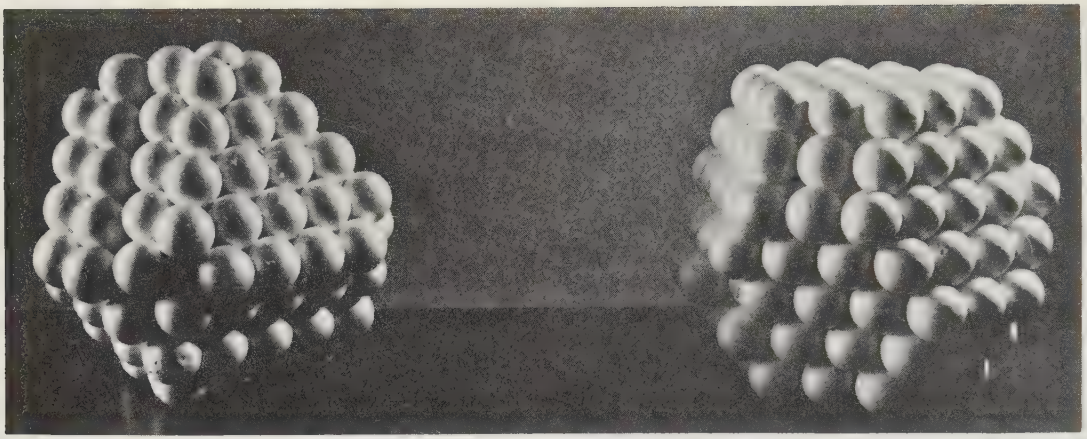


Fig. 15.4 Close-packed structures made up from polystyrene spheres. (a) Hexagonal close packed. (b) Face centred cubic.

close packing can be denoted by the letter sequence:

$$ABABABA \ldots \text{ etc.}$$

Alternatively, each sphere in the third sheet can be placed so that it is directly above neither the spheres in *A* nor in *B*. Denoting their relative positions by the letter *C*, a three-dimensional structure continuously repeating this form of close packing can be denoted by the letter sequence:

$$ABCABCABCA \ldots \text{ etc.}$$

Figures 15.4(a) and 15.4(b) show photographs of two regular structures built up from 5 cm polystyrene spheres using (in Fig. 15.4a) the sequence *ABABA* . . . and (in Fig. 15.4b) the sequence *ABCABCABCA* . . .). We shall find that these two structures turn out to be rather important. They underlie quite a few crystal structures. Theoretically it is possible to have mixtures like *ABCBCBCABCABABABC*, etc., but, in practice, they do not occur at all frequently. It will be beyond our present model to explain this. The structure *ABA* . . . is usually referred to as *hexagonal close packed* (h.c.p.) and the structure *ABCA* . . . is called *face centred cubic* (f.c.c.).

15.2 ORDERLY ARRANGEMENTS

Orderly arrangements of any sort can be described in terms of a pattern. The word pattern is commonly used when describing such things as wall paper or dress fabrics. The essence of order is the constant relationship of any basic unit to its near neighbours. Wall papers and dress fabrics can differ from each other not only in the design of the basic unit, but also in the relationship of the basic units to each other. The smallest piece of the whole design which displays both the basic unit and its inter-relationship with other units is called the pattern.

This idea of a pattern can be carried over to any orderly arrangement. Look at the two-dimensional arrangements of discs in Fig. 15.5(a). In each case, any one shaded disc bears the same relationship to its neighbours as every other disc within its own arrangement. But the patterns in the two cases are different. Figure 15.5(a) illustrates the basic pattern in the first case and Fig. 15.5(b) in the second case. In both cases some discs are shared with neighbouring repetitions of the pattern, so the basic pattern may be drawn as in Fig. 15.5(c).

In three-dimensional structures, the pattern of the arrangement is the smallest three-dimensional structure which shows the relationship between each atom and its nearest neighbours. The pattern is called the *unit cell* and in the case of the *ABA* . . . structure, is (using dots for the centre of each sphere) shown in

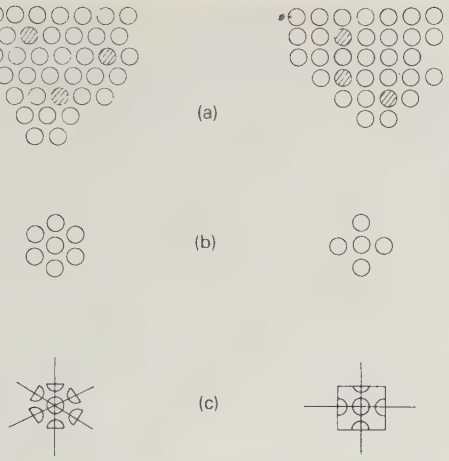


Fig. 15.5

Fig. 15.6. Hence the term hexagonal close packed; "hexagonal" describes the shape of the unit cell.

Figure 15.7 shows the case of the *ABCA* . . . structures. Fig. 15.8 shows it as made up solid spheres. (The letters in each diagram refer to the planes identified by these letters in Fig. 15.1.) In this last case, the unit cell is a cube with spheres at each corner and one in each face—hence the term *face centred cubic*.

The shape of the unit cell provides an effective description of the many possible arrangements of the spheres.

15.3 CRYSTAL STRUCTURE AND THE PROPERTIES OF SOLIDS

We have seen that many of the properties of solids, and, in particular, the wide variations in their properties, cannot be explained in terms of the simple inter-particle force curve. Will the possibility of an ordered arrangement of the particles help?

15.4 ELASTIC DEFORMATION OF AN ORDERED STRUCTURE

In Fig. 15.9, *A* and *B* represent sections through two imaginary sheets of atoms in a metal under stress. For

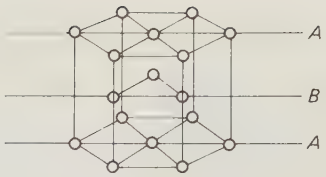


Fig. 15.6 Schematic view of a hexagonal close packed structure showing the location of atom centres. (From L. H. Van Vlack, *Elements of Materials Science*, Addison-Wesley, 1964.)

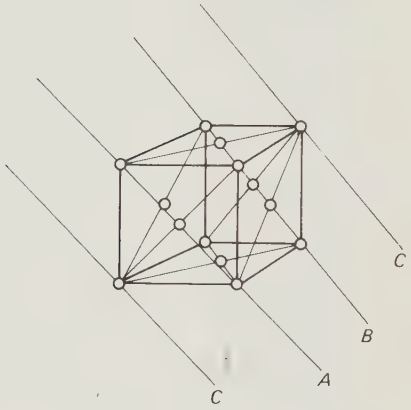


Fig. 15.7

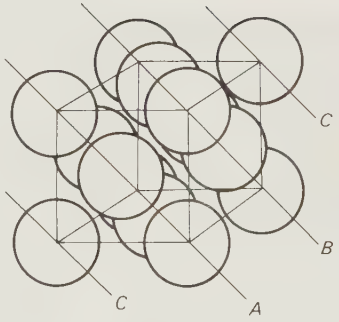


Fig. 15.8

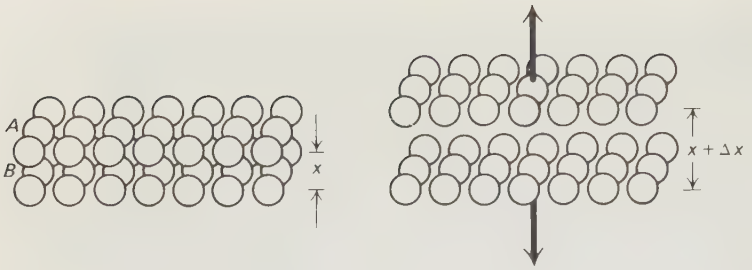


Fig. 15.9

simplicity, the stress is shown at right angles to the sheets of the atoms which are arranged in a simple cubic array. The distance between the atoms' centres is x. Under the action of stress the atoms will move apart, but as little as possible, the distance between the atom centres in the direction of stress increasing to $x + \Delta x$.

If the stress is removed, the atoms will return to their original equilibrium position and the deformation will vanish. This represents the process of elastic deformation which we have already considered.

In such a case the elastic strain is $\Delta x/x$. Since the deformation is elastic, pairs of atoms in adjacent planes will be pulled together by a force $nk\Delta x$ where n is the number of atoms in such a layer and k is the force constant. The force per unit area between the two layers will be $(n/A)k\ \Delta x$.

Now the array was a cubical one so that the number of atoms per unit area is $(1/x^2)$. Therefore,

$$\text{force per unit area or stress} = \frac{1}{x^2}\, k\ \Delta x.$$

$$\text{Young's modulus } E = \frac{\text{stress}}{\text{strain}}$$

$$= \frac{\frac{1}{x^2}\, k\ \Delta x}{\frac{\Delta x}{x}}$$

$$= \frac{k}{x}.$$

In the case of steel, E is about 2×10^{11} N m^{-2} and x is about 3×10^{-10} m. So the force constant k is about $2 \times 10^{11} \times 3 \times 10^{-10}$ N m^{-2} m or 60 N m^{-1}.

The reality of this physical separation of the atoms in a metal under tension has been confirmed by measurements of the inter-atomic spacing in stretched and unstretched specimens, using X-ray diffraction techniques. However, the values obtained are not usually those predicted from the inter-atomic force-separation curve.

15.5 SOLIDS AS ORDERED STRUCTURES

Evidently the presence of order in solid structures could help to explain some of their important differences. What evidence is there that such order exists? If it does exist, what are its basic patterns?

First, let us attempt to estimate the size of the particles involved in these basic patterns (see also Unit 4). We will accept the basic assumptions of the atomic model and apply them to a metal, for example, copper. The density of copper is 9000 kg m^{-3} and its atomic mass 63.6.

Calculate the volume occupied by 63.6 g of copper. Divide this volume by the number of atoms in 63.6 g. This will give you the volume, V, occupied by each atom of copper. If all this space were occupied by the copper, the atoms would be small cubes of side 2.3×10^{-10}m. If, on the other hand, we assume some close packed structure of copper spheres (it turns out to be face-centred cubic) and so allow for the space not occupied by the copper, the diameter of each atom comes out nearer to 1.3×10^{-10}m.

Either way, this result supports our assertion as to the small size of these ultimate particles. It is

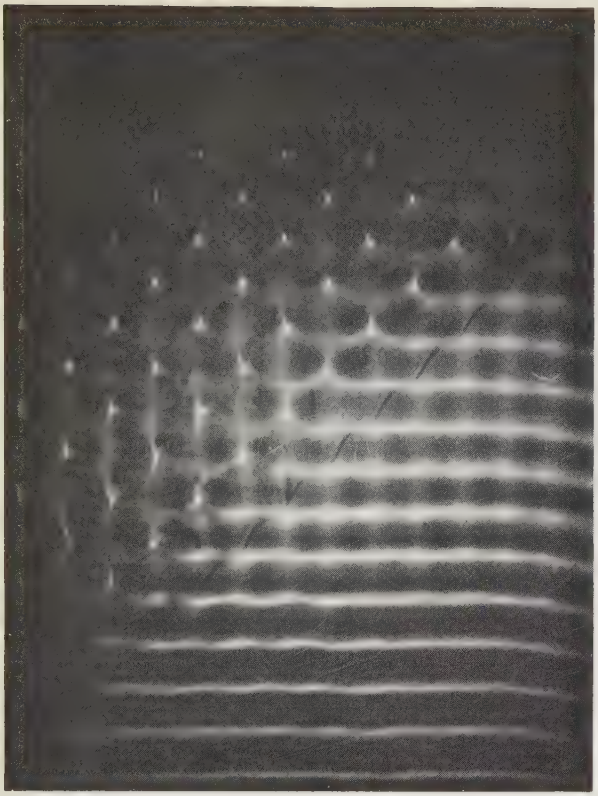


Fig. 15.10 The reflection of plane ripples from a regular array of obstacles in a ripple tank. (Photograph by E.J.W.)

beyond the range of even the most powerful microscopes. We cannot expect to see this ordered arrangement, if it exists, in any direct sense. But we may be able to infer it from some indirect observations which themselves depend on this order. Crystalline structure has already given some hint of underlying order. But the most important materials do not, in general, show any visible external regularity of form.

Instead of trying to see the particle directly, we use a different effect.

15.6 INTERFERENCE PATTERNS AND CRYSTALS

Look at Fig. 23.15. It shows a cylinder placed in the path of plane waves in a ripple tank. The plane waves are diffracted by the cylinder. The pattern it produces is best described as a plane wave with another cylindrical wave, centred on the cylinder, superimposed upon it.

This pattern is produced only if the cylinder's diameter is of the same order of magnitude as the wavelength of the incident wave. Try it for yourself. What happens if

a) the cylinder is much larger than the wavelength of the plane wave?
b) much smaller?

(See also Unit 5. To answer, imagine a sea wave passing (a) a large rock and (b) a post.)

In three dimensions, a sphere will set up a spherical secondary diffracted wave centred on the sphere.

An atom will scatter waves in this way if they are of the right order of magnitude of wavelength. We have seen already that atoms are about 10^{-10} m across. We would need radiation of about this wavelength to produce a similar effect to that observed in the ripple tank—X-rays, in fact.

A crystal, we suspect, is a regular array of atoms. Let us see how a regular array of scattering centres would affect radiation whose wavelength is about that of the diameter of a scattering centre.

First consider a simple model in a ripple tank (Fig. 15.10). The scattered radiation is concentrated into specific directions. It is no longer evenly distributed as it was from the single cylinder.

In another model (Fig. 15.11) the 3 cm radiation is scattered into sharply defined directions only when the model crystal makes particular angles with the incident beam.

This strong reinforcement of the radiation in particular directions is a characteristic behaviour of waves when several waves are superimposed on each other—a behaviour we call *interference*. Each polystyrene sphere in the model crystal scatters the electromagnetic radiation in all directions. These scattered wavelets will interfere with each other. We will now see if we can calculate the directions in which we would expect strong reinforcement. To simplify matters, consider a section through a regular cubic array of spheres which is placed in the path of an



Fig. 15.11 Experimental arrangement for the investigation of the scattering of 3 cm waves in a model crystal structure. (Reproduced by permission of the Unilab Division of Rainbow Radio, Ltd.)

on-coming plane wave, ONM (Fig. 15.12). AB is supposed parallel to the surface of the crystal and the wave normal NP makes an angle θ with AB.

Let us further suppose that strong reinforcement of the scattered radiation takes place in a direction θ' to AB. If this is so, all the scattered wavelets, from scattering centres in all the planes AB, CD, etc., must be in phase. It may seem a formidable task to write down the conditions for this, but a moment's inspection of the diagram shows that:

a) if PS is in phase with QT, then all waves from scattering centres in plane AB will be in phase with each other. So will all waves from scattering centres in plane CD be in phase with each other.
b) if QT is also in phase with RU, then all waves from scattering centres in AB will be in phase with all waves from scattering centres in CD and so on.

For PS to be in phase with QT,

$$VQ - PW = k\lambda,$$

where k is an integer. In triangle PVQ:

$$VQ = a\cos\theta,$$

In triangle PWQ:

$$PW = a\cos\theta'.$$

So the first condition to be satisfied is

$$a(\cos\theta - \cos\theta') = k\lambda.$$

For QT to be in phase with RU,

$$XR + RY = l\lambda,$$

where l is an integer. In triangle RXQ:

$$XR = b\sin\theta.$$

In triangle RYQ:

$$YR = b\sin\theta'.$$

So the second condition to be satisfied is

$$b(\sin\theta + \sin\theta') = l\lambda.$$

So for a strong reinforcement of the radiation to take place in a direction θ' to the crystal surface:

$$a(\cos\theta - \cos\theta') = k\lambda, \qquad (15.1)$$

$$b(\sin\theta + \sin\theta') = l\lambda. \qquad (15.2)$$

These equations are due to von Laue, the discoverer of X-ray diffraction patterns produced by crystals.

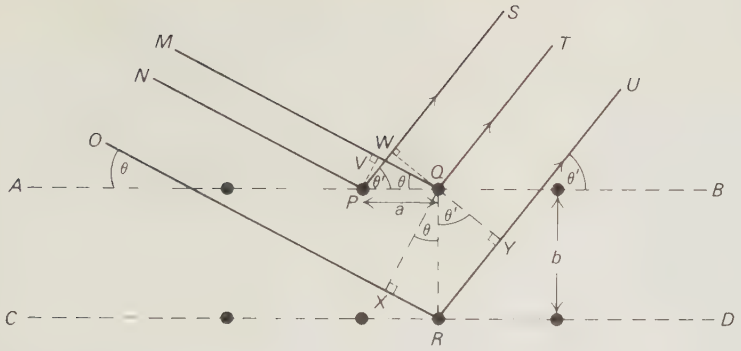


Fig. 15.12

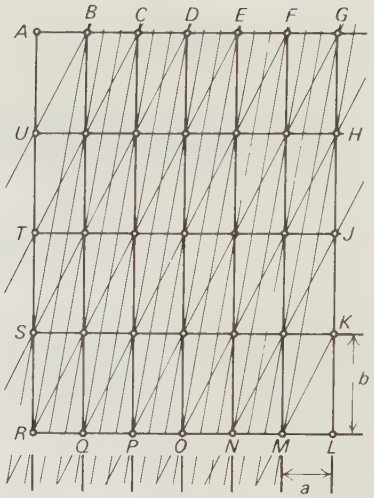


Fig. 15.13

The existence of strong reinforcement of radiation falling on a crystal would itself be sufficient to justify our belief in a regular structure. A measurement of angles θ and θ' should enable us to calculate the internal dimensions of the structure. This would be difficult, if it were not for a curious geometrical result which was first worked out in detail by Sir Lawrence Bragg.

He pointed out that there were a large number of ways of dividing up a regular crystal into a number of parallel planes. Let us look again at the simple array of points we have considered.

In Fig. 15.13, the array of points has been divided up in four different ways. You will be able to spot even more. We could describe each of these parallel lines of points in terms of the angle they make with AB.

If α is the slope of lines BU, CT, etc., $\tan \alpha = \dfrac{b}{a}$.

For rows AG, UH, etc., $\tan \alpha = 0$.

For rows SB, MH, etc., $\tan \alpha = \dfrac{3b}{a}$.

For rows RD, etc., $\tan \alpha = \dfrac{4b}{3a}$.

For the vertical rows $\tan \alpha = \infty$.

In general, each set of rows can be represented by lines of slope α such that

$$\tan \alpha = \frac{mb}{na}, \text{ where } m \text{ and } n \text{ are integers.}$$

Bragg pointed out that in a three-dimensional array one could *always* find a set of planes such that the angle made by the incident radiation with the plane was equal to the angle made by the reinforced radiation with the plane. Let us look again at our regular cubic array.

Suppose $A'B'$, $C'D'$ (Fig. 15.14) is such a set of planes. Then $\theta + \alpha = \theta' - \alpha$. We shall try to prove that these planes include rows of points within the regular array.

Let

$$\phi = \theta + \alpha = \theta' - \alpha.$$

Then

$$\theta = \phi - \alpha, \theta' = \phi + \alpha,$$

$$\sin \theta = \sin (\phi - \alpha) = \sin \phi \cos \alpha - \cos \phi \sin \alpha,$$

$$\sin \theta' = \sin (\phi + \alpha) = \sin \phi \cos \alpha + \cos \phi \sin \alpha.$$

Adding, we obtain

$$\sin \theta + \sin \theta' = 2 \sin \phi \cos \alpha.$$

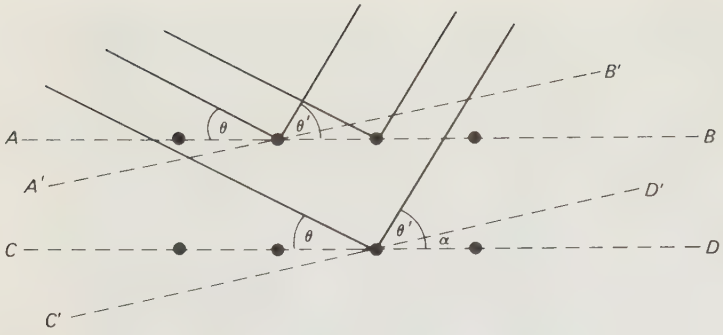


Fig. 15.14

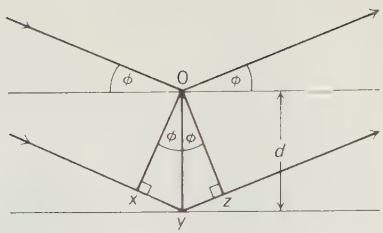


Fig. 15.15

But

$$b\,(\sin\theta + \sin\theta') = l\,\lambda \qquad \text{(from Eq. (15.2))},$$

$$\therefore\ 2b \sin\phi \cos\alpha = l\,\lambda. \tag{15.3}$$

Similarly we can show that

$$a\,(\cos\theta - \cos\theta') = 2a \sin\phi \sin\alpha,$$

and hence

$$2a \sin\phi \sin\alpha = k\lambda. \tag{15.4}$$

Dividing Eq. (15.4) by Eq. (15.3)

$$\frac{2a \sin\phi \sin\alpha}{2b \sin\phi \cos\alpha} = \frac{k\lambda}{l\lambda}$$

$$\therefore\ \frac{a}{b}\tan\alpha = \frac{k}{l}.$$

Hence

$$\tan\alpha = \frac{kb}{al}. \tag{15.5}$$

Since we have already said that k and l must be integers, this is just the condition for the lines to be rows of points within the array.

Let us look again at the condition for constructive interference in terms of a plane for which the incident angle is equal to the angle made by the reinforced beam (Fig. 15.15). $XYZ = \eta\,\lambda$ for reinforcement at incident and scattering angles ϕ. In triangles OXY, OYZ,

$$XY = YZ = d \sin\phi.$$

Consequently

$$2d \sin\phi = n\,\lambda. \tag{15.6}$$

This is the Bragg law for crystal diffraction.

All the strongly reinforced scattered beams can be treated as though they have been reflected from a succession of parallel layers of spacing d. The spacings so worked out will be characteristic of the regular array from which scattering has taken place.

From the set of plane spacings so obtained, it is possible to reconstruct the internal pattern of the array.

15.7 INVESTIGATING THE STRUCTURE OF REAL SOLIDS

In our discussion above we have limited our attention to an array of known order—a large-scale model. In so doing, the radiation used had a wave of about the same length as the diameter of a particle. Atoms are about 10^{-10} m across. To investigate the structure of solids, we shall need a radiation of similar wavelength. Such radiation exists within the electromagnetic spectrum. It is called X-radiation.

The interaction of X-rays with solids has been used both to demonstrate the wave-like nature of X-rays and the regular structure of solids. Here is a case where two models have provided support for each other.

Figure 15.16 may help to explain how this can be so. The affirmative answer to both these problems by

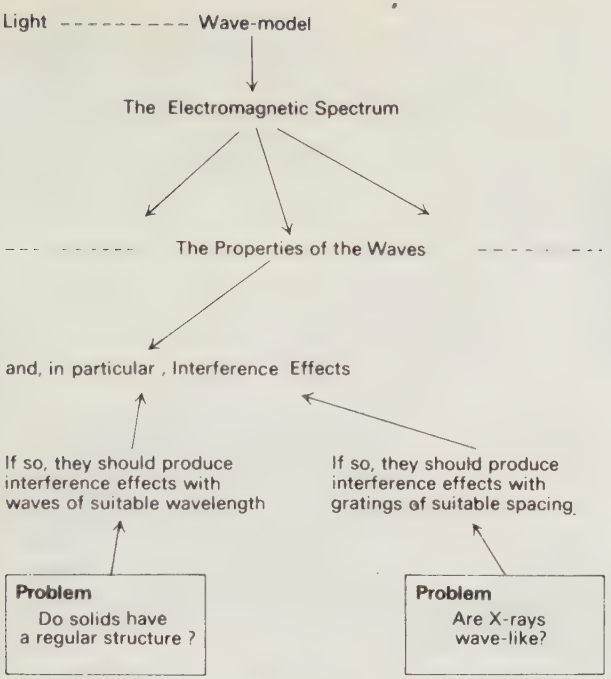


Fig. 15.16

a single experiment serves to confirm both models so long as it is consistent with other evidence. This is another interesting example of the development of theories and models in physics.

The techniques of crystal analysis using X-rays are very important. We will content ourselves here by just noting the various methods which can be used.

a) Rotating crystal

The experiment with the 3 cm wave and model crystal is an analogue for this method which uses a monochromatic beam of X-rays. These may be detected electronically or, more usually, photographically. The interference maxima form a series of spots which indicate the internal regular structure of the crystal. The basic crystal pattern turns out to be a face-centred cubic. In X-ray diffraction patterns further information about the structure can be gained from the intensity of the interference maxima. The intensities will differ because firstly there may be more than one set of planes in the crystal with the same spacing leading to the same angle θ for the direction of an intense beam, and secondly, scattering off layers which are not close packed will in general lead to weaker intensity and thus help to identify the particular planes from which scattering has taken place.

b) Powder method

Instead of rotating the crystal to bring about the necessary conditions for scattering, the crystalline material can be dispersed so that minute crystals are oriented in all possible directions. Some of the planes will always be at the correct angle for their particular spacing. The radiation is scattered into cones as shown in Fig. 15.17. This is a much better method when single crystals are hard to obtain. Figure 15.17 shows a photographic record obtained from copper. The strong interference maxima give evidence of its regular structure: analysis of the photograph shows that copper is a face-centred cubic structure. These photographs provide our first real evidence that the basic particles which go to make up metals are packed together in an ordered way just like the spheres of the large-scale model.

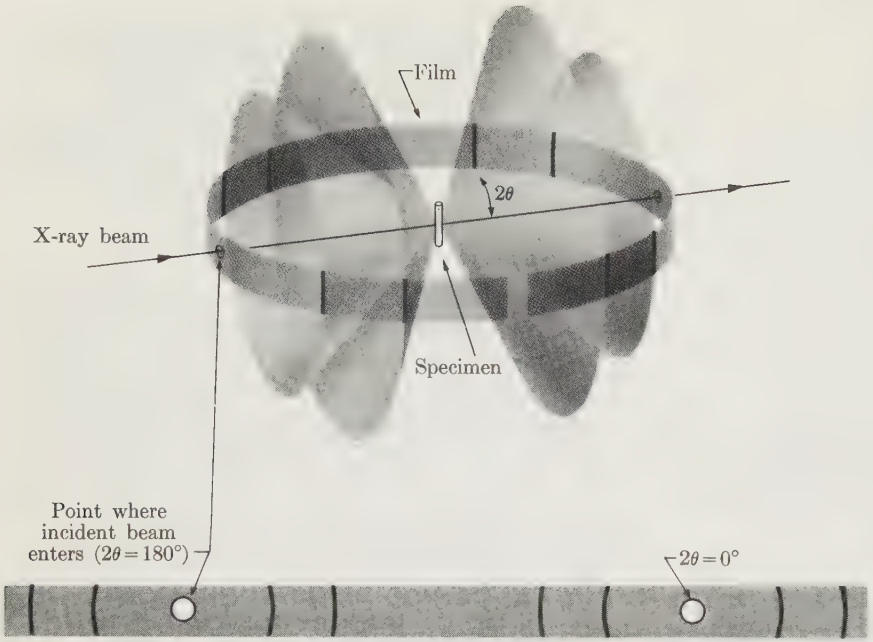
c) Von Laue's method

Instead of varying θ, it is possible to use X-rays of a wide range of wavelengths, so varying λ. In particular directions there will be a strongly scattered radiation such that

$$n\lambda = 2d \sin\theta.$$

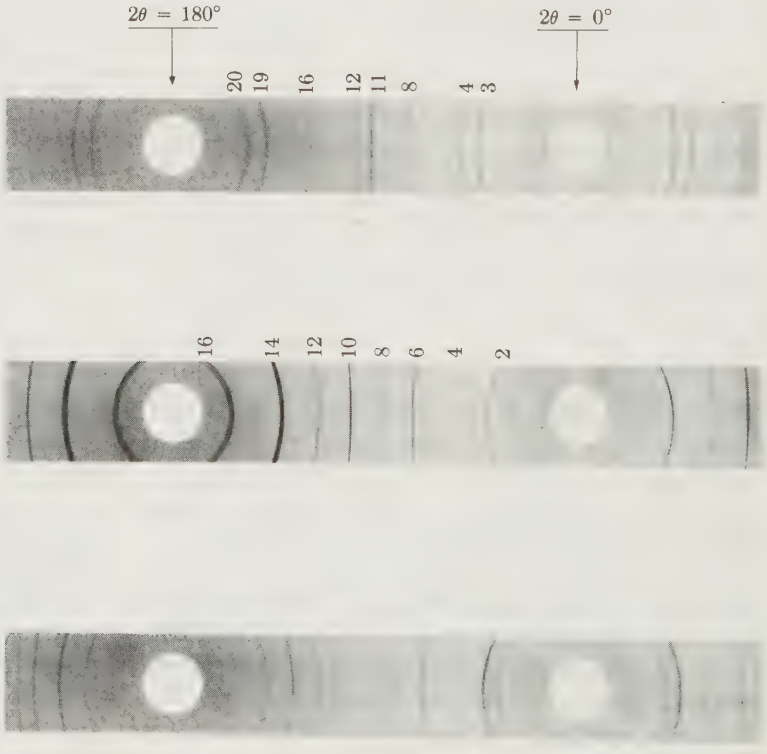
Such patterns are very difficult to analyse as λ will be unknown. They are of great historical interest, being the first X-ray diffraction patterns obtained from crystals and thus serving to justify the model of solids and X-rays we have been using.

In fact, we have been wrong, historically, in suggesting that a model for X-rays and a model for solids pointed the way to an interference experiment which was then performed. Von Laue obtained these curious photographs when irradiating a crystal with X-rays but he was unable to interpret them. It was Sir Lawrence Bragg who proposed an interpretation based on his assumed models for solids and X-rays.



(a)

Fig. 15.17(a) The exposure of X-ray diffraction patterns. Angle 2θ is precisely fixed by the lattice spacing d and the wavelength λ as seen in Eq. (15.6). Every cone of reflection is recorded in two places on the strip of film. (From B. D. Cullity, *Elements of X-ray Diffraction*, Addison-Wesley, 1956.)



(b)

Fig. 15.17(b) X-ray diffraction pattern for copper. (From B. D. Cullity, *Elements of X-ray Diffraction*, Addison-Wesley, 1956.)

Chapter 16
APPLYING THE MODEL FOR MATTER

16.1 Surface energy and surface tension
16.2 Plastic deformation
16.3 The thermal and electrical properties of metals
16.4 Conclusion

How does the model which we have now developed apply to particular cases: to a wire hawser under tension, for example?

We will need to bear in mind the force/separation and energy/separation diagrams (Fig. 14.9). These indicate that potential energy must be provided if the strong attractive forces which develop under tension are to be overcome at all.

We will need to examine again the stress/strain graph (Fig. 14.12) for a typical metal. The first part of this graph is linear; the stretching is elastic and Hooke's law

$$\text{stress} \propto \text{strain},$$

applies.

Stress has been defined as the force per unit cross-sectional area; strain as the extension per unit length.

$$\text{stress} = \frac{F}{A} \qquad \text{strain} = \frac{x}{l}.$$

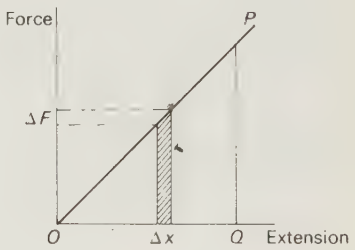


Fig. 16.1

The graph of force against extension (x) must also be a straight line so long as the extension is elastic (Fig. 16.1). Then $F = -kx$ where F is the force, x is the extension and k is a constant. If, as a result of increasing the force F_1 by a small amount ΔF_1, the extension also increases from x to $x + \Delta x$, the energy transformed may be taken to be $F_1 \Delta x$ since F_1 is, to a close approximation, the mean force acting during this change. It can be seen that this is equal to the shaded area on the graph in Fig. 16.1.

The total energy involved in the extension of the solid by a length x is the area OPQ on the graph, i.e.,

$$\tfrac{1}{2}OP.OQ,$$

or

$$\tfrac{1}{2}Fx.$$

So the energy stored $= \frac{1}{2}Fx$

$$= \frac{1}{2}\left(\frac{F}{A}\right)\left(\frac{x}{l}\right) Al \qquad (16.1)$$

or $\frac{1}{2}$ stress $\times$ strain $\times$ volume.

It is often convenient to express this as the strain energy per unit volume.

$$\text{Strain energy per unit volume} = \tfrac{1}{2} \text{ stress} \times \text{strain}.$$

$$\text{Since } \frac{\text{stress}}{\text{strain}} = \text{Young's modulus}$$

$$\text{strain energy per unit volume} = \frac{1}{2}\frac{(\text{stress})^2}{\text{Young's modulus}}. \qquad (16.2)$$

Energy stored as strain energy may be quite considerable. Consider, for example, a steel hawser of cross-sectional area 10^{-3} m^2 and length 10 m which is experiencing a tension of 2×10^6 newtons. If Hooke's law can be applied (and the breaking force for such a hawser might be 3×10^6 newtons so this is not an unreasonable assumption) then

strain energy per unit volume

$$= \frac{1}{2} \cdot \left(\frac{2 \times 10^6}{10^{-3}}\right)^2 \times \frac{1}{2 \times 10^{11}} \text{ J m}^{-3}$$

since Young's modulus for steel is about 2×10^{11} N m^{-2}

$$= \frac{1}{2} \cdot \frac{4 \times 10^{12}}{10^{-6}} \times \frac{1}{2 \times 10^{11}}$$

$$= 10^7 \text{ J m}^{-3}.$$

Total energy stored in the hawser whose volume is $(10^{-3} \times 10)$ m^3 is

$$10^7 \times 10^{-2} \text{ J} = 10^5 \text{ J}.$$

Problem 1 What is the speed of a truck (mass 2000 kg) which possesses the same kinetic energy as the hawser?

This is a very considerable amount of energy and, should the hawser break under tension, considerable damage might be done as this strain energy is released and, almost entirely, converted to kinetic energy.

Should this happen, two adjacent layers of atoms must become separated from one another. Two new surfaces are formed. This also involves an energy change.

Each particle in the solid lattice will have a number of near neighbours. Within the body of the solid, these are likely to be regularly distributed in the space around the particle we are interested in. The potential energy of the particle is almost entirely due to the effects of the nearest neighbours, for we have already seen that the energy changes very rapidly with distance. If there are q near neighbours and E_0 is the energy required to separate a particle from its neighbour, the energy which binds it in position is qE_0. To remove such a particle qE_0 units of energy must be supplied. The simplest way of supplying such energy is by heating—and this may result in the sublimation of the solid. In this case the heat of sublimation is a measure of the energy required to break the inter-particle bonds in the solid.

If E_0 is the bond energy and q is the number of near neighbours which each atom possesses, the energy of sublimation (L_s) per mole is given by

$$L_s = \tfrac{1}{2}N_A q E_0.$$

The factor of $\frac{1}{2}$ takes account of the fact that each bond is shared between a pair of atoms. Now for a close-packed structure q is 12 (see Fig. 16.2), so that

$$L_s = 6N_A \cdot E_0. \qquad (16.3)$$

This simple relationship gives a very good guide to the value of the bond energy for certain substances (notably solid neon, argon, krypton, and xenon) but unfortunately not for the metals. This suggests that our model is not without faults and may well require modification later. Another approach must be adopted for metals.

A particle which is near to the surface of a solid has fewer near neighbours than one which is embedded in the interior. We have just seen that, in a close-packed structure, an atom which is well within

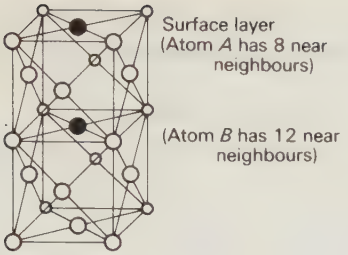


Fig. 16.2

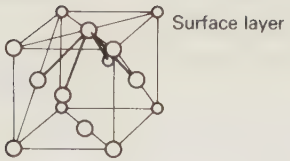


Fig. 16.3

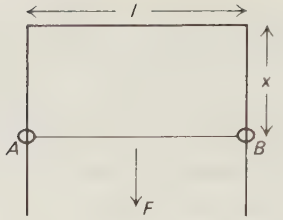


Fig. 16.4

the crystal lattice has 12 near neighbours. One such atom on the surface of the lattice will have only 8 near neighbours (see Fig. 16.2). Such an atom will have a different potential energy from one in the body of the structure. Now, we have seen that it is convenient to regard the potential energy of an atom as zero when it is at a very large distance from any neighbours; and we have seen how an atom within a lattice is in equilibrium about a minimum potential energy E_0. So the atom in the surface has a higher potential energy than the one within, and the surface as a whole must possess surface energy. It follows that the creation of new surfaces by, say, breaking the specimen in two, requires the supply of this energy.

Consider a rectangular bar of metal of cross sectional area A. To break it a total of four bonds must be broken for each atom which exists in the surface (see Fig. 16.3) so that a total energy of $4E_0N_s$ must be supplied. (N_s is the number of atoms in the surface layer.) The area of surface which is created by breaking the bar is $2A$ so that the surface energy per unit area (γ) is given by

$$\gamma = \frac{4E_0N_s}{2A} = 2E_0\frac{N_s}{A}.$$

Now we have already seen that the energy needed for the sublimation of N_A atoms is

$$L_s = 6E_0N_A.$$

So the energy needed for the sublimation of N_s/A atoms is $6E_0(N_s/A)$ which is just three times the surface energy of the single layer.

16.1 SURFACE ENERGY AND SURFACE TENSION

The ability of a liquid to form sperical drops is a well known property which must relate to the property of surface energy. The spherical shape is the one which contains a maximum volume for a minimum of surface area, so that the liquid in the drop has a minimum surface energy.

It is often convenient to present this in another way. Consider, for example, a liquid film (with two surfaces) of width l and length x, bounded by a wire AB which is free to slide (Fig. 16.4). If a force is applied to the wire, the surface may be extended by a length Δx so that a new surface of area $2l\,\Delta x$ appears. The energy provided in this process is $\gamma.2l\,\Delta x$. The work done by the applied force is $F\,\Delta x$.

So

$$F = 2l\gamma$$

or

$$\gamma = \frac{F}{2l}. \qquad (16.5)$$

The surface energy is then equal, numerically, to the force per unit length. This line tension (force per unit length at right angles to a line drawn in the surface) provides an alternative measure, which in the case of liquids, is easy to make. Known as the surface tension of the liquid, it will be quoted in newtons per metre. For water it is about 7×10^{-2} N m^{-1} so that the surface energy of water is about 7×10^{-2} J m^{-2}.

In this case, we have assumed that no changes have occurred in the surface structure—and, because

the molecules within the liquid are so mobile, this is likely to be true. But it will not be true where a solid is concerned. Consequently the surface energy and the surface tension for a solid will be related but in a more complex way than for the liquid.

Indeed, in this case we may write: energy transferred per unit area in increasing the area of a solid surface is

$$\frac{\mathrm{d}}{\mathrm{d}A}(A\gamma) = \gamma + A\frac{\partial\gamma}{\partial A}.$$

In the case of the liquid with its mobile molecules, $\partial\gamma/\partial A$ is zero but in the case of a solid it is not. As the surface stretches so the atoms are pulled apart and the surface energy diminishes so that $\partial\gamma/\partial A$ is negative.

We would expect, then, that the measurement of the surface energy of a solid would be difficult. In the case of copper a reliable estimate has been made by Udin, Shaler, and Wulff (*Metals Transactions*, 1949, p. 186). Fine copper wires support small masses in an evacuated chamber at a high temperature so that the metallic creep is appreciable even under small stresses. If the weight exceeds the surface tension the wire stretches; if not, the wire shrinks. At the critical load, the wire does not creep at all and the surface tension (which, as we have seen, is related to the surface energy) is equal to the weight. The surface energy of copper was found by this method to be about 1.4 J m^{-2}.

In spite of these practical difficulties, a knowledge of the surface energy is well worth having, for, not only does it help us to determine the bond energy but it also enables us to calculate the breaking stress (tensile strength) of a metal and so check the application of the model to a real material.

Consider two adjacent layers of particles separated by a distance x. Then a layer of unit area will have a volume of x. If this material is under stress, the strain energy per unit area of either layer will be

$$\frac{1}{2}\frac{1}{E}(\text{stress}^2)\ \text{volume} = \frac{1}{2}\frac{1}{E}(\text{stress}^2)\,x.$$

If two new surfaces are to be created as the layers are pulled apart, then this is the amount of energy which must be provided and shared to give surface energy of γ to each. We may write

$$2\gamma = \frac{1}{2}\frac{1}{E}(\text{stress}^2)\,x$$

i.e.,

$$(\text{stress}^2) = \frac{4\gamma E}{x}$$

and

$$\text{breaking stress} = 2\sqrt{\frac{\gamma E}{x}}. \tag{16.6}$$

In making this estimate we have assumed that Hooke's law applies right up to the moment of fracture: we know that this is not so. The stress must therefore be too high—but it is unlikely to be more than two or three times too high. If it is twice too high we would expect to find that

$$\text{breaking stress} = \sqrt{\frac{\gamma E}{x}}.$$

To test the argument, let us insert the known figures for steel.

Surface energy of steel is about 1 J m^{-2}.

Young's modulus for steel is about 2×10^{11} N m^{-2}.

Atomic spacing in steel is about 2×10^{-10} m.

So tensile strength (breaking stress)

$$= \sqrt{\frac{1 \times 2 \times 10^{11}}{2 \times 10^{-10}}}\ \mathrm{N\ m^{-2}}$$

$$= \sqrt{10^{21}}\ \mathrm{N\ m^{-2}}$$

$$\approx 3 \times 10^{10}\ \mathrm{N\ m^{-2}}.$$

Unfortunately, a typical breaking stress for steel is only about 4×10^{8} N m^{-2} and even a specially prepared steel can claim no more than about 3×10^{9} N m^{-2}.

Moreover, calculations for other metals reveal similar discrepancies. Either the logic of the argument or the model of atoms in layers which we have used must be at fault. Of the two, the argument is the easiest to check, and, even allowing for the guess which we made when we observed that Hooke's law

cannot be applied in this case, it does not seem likely that our error could be as much as two orders of magnitude out.

Strangely, the problem, which had been well known for a long time, was not resolved until about 1920. Then A. A. Griffith made a very important change in the model.

Griffith had been experimenting with thin glass fibres and these were found to get stronger as they became thinner. Indeed the strength of a fibre of diameter 2.5×10^{-3} mm was as much as a quarter of the calculated strength, whereas a glass rod of about 1 mm diameter broke at about one eightieth of the calculated strength.

Griffith suggested that some mechanism must be responsible for concentrating the stress so that, locally in the glass, the stress rose to higher values than expected. This idea of stress concentration had risen seven or eight years before when Professor Inglis was investigating the causes of failure in the hulls of steel ships. Now Griffith extended the idea from the macroscopic to the microscopic scale.

Consider an atomic lattice with its regular array of bonds and particles under tension (Fig. 16.5). Each bond contributes to the whole strength of the material. But suppose that there is a fault or a crack as at C in Fig. 16.6.

At the tip of this crack there will be an increase in the concentration of the stress. The inter-particle bonds are broken along the length of the crack; the whole stress is taken up at the tip. Although the stress which is applied to the material is well below the anticipated breaking stress, just at the tip of the crack the stress will be much higher than we imagine. The bonds between the particles in the vicinity of the crack tip are the ones which are most likely to fail under stress.

Griffith suggested that minute cracks, smaller perhaps than a wavelength of light, would serve to reduce the strength of his glass fibres. It was many years before suitable techniques were developed which enabled men to reveal the crack patterns, but Griffith's analysis had been accepted by many long before this. The particle model could now accommodate brittle fractures.

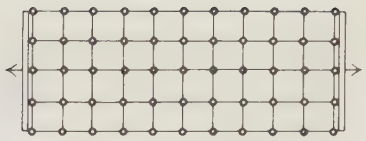


Fig. 16.5

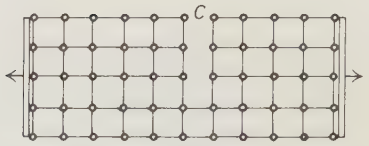


Fig. 16.6

An excellent example of the effect of a crack in weakening a material is to be found in the breaking of a glass tube or the cutting of a pane of glass. All that the glass worker does is to initiate a crack with a suitable knife and then to apply a suitable stress. The glass breaks cleanly along the line of the crack.

16.2 PLASTIC DEFORMATION

Metals usually behave in quite a different way however. The stress/strain curve (Fig. 14.12) for a typical metal bends over somewhat after the linear reversible region (in which Hooke's law applies) is passed. If now the stress is relaxed, the metal returns along a new curve; it has suffered permanent, or plastic, deformation.

This process is often accompanied by a well-known phenomenon called work-hardening. This is easily illustrated. Take about 20 cm of new copper wire of diameter about 3 mm. Bend the end to make a small hook and clamp the other end firmly so that the wire is horizontal. Hook a lightweight carrier to the end and observe what happens as you increase the load. The copper wire will bend readily. Now take it between your fingers and bend it between them along its whole length several times. Repeat your test. Now the wire will support a much increased load before yielding. This is known as work-hardening—and its explanation demands yet another modification of the model.

These two phenomena—that of plastic deformation and of work-hardening—can hardly be accounted

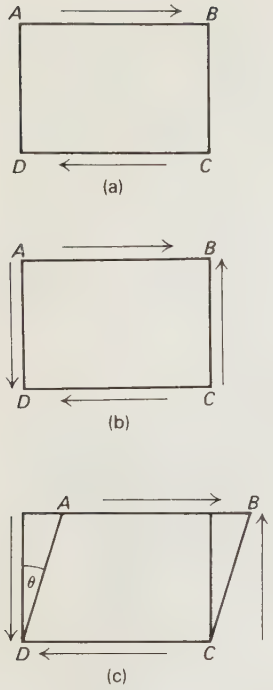


Fig. 16.7

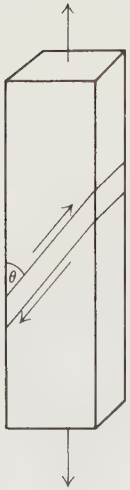


Fig. 16.8



Fig. 16.9

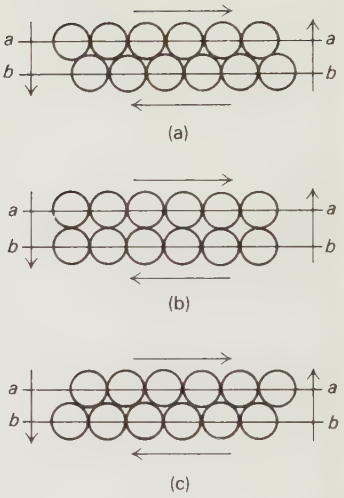


Fig. 16.10

for in terms of the simplified structure of Chapter 15. When discussing that structure we noted that the model used was simple in the extreme. The layers of atoms which were being examined were assumed to be stressed at right angles to the direction of the applied force. It is far more likely that we should be considering layers of atoms which are inclined to that force.

Layers so stressed are under a shearing stress and, before proceeding, it will be necessary to consider such stresses in general. For simplicity let us take a block of material *ABCD* and imagine that a pair of forces is applied as shown in Fig. 16.7(a). If the block is not merely to turn round, a second pair of forces must be applied as in Fig. 16.7(b). The result may be that the material is sheared through the angle θ. This is a measure of the strain.

The shearing stress is F/A where A is the area of the surface of the block through which the force F is applied. Experiment shows that the ratio

$$\frac{\text{shearing stress}}{\text{strain}}$$

is a constant. The name *shear modulus* or *rigidity modulus* (G) is given to this ratio.

Returning now to the case of the metal bar under tension (Fig. 16.8). Suppose that the layer shown represents two such planes of atoms in a bar subject to a tensile stress of F/A where A is the cross-sectional area and F the applied force. On one side of the layer shown there is a force $F \cos \theta$; on the other side $F \cos \theta$ in the opposite direction. As we have seen, such a pair of forces constitutes a shear.

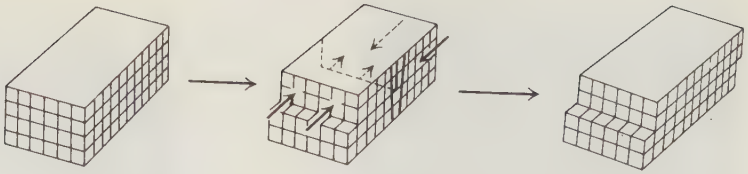


Fig. 16.11 Formation of an edge dislocation. (From L. H. Van Vlack, *Elements of Materials Science*, Addison-Wesley, 1964.)

Should two such layers move relative to one another under this shearing action, then the material could well acquire a permanent elongation (Fig. 16.9).

Let us examine this shearing action in atomic terms, considering a close packed structure (Fig. 16.10a).

If layer *a* slips relative to layer *b* by the application of pairs of forces as shown, the resistance to slip will increase to a maximum at some position intermediate between the initial position of Fig. 16.10(a) and the position of Fig. 16.10(b). In this latter position all resistance to the shear has gone and the layer *a* is quite likely to slip down to the position of Fig. 16.10(c). The angle of shear is now 60°. As a first approximation we can take about one quarter of this as the angle of shear corresponding to the maximum resistance to shear. This is 15° or, in radians, about 0.25. If the material was steel, with a shear modulus of about 0.8×10^{11} N m^{-2}, we would expect the maximum shearing stress to be

$$\text{shear modulus} \times \text{shear angle}$$

$$\approx 0.8 \times 10^{11} \times 0.25 \text{ N m}^{-2},$$

$$\approx 0.2 \times 10^{11} \text{ N m}^{-2}.$$

Unfortunately, experiment reveals that the maximum shearing stress for steel is about one fiftieth of this and so the angle of shear must be much less than 15°. Yet another major discrepancy between a prediction made on the basis of the model and the reality is revealed.

Metals start to deform plastically under stresses which are far smaller than those predicted by theory. It is, in fact, this effect which confers upon them the very important property of ductility. Ductile materials will undergo a considerable amount of irreversible distortion before breaking. In some cases, for example the work-hardening of copper, this may even strengthen the metal. In any case, this ability of ductile materials to accept permanent distortion and therefore to be shaped (whether by cold or by hot working) is a very important property.

It was G. I. Taylor who, in 1934, advanced an explanation of the discrepancy between the calculated and the actual shear strength for metals. He pointed out that the theoretical calculation presupposes a perfect crystal structure. If, however, there are imperfections within the crystal, these imperfections might themselves be able to travel through the material of the crystal and, in so doing, lower the shear strength.

If the shearing forces act as shown in Fig. 16.11 the right-hand rows of atoms can move relative to the left-hand rows and the extra row will be moved to the edge of the crystal. The bonds at the end of that extra row will be highly stressed so that the application of a small shearing force will cause them to break down and slip to occur. But immediately the bonds are re-established with the next pair of rows and so on until the extra row reaches the edge. Such an extra row of particles is known as an edge dislocation. Other forms are known. Figure 16.12 shows a photograph of a dislocation.

Such a defect in the structure can move around quite freely and, sooner or later, will meet another dislocation. Usually, however, dislocations tend to keep away from one another and this impedes their freedom of movement. A metal with many interlocked dislocations in its crystal structure will be harder than a metal with only a few. It is this which accounts for the work-hardening process.

This photograph illustrates the difficulty of presenting the evidence for this view of G. I. Taylor's.

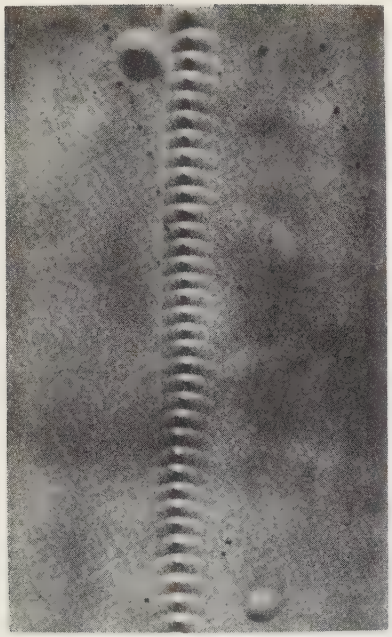


Fig. 16.12 Germanium crystal etched to show the ends of edge dislocations (×1000). (From L. H. Van Vlack, *Elements of Materials Science*, Addison-Wesley, 1964.)

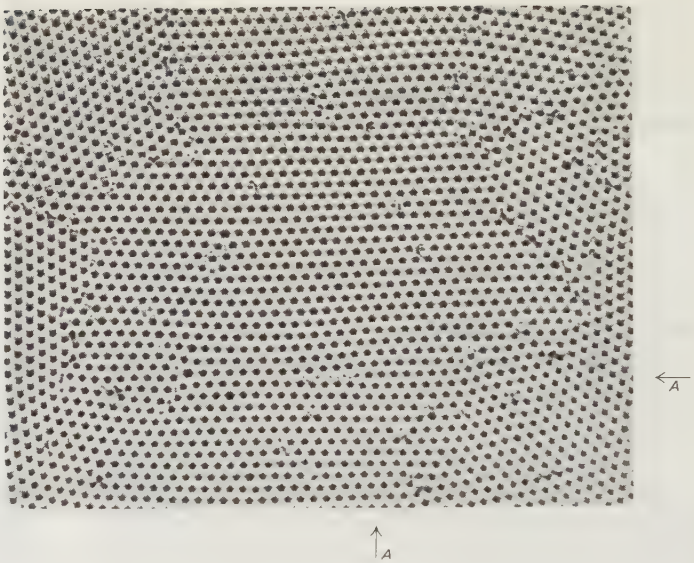


Fig. 16.13 A bubble raft showing grain boundaries and, at *A*, a dislocation. (Photograph by E.J.W.)

But a simple analogue in two-dimensions is available in the bubble raft of W. L. Bragg (Fig. 16.13).

A raft of uniform sized bubbles about a millimetre in diameter is easily formed by blowing gently through a jet held just below the surface of a detergent solution. The depth of the jet below the liquid surface should be kept fixed. These bubbles will adhere to one another, i.e., inter-bubble attractive forces and also forces of repulsion will exist. The close-packed structure of the raft is clearly seen and so is the existence of "bubble crystal grains." At a point such as *A* in Fig. 16.13 a bubble is missing—and a dislocation exists. If, now, shearing forces are applied to such a raft, the motion of the disclocation is very clear to see.

The presence of atoms of a different size from the majority of the atoms which make up the lattice will also play an important part in the interaction of dislocation with dislocation.

16.3 THE THERMAL AND ELECTRICAL PROPERTIES OF METALS

We can yet again refer to the force/distance graph (Fig. 16.14a) to seek information about a different set of properties of solid materials. There is, we saw, an equilibrium separation (r_0) at which the energy of a particle with respect to its neighbour is a minimum. At this distance, the particle is subject to a net zero force; but should it be displaced, there will immediately come into action a restoring force. If the particle is displaced towards the y axis this force is one of repulsion, but if the particle is displaced away from the y axis, then the restoring force is one of attraction. Displacement under these conditions can only cause the particle to oscillate.

The addition of further energy to this system will cause the amplitude of the oscillations to increase. Consider the potential energy/distance graph (Fig. 16.14b). When the particle has its minimum of

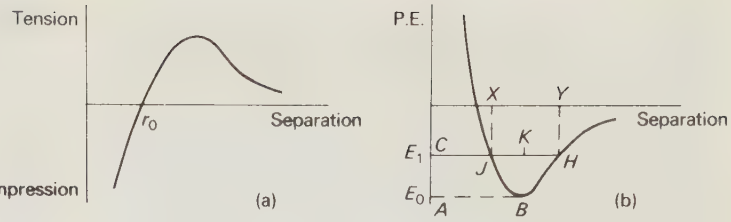


Fig. 16.14

energy E_0 its separation from its neighbour is r_0 (or AB). Now suppose that the total energy of the particle rises to E_i. The particle can now exist in oscillation at any point along the line JH. If K represents the average position of this particle, we see that the displacement KC exceeds the original displacement AB because the energy/displacement curve is not symmetrical. So the particle is, on average, a little further away from A than it was before the energy was supplied. As the energy is further increased (and this corresponds to a further rise in temperature) the asymmetry of the curve, which is more pronounced, will ensure that the particle moves even further away, on average, from its initial position. Extending the whole argument to the complex lattice structure of a metal, it can be seen that the model leads us to the expectation that metals will expand on heating and also to the conclusion that the coefficient of this expansion would itself increase with temperature. And this it does.

As a metal lattice absorbs energy from a heat source, its internal energy is increased. This energy is available within the vibrations of the atoms as the sum of the potential and kinetic energies of the atoms. (See Unit 5.)

As long ago as 1819, Dulong and Petit observed that most solid elements absorbed nearly the same amount of energy for unit temperature rise if the masses chosen for comparison were proportional to the atomic masses. Consider, for example, the molar heat capacities of some metals (Table 16.1). The product in each case lies around the value 25 J mol^{-1} K^{-1}. This is often referred to as Dulong and Petit's law.

However, experiments show that the value of the specific heat capacity per mole falls as the solid is cooled. This suggests that there must be some restrictions on the way in which the atomic oscillators can accept the energy proffered to them. This points to a need to modify one of the two models we are using at this point—either the model of the solid itself or the model of energy. That it is the model of energy which needs modification is a very important matter to which we must return after the properties of oscillators themselves have been examined. (See Unit 5.)

Table 16.1

Element	Specific heat capacity (c) J kg^{-1} K^{-1}	Molar mass (m) kg	Molar heat capacity ($m \times c$) J mol^{-1} K^{-1}
Aluminium	923	0.027	24
Copper	390	0.064	24.8
Lead	126	0.207	26.1

If one end of a solid bar is heated, it is evidently possible for the vibration of the atomic oscillators near to the heat source to be transmitted from particle to particle through the solid; energy is conducted away from the heat source mainly by the vibration of the particles which constitute the lattice. On such a view, one would expect that the thermal conductivities of all solids would not differ widely from one another. And yet there are excellent solid thermal insulators and there are excellent solid thermal conductors (especially such metals as copper and silver). The range of thermal conductivities, written λ in the equation

$$\frac{\Delta Q}{\Delta t} = \frac{\lambda A(T_2 - T_1)}{L}, \qquad (16.7)$$

where $\Delta Q/\Delta t$ is the rate of flow of energy through a slab of the material of cross sectional area A and thickness L when the temperature difference between the faces is $(T_2 - T_1)$ is about 1 to 1000.

Whilst the model is adequate for thermally insulating solids, it is clearly inadequate for good thermal conductors. This becomes even clearer when the electrical properties of solids are considered. On the one hand are the electrical insulators; on the other, the electrical conductors (again headed by such metals as silver and copper). The range of electrical conductivities written σ in the equation

$$i = \frac{\sigma A(V_2 - V_1)}{l}, \qquad (16.8)$$

where i is the current through the conductor of cross sectional area A and length L when the potential difference between the two ends is $(V_2 - V_1)$, is even greater than in the case of heat conductivity. It is about 1 to 10^{10}.

Such discrepancies demand a mechanism for both electrical and thermal conduction in metals and other good conductors which is far more effective than lattice vibrations. This is a matter to which we must return after some consideration of the electrical nature of matter in later units.

Problem 2 An aluminium kettle containing boiling water is in contact with a gas flame at a temperature of about 550° C. The water is found to be evaporating at a rate of 0.1 kg per min. The base of the kettle is 1 mm thick and its area is 0.015 m^2. Estimate the temperature of the outer surface of the base of the kettle. (Specific latent heat of vaporization of water is 2.27×10^6 J kg^{-1} and the coefficient of thermal conductivity of aluminium at this temperature is 2.1×10^2 J m^{-1} s^{-1} K^{-1}.)

Since the water is vaporizing at the rate of 0.1 kg per min and the specific latent heat of vaporization of water is 2.27×10^6 J kg^{-1}, energy is being transmitted through the aluminium base at a rate

$$\frac{\Delta Q}{\Delta t} = \frac{0.1}{60} \times 2.27 \times 10^6 \text{ J s}^{-1}.$$

Substituting in the equation, we have

$$\frac{\Delta Q}{\Delta t} = \frac{\lambda A(T_2 - T_1)}{L}$$

$$\frac{0.1}{60} \times 2.27 \times 10^6 = \frac{2.1 \times 10^2 \times 0.015 \times (T_2 - T_1)}{10^{-3}}$$

$$T_2 - T_1 = \frac{0.1}{60} \times \frac{2.27 \times 10^6 \times 10^{-3}}{2.1 \times 10^2 \times 0.015}$$

$$= 1.2°\text{C}.$$

The temperature of the outside surface of the base of the kettle is about 1°C higher than the water temperature within. The metal is "in contact" with a hot flame at about 500°C. This suggests that the major temperature change must occur in the surface layers of gas.

16.4 CONCLUSION

The development of theories and models in physics is a very complex process. It is a creative act which goes on alongside the process of observation and experiment. In retrospect, it is always easy to describe a model or theory and then produce an amount of experimental evidence in support of this model.

When a particular model is well-established, this is often the most economical method of approaching it. Such is the case with the model for matter. In a future chapter we shall go on to investigate this model further. We shall find that further modifications are necessary in order to produce something whose properties correspond to the properties of real materials.

The form of the model with which we shall end is now well established in physical theory. An engineer, interested in using materials; a technologist interested in improving their properties; a chemist interested in producing new materials will not concern themselves with the details of the whys and wherefores. They will take, and learn to use, the best model available to achieve the end in view. But these terms *engineer*, *technologist*, and *chemist* are in many ways shades of meaning applied to those involved in the study of physical science. They reflect the particular interest

of these people. But all are involved in the processes typified by the work of Boyle. Every so often new observations will be made which do not fit in with accepted models and ideas. The model or theory must be modified. This is the process of creative science. And at these moments it is necessary to ask why science has accepted its present ideas.

There are those scientists today who are engaged in what is often called fundamental research. These are people who continually ask and probe the question "Why?" There are others—the great majority—who are engaged on the application of present knowledge, extending and using the models already developed. But it is only in recent years that we have seen such an apparent division of "scientific" study. The history of science suggests that these two activities cannot be successfully divided. Boyle, who is chiefly remembered today for his law, must have been well known in his own day as a theorist and speculator, and not just an investigator and observer. His friend, Robert Hooke, again remembered for his law, was perhaps better known in his time as an architect—a man to whom the *application* of stresses and strains was everyday business.

PROBLEMS: Unit Four

4.1 What volume does a mole of an ideal gas occupy at a temperature of 273K and a pressure of 1.01×10^5 N m^{-2}? The ideal gas constant R is 8.3 J mol^{-1} K^{-1}.

4.2 A gas cylinder has a volume of 0.4 m^3. It contains nitrogen at an absolute pressure of 20×10^5 N m^{-2} and a temperature of 20° C. Nitrogen is withdrawn from the cylinder at a steady rate and the absolute pressure falls to 10×10^5 N m^{-2} whilst the temperature of the gas remaining falls to 10° C.

a) What was the mass of nitrogen in the cylinder initially?
b) How much was withdrawn?
c) What volume would the withdrawn nitrogen occupy at the normal atmospheric pressure of 10^5 N m^{-2} and a temperature of 20° C?

4.3 A motor car tyre is stated to be correctly inflated at a pressure of 26 lb/in^2. In this system the atmospheric pressure is about 15 lb/in^2, which in the S.I. is about 10^5 N m^{-2}.

a) What is the absolute pressure within the tyre, in S.I. units?
b) A pump is full of air at normal atmospheric pressure and the length of the stroke is 15 cm. At what point in the stroke does air begin to enter this tyre? Assume the compression to be isothermal.

4.4 A motor car tyre is correctly inflated to an absolute pressure of 2.7×10^5 N m^{-2}.

a) What is the reading of a tyre gauge connected to it if the atmospheric pressure is 10^5 N m^{-2}?
b) If the volume of the tyre is 0.035 m^3 and the temperature of the air is 20° C, how many moles of air does the tyre contain?
c) If the tyre supports one quarter of the weight of the car (mass 1500 kg) what area of surface of the tyre must be in contact with the road?
d) If the car is driven at high speed in hot weather what changes will occur to the air in the tyre? What steps would you advise the driver to take?

4.5 Compute a value for the r.m.s. velocity of molecules in the air at a pressure of 10^5 N m^{-2} given that the density of air is 1.2 kg m^{-3}.

A tightly-fitting cork is pulled from a bottle of ammonia at the far end of a draught-free corridor. How does the kinetic theory of gases account for

a) the time taken for the sound to reach the other end of the corridor,
b) the time taken for the smell of the ammonia to reach the same point?

Account for the wide disparity in these times.

Would these times be increased or decreased if the experiment were performed in a building on a mountain at such a height that the pressure was reduced to one-half? Explain your reasoning carefully. You may assume that the air temperature has remained constant (Velocity of sound in air is 340 m s^{-1}).

University of Birmingham
Degree of B Ed: 3rd year Paper.

4.6 How high would the atmosphere be if the density of the air were maintained at the value it has at s.t.p.?

Calculate the velocity of a particle which falls from rest through this height assuming that no retarding forces act.

4.7 What is the total random molecular kinetic energy in a mole of helium gas at temperatures of 300 K and 301 K?

Table P4.1

Speed range in m s^{-1}	0 – 99	100 – 199	200 – 299	300 – 399	400 – 499	500 – 599	600 – 699	700 – 799	800 – 899	900 – 999	1000 – 1099	1100 – 1199
% of molecules within the range	1	5	14	20	21	16	10	6	3	2	1	<1

How much energy has to be supplied to the gas in order to raise its temperature through 1 K?

4.8

a) Plot a histogram to display the distribution of speed among the molecules of nitrogen at 0° C from the information given in Table P4.1.

b) Calculate the r.m.s. velocity of nitrogen at this temperature. Does this value agree with the value you would expect to get by considering the graph? Explain.

4.9 The air in a room is to be maintained at a temperature of 20° C when the air outside is at a temperature of 0° C. The window of the room has an area of 2 m^2 of glass of thickness 2 mm. The thermal conductivity of glass is 1.05 J m^{-1} s^{-1} K^{-1}. The temperature drop across the window is found to be 0.2° C. At what rate must energy be supplied to the air in the room to maintain this temperature assuming that all the heat losses are confined to the window? Is your answer a realistic one?

4.10 A student is told that the air in a room is at a temperature of 20° C and the air outside is at 0° C. The window has an area of 2 m^2 and the glass, which is 1.5 mm thick, has a thermal conductivity of 1.0 J m^{-1} s^{-1} K^{-1}. He computes the rate of loss of energy and finds it to be 27 kW. How did he arrive at this figure? Discuss whether or not this is a probable rate of loss of energy. What additional information would you need to compute the real rate of loss?

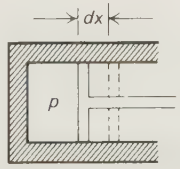


Fig. P4.1

4.11 Figure P4.1 shows a well insulated cylinder in which some fluid is contained by a piston. The pressure inside the cylinder is p, the cross-sectional area of the piston is A. The piston moves out an infinitesimally small distance dx. How much work is involved in this expansion and in an expansion from V_1 to V_2?

The cylinder contains 1 mole of nitrogen gas at a temperature of 273 K and the piston maintains a constant pressure on the gas of 10^5 N m^{-2}. The cylinder and the gas are then heated to a temperature of 546 K.

a) How much work is done?

b) How much energy was supplied?

c) By how much did the internal energy of the gas change?

4.12 At atmospheric pressure (assumed to be 10^5 N m^{-2}) 1 kg of water occupies 1000 cm^3 and 1 kg of water vapour occupies 1 670 000 cm^3. The specific latent heat of water is 2.27×10^6 J kg^{-1}. Find:

a) the work done if 1 kg of water is converted into steam at this pressure

b) the increase in the internal energy of the system.

4.13 A steel wire 3 m long and with cross sectional area 0.20×10^{-6} m is used to support a mass of 5 kg. By how much will the wire stretch? (Young's modulus for steel = 200×10^9 N m^{-2}).

4.14 The elastic limit of a steel cable used in a lift or elevator is 200×10^6 N m^{-2}. If the stress is not to exceed 25% of this figure what is the maximum upward acceleration which can be given to the car of total mass 1000 kg? The cable has a cross-sectional area of 3 cm^2.

4.15 A steel rod is used to secure a vessel to a motor so that the vessel and its contents can be whirled around in a horizontal circle. The rod has a cross sectional area of 1.5 cm^2 and the stress must not exceed 25% of the elastic limit for the steel (200×10^6 N m^{-2}). What is the maximum speed of rotation which can be reached safely if the total mass of the vessel is 0.5 kg and the centre of mass of the system is 20 cm from the axis of rotation?

4.16 The tensile strength of a material depends not only upon the nature of the material but also upon its cross-sectional area.

The legs of animals have sufficient strength to withstand the weight of the animal.

Use these facts to explain why the legs of a mouse are relatively much thinner than the legs of an elephant.

An explorer of the planet Jupiter reports seeing ant-like creatures 1 m high. Would you accept this? Give a reason for your answer.

4.17 The linear dimensions of a simple pendulum made up of a spherical bob and a steel wire are scaled up by a factor of three. By what factor do each of the following change

a) the volume of the bob,
b) the mass of the bob,
c) the surface area of the bob,
d) the volume of the wire,
e) the mass of the wire,
f) the surface area of the wire,
g) the tensile strength of the wire,
h) the period of the pendulum?

4.18 The wire supporting the bob of Problem 4.17 will be stretched. What change will occur in the stretch of the wire when the pendulum is scaled up by the factor of 3? ($E = 200 \times 10^9$ N m^{-2}).

BIBLIOGRAPHY: Unit Four

On kinetic theory

Brush, S. G. (1965). *Selected Readings in Physics. Kinetic Theory* **1**, Pergamon Press.

Clausius, R. (1859). "On the mean lengths of the paths described by the separate molecules of gaseous bodies," *Phil. Mag.* **17**, 81–91.

Feather, N. (1959). *Mass, Length and Time*, Edinburgh University Press.

Provides a further account of the theory.

Project Physics, Volume 3. (1968). Holt, Rinehart and Winston.

Includes a very clear and detailed account of the kinetic theory of gases at the level of this Unit.

On materials

Gordon, J. E. (1968). *The New Science of Strong Materials*, Penguin.

A paperback providing an insight into materials science with little mathematical analysis. Chapters 2 and 3 will be found especially useful as a supplement to the coverage of the present Unit.

Holden, A. (1965). *The Nature of Solids*, Columbia University Press.

This gives useful background reading about the development of the model of a solid from simple to complex as more features have to be incorporated. It is at pains to show how increasing complexity of a model does not necessarily invalidate earlier, simple models.

Hurst, M. M. (1969). *Crystals*, Longmans.

An attractive account providing useful background reading.

Rogers, B. A. (1964). *The Nature of Metals*, Pergamon.

A useful reference concerning the nature and the behaviour of metals.

Tabor, D. (1969). *Gases, Liquids, and Solids*, Penguin.

Although going well beyond the standard of the present book, the early chapters will be found relevant to the work on the force and potential energy curves.

Vlack, L. van, (1966). *Elements of Materials Science*, Addison-Wesley.

Chapters 1 and 3 are a valuable reference for material properties and for the relationship between X-ray analysis and structure.

On X-ray diffraction methods

Elton, L. R. B., and Jackson, D. L. (1966). "X-ray Diffraction and the Bragg Law", *Am. J. Phys.*, **34** 1036.

This is the source for the approach to the Bragg law used in this Unit.

Diamond, J. (1963). "Microwave Optics," *School Science Review*, **45**, 23.

This paper describes a number of analogue experiments including the 3 cm wave analogue to X-ray diffraction.

James, R. W. (1957). *X-ray Crystallography*, Methuen.

This, and the reference to Quinn (1970), go well beyond the needs of the Unit but Chapter 2 will be found useful.

Quinn, T. F. J. (1970). *X-rays, Electrons, and Crystalline Materials*, Butterworths.

UNIT FIVE
Vibrations and Waves

17 Periodic Motion
18 The Motion of an Oscillator
19 Damped Oscillations
20 Forced Oscillations
21 Coupled Oscillators and the Transmission of Energy
22 Waves on Water
23 Wave Behaviour

Chapter 17
PERIODIC MOTION

17.1 Introduction
17.2 Describing periodic motion
17.3 Isochronous oscillation
17.4 Sinusoidal motion
17.5 Phase

17.1 INTRODUCTION

Before launching into a detailed study of vibrations we should be aware that all motion in the physical world can be divided into two broad categories: *periodic* motion and *non-periodic* motion. In periodic motion an object repeats its pattern of motion at regular intervals of time: it is rhythmic motion. The movement of a bird's wing, and a fly's wing, in flight, are examples of periodic motion. One wing-beat of a rook, lasts about half a second. The wing-beat of a bluebottle fly lasts about one hundredth of a second. The water level in the Pool of London changes in a periodic way, the time interval between one high tide and the next being just under thirteen hours. Periodic motion is very common in nature although often very complex in form: the precise motion of a rook's wing in flight is a subtle and beautiful thing, and far from easy to describe exactly. In the world of machinery periodic motion is commoner still, and vibrations can be an engineer's nightmare: the violent vibrations of an engine during trial runs may succeed in shattering it if the vibrations cannot be suppressed.

Rotational motion too is a form of periodic motion, but in this chapter we shall be concerned mainly with periodic motion back and forth along a straight line: how to describe this kind of motion, the forces which cause periodic motion, and the analysis of a variety of simple examples of periodic motion.

17.2 DESCRIBING PERIODIC MOTION

The graph in Fig. 17.1 shows the pulse of a healthy human being. It is in fact a record of the movement of the radial artery in the wrist, in response to the pumping action of the heart. It can be seen that the pattern of pressure variation repeats itself regularly. When a doctor feels a patient's pulse he is detecting, with his fingertips, the maxima of blood pressure which distend the artery in the patient's wrist in a periodic way. The unit of the pattern in Fig. 17.1, which is repeated regularly, is termed one *cycle* of the motion. The time duration of the cycle is called the

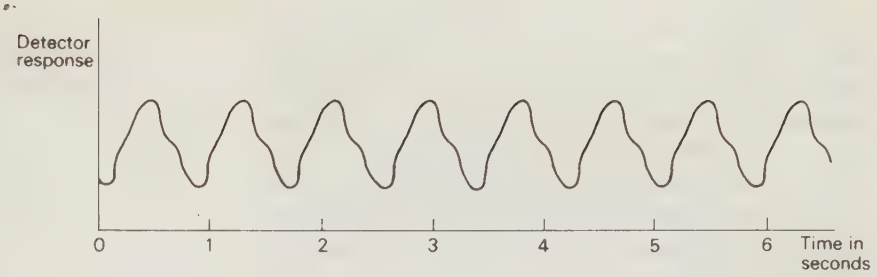


Fig. 17.1 Periodic motion of the human pulse.

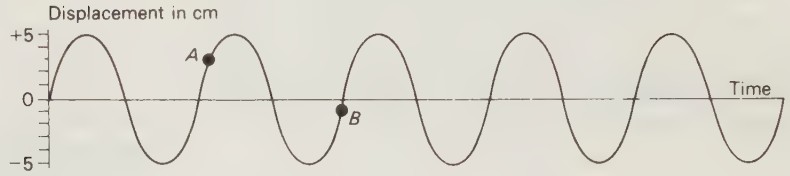


Fig. 17.2 Periodic motion of a mass attached to a spring.

periodic time, or more commonly, the *period* of the motion. Thus, a person with a pulse of 72 per minute has a period of blood-pressure variation of $\frac{60}{72} = \frac{5}{6}$ second.

Compare the pattern of Fig. 17.1 with that of Fig. 17.2. In Fig. 17.2 the graph shows a simpler example of periodic motion: the movement of a weight suspended on a spring and oscillating freely up and down, above and below its equilibrium position. Clearly, the pattern of the motion shown in Fig. 17.2 is simpler than that of Fig. 17.1. In fact, as will be discussed later, the motion illustrated in Fig. 17.2 is an example of the simplest kind of periodic motion.

We shall use the motion shown in Fig. 17.2 to define some of the terms used in describing periodic motion. The rhythmic movement of an object back and forth about an equilibrium position, or the rhythmic variation of a measurable physical quantity above and below an equilibrium value, is often termed *oscillation*. The moving object, and the varying physical quantity, can be said to perform oscillations or to execute oscillatory motion. The term *period of oscillation* has already been defined and we may see that, although the oscillatory motions shown in Figs. 17.1 and 17.2 could have the same period, they would still be quite different. We now have to discuss how the precise pattern of an oscillatory motion can be described.

Referring again to Fig. 17.2, we say that when the weight is at its equilibrium position its displacement is zero. The distance of the weight from its equilibrium position, at any given instant, is termed its *displacement*. If the weight is above the equilibrium position its displacement is reckoned as positive; if below equilibrium, it is reckoned as negative. Thus, at the instant represented by point A, the displacement is $+3$ cm. At point B the displacement is -1 cm. The maximum value of the displacement is termed the *amplitude* of the oscillation, and in this example we see that the amplitude is 5 cm.

The graph of Fig. 17.2 does not perfectly represent the free oscillation of a weight suspended by a spring because it does not show the gradual dying-down of the oscillation. We will assume for the present that this dying-down is negligible.

The *frequency* of an oscillatory motion is the number of cycles of oscillation performed per unit time, and the most commonly used time unit is the second. The basic unit used for measuring frequency is known as the *hertz* (Hz). A frequency of one hertz means simply one complete oscillatory cycle per second. Scientists and engineers often have to deal with very high frequencies and use the standard

prefixes (kilo, mega, etc.) for these (see Appendix 2). The vibrations associated with sound waves range from about 20 Hz to 20 kHz; the electrical oscillations associated with radio waves range from about 100 kHz to 100 MHz. When you get the opportunity, look at the tuning scale on a radio receiver, and notice the ranges covered. Not all such scales are marked in frequency units, however: they may be scales of wavelength.

The relationship between the period and the frequency of an oscillatory motion is simple. The symbol generally used for frequency is f, and the symbol for period is T, and one is the reciprocal of the other, that is,

$$T = \frac{1}{f}. \qquad (17.1)$$

For example, if you swing a small weight tied on a thread about 25 cm long like a pendulum you will find that the period is almost exactly half a second: thus the frequency is almost exactly 2 Hz. To take another example: in a certain radar transmitting antenna the frequency of the alternating current is 200 MHz. This means that the period of oscillation of the electrons is

$$\frac{1}{200 \times 10^6}\ \text{s} = 5 \times 10^{-9}\ \text{s}$$

$$= 5\ \text{ns (five nanoseconds).}$$

17.3 ISOCHRONOUS OSCILLATION

A pendulum is a simple form of oscillator. The pendulum has been used for about three centuries as the basis of reliable time-keeping instruments. It is probably not surprising that a pendulum, swinging back and forth with a constantly maintained amplitude, should keep regular time. But it is perhaps rather surprising to observe that, if a pendulum is set swinging freely, as it dies down and the amplitude decreases, the rhythm does not alter appreciably. This is to say that the period of oscillation is very nearly independent of the amplitude of oscillation. Table 17.1 shows how very nearly constant the period of a simple pendulum is for small amplitudes.

Table 17.1

Total angle of swing	Period (s)
10°	1.0005
20°	1.0019
30°	1.0043
40°	1.0076
50°	1.0117
60°	1.0168
70°	1.0227
80°	1.0293
90°	1.0387

This property of a pendulum swinging with small amplitudes, namely, that its period is independent of its amplitude of swing, is called *isochronism*. It was first observed, so the tradition has it, by Galileo Galilei in the cathedral of Pisa. He was attending mass and, during a tedious sermon, timed the swing of a lantern suspended by a long chain from the roof, by counting how many of his own pulse-beats corresponded to each swing of the lantern. He discovered that the time of swing was the same for large and small amplitudes alike.

A pendulum is not a perfectly isochronous oscillator. One might wonder if in fact there is such a thing as a precisely isochronous oscillator. As will be made clear later, an oscillator consisting of a mass attached to elastic springs is very nearly isochronous. The distinction between isochronous and non-isochronous oscillations is very nicely demonstrated by a simple piece of apparatus (Fig. 17.3) consisting of pieces of bent metal curtain-track with a steel ball rolling in each. One is a V-shaped track with a curved apex. Another is in the form of a circular arc. The third is parabolic. A ball is released near the top of one side of the V-shaped track and allowed to roll back and forth. The amplitude of the motion decreases rapidly as a result of energy-loss through friction. The period of oscillation is long at first; then, as the amplitude decreases, it gets shorter until finally the ball is rattling back and forth rapidly in the apex of the track.

By contrast, the motion of the ball in the circular track is quite different. As the amplitude decreases

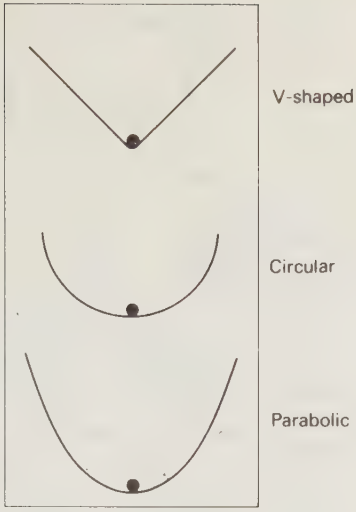


Fig. 17.3

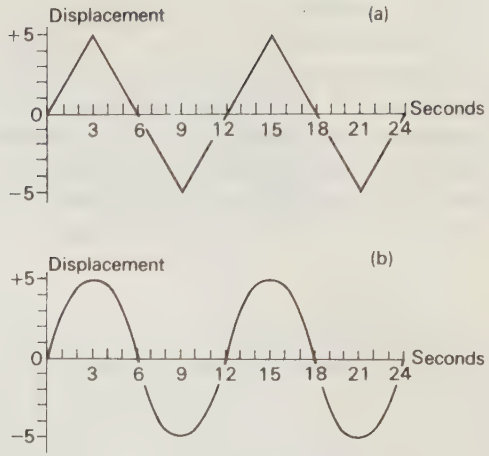


Fig. 17.4

there is no noticeable change in the rhythm of the movement, and the period appears to remain constant right up to the stage at which the motion becomes imperceptible. Thus the ball rolling in a circular arc, like a pendulum bob moving in a circular arc, appears to be a nearly perfect isochronous oscillator. The ball rolling in the parabolic track is found not to be isochronous, although as the amplitude becomes smaller, the period, initially long, becomes less, and seems to approach a limiting constant value.

The perfectly isochronous oscillator is an ideal to which a number of simple oscillatory systems approximate very well in practice. Isochronous oscillators are, for our purposes, the most important category of oscillators because their mathematical description and analysis is simple.

17.4 SINUSOIDAL MOTION

We have already gone some way towards the description of oscillatory motion, defining the terms *cycle*, *period*, *frequency*, *displacement*, *amplitude*, and *isochronous*. But as yet we have not concerned ourself with the pattern of oscillatory motions in particular. Consider the two oscillatory motions represented by the displacement-time graphs of Fig. 17.4. The two oscillatory motions have the same amplitude and frequency, but there is a difference in the way in which the displacement varies with time. Whereas motion like Fig. 17.4b is easy to produce in practice (a weight suspended on a spring), the motion of 17.4a would be impossible in any mechanical system comprising a mass and agencies for producing forces to act upon the mass. (Why is such motion impossible?) But for the time being we are not concerned with the nature of the forces causing the motion, only with how the motion can be described.

We will use the symbol x for displacement and t for time. In the motion of (a) in Fig. 17.4 the displacement is not a continuous function of time, and the motion cannot be described by one single, simple mathematical expression. We could describe it thus:

$$x = \tfrac{5}{3}t \qquad \text{(from } t = 0 \text{ to } t = 3\text{)},$$

$$x = 10 - \tfrac{5}{3}t \qquad \text{(from } t = 3 \text{ to } t = 9\text{)},$$

$$x = -20 + \tfrac{5}{3}t \qquad \text{(from } t = 9 \text{ to } t = 15\text{)},$$

and so on. But the motion of (b) can be described simply by the expression

$$x = 5 \sin \frac{\pi}{6} t.$$

Motion (b) is an example of *sinusoidal* motion, whereas (a) is one type of *non-sinusoidal* motion. It is the sinusoidal motion which we now need to examine more closely.

The expression $x = a_0 \sin \omega t$ represents the displacement of a particle moving with sinusoidal motion of amplitude a_0. ω is a constant which is related to the frequency (and hence the period) of the motion as will be explained shortly. Table 17.2 and Fig. 17.5 illustrate successive stages of the motion.

Table 17.2

Particle at	ωt	$\sin \omega t$	x
O	0	0	0
P	$\pi/2$	+1	$+a_0$
Q	π	0	0
R	$3\pi/2$	−1	$-a_0$
S	2π	0	0
T	$5\pi/2$	+1	$+a_0$

When the particle reaches D it has completed one cycle of the oscillation, so the time elapsed (OD on the time scale) is the period, T, and

$$\omega T = 2\pi \tag{17.3}$$

hence

$$T = \frac{2\pi}{\omega} \tag{17.4}$$

and the frequency

$$f = \frac{1}{T} = \frac{\omega}{2\pi} \tag{17.5}$$

alternatively we can write

$$\omega = 2\pi f. \tag{17.6}$$

Thus, instead of writing $x = a_0 \sin \omega t$ we can write $x = a_0 \sin 2\pi f t$. The name *angular frequency* is generally given to the quantity represented by ω, and it will have greater significance when we come to discuss the relationship between sinusoidal motion and motion in a circular orbit. For the time being we need think of ω as being simply a shorthand way of writing $2\pi f$.

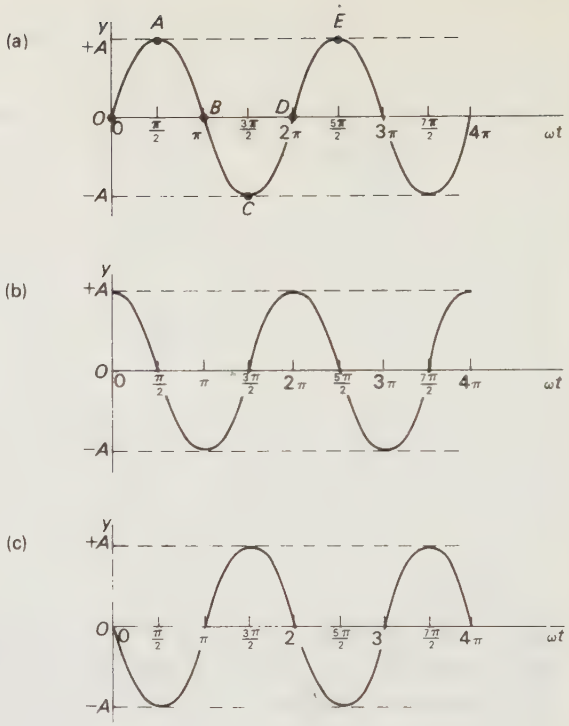


Fig. 17.5

In the expression $x = a_0 \sin \omega t$, as has been already explained, x represents the displacement of a particle and t represents the time: but we must bear in mind that, when we say "t represents the time", there must be a starting point for this time—an instant at which we start our imaginary clock. This instant we can call *time zero*. The expression $x = a_0 \sin \omega t$ then implies that at time zero the particle had zero displacement, and that the particle is just about to be displaced in the *positive* direction. The expression is plotted as a graph in Fig. 17.5(a). Now, we could just as well have described the same oscillatory motion in a slightly different way, choosing to start our clock at the instant when the particle has maximum positive displacement: then the expression would be $x = a_0 \cos \omega t$. This expression is plotted in Fig. 17.5b. The pattern of the motion is precisely the same in each case: the only difference is in the

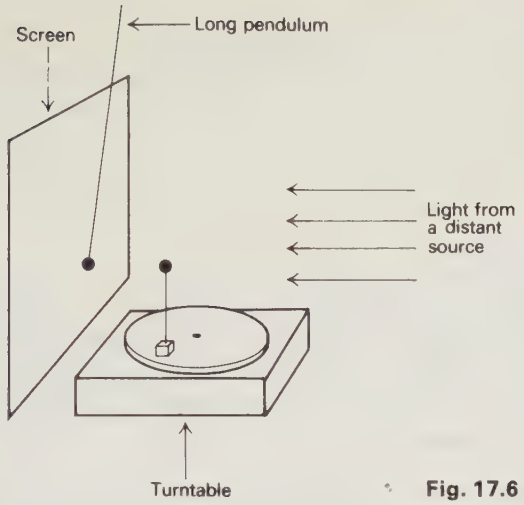


Fig. 17.6

stage, during the oscillation, at which we decide to start our clock. We could describe the same motion in a third way which is different again. Figure 17.5c shows a graph representing precisely the same pattern of motion. You should be able to write down the expression which is represented by this graph.

Let us summarize at this point. The same pattern of sinusoidal motion, as illustrated in Fig. 17.5, can be represented in different ways. The three ways we have chosen can be written as algebraic expressions thus:

$$x = a_0 \sin \omega t, \qquad (17.7a)$$

$$x = a_0 \cos \omega t, \qquad (17.7b)$$

$$x = -a_0 \sin \omega t. \qquad (17.7c)$$

Now these three expressions can also be written (taken in the same order) thus:

$$x = a_0 \sin \omega t, \qquad (17.8a)$$

$$x = a_0 \sin \left(\omega t + \frac{\pi}{2}\right), \qquad (17.8b)$$

$$x = a_0 \sin (\omega t + \pi). \qquad (17.8c)$$

You should check that these are correct alternative ways of writing the three original expressions, either by using your knowledge of trigonometrical functions, or by drawing up a table of values as in Table 17.2. Observe that the three expressions all contain the "$a_0 \sin \omega t$" part, and that the information about the choice of the instant at which we start the imaginary clock is contained in an extra term which has the value $\pi/2$ in Eq. (17.8b) and π in Eq. (17.8c).

17.5 PHASE

There is a very close connection between sinusoidal motion, back and forth in a straight line, and the motion of an object at constant speed around the circumference of a circle. With the apparatus of Fig. 17.6 we can show this correspondence, and use it also to introduce the concept of *phase*. We set up a slow-running turntable close to a vertical screen. On the turntable we place a thin vertical rod mounted upon a heavy base and supporting a small ball at the top end. A small, intense light source is placed several metres away, on the same horizontal level as the ball, so that it casts a shadow of the ball on the screen. The light from the distance source is very nearly parallel and so the shadow of the ball should be sharp and negligibly larger than the ball itself. When the turntable is set moving at a steady speed the shadow is seen to move back and forth in a straight line on the screen. Is the motion sinusoidal? Well, it is easy, half-closing one's eyes, to imagine that the shadow is not that of a ball going in a circle, but of the bob of a long pendulum swinging to and fro. Later in Section 18.7, it is proved that the motion of a simple pendulum approximates very closely to sinusoidal motion provided that the angular amplitude of swing is small. We can mount a long pendulum—a small heavy ball suspended by a thread—to swing near the screen in a plane parallel to it, and by adjusting its length arrange that the shadow of the pendulum bob moves with the same period as the period of revolution of the ball on the turntable. This needs some patience to achieve exactly, but makes an impressive demonstration. The two shadows will be seen to move with almost exactly the same motion.

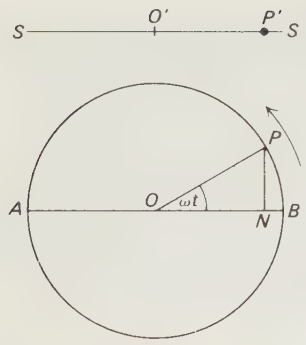


Fig. 17.7

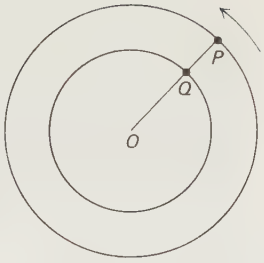


Fig. 17.8

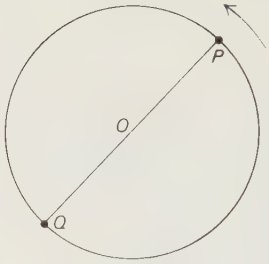


Fig. 17.9

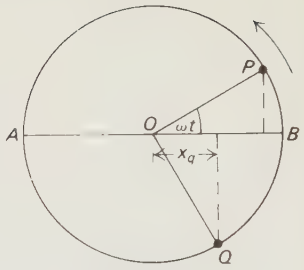


Fig. 17.10

We can explain the correspondence demonstrated above by a simple geometrical argument. In Fig. 17.7 imagine that the small ball is travelling anticlockwise around the circumference of the circle. A clock is started when it is at B. At time t the ball is at point P. The ball is travelling around the circle with constant speed. If its angular velocity is ω then the angle $P\hat{O}N$ is ωt. A shadow of the ball is being cast on a screen SS, the line SS being parallel to AB, by a parallel beam of light in a direction perpendicular to the screen. Thus the shadow of the ball is at P'. As the ball travels around the circle the shadow P' travels back and forth in a straight line on the screen. Clearly, the displacement of the shadow-ball, measured from the central point (O') of this motion, is equal to the distance ON at every instant. It is easier now for us to fix our attention upon the motion of N rather than on P'. The amplitude a_0 of the motion of N is equal to the radius of the circle, OA. Thus the displacement of N, measured from O, (x) is given by

$$x = a_0 \cos \omega t$$

and since a_0 and ω are constants, this means that x is varying sinusoidally. Thus the motion of N, and hence of the shadow P', are truly sinusoidal, as we expected, and the displacement-time graph would be that of Fig. 17.5(b).

Now return to the apparatus of Fig. 17.6. On the turntable we place another ball on a stick as shown in the plan view of Fig. 17.8 at Q. With the turntable rotating the two shadows moving back and forth. For convenience we will use the symbol P' for the shadow of P, and Q' for the shadow of Q. When P' is at its maximum leftwards displacement so also is Q'. Likewise, P' and Q' both reach their maximum displacement to the right at the same instant, and have zero displacement at the same instant. The amplitudes are different: that of Q' is smaller than that of P'. But the two motions are *in phase*, and therefore we could represent them by the equations

$$x_P = a_P \cos \omega t,$$

$$x_Q = a_Q \cos \omega t.$$

Now arrange P and Q as shown in Fig. 17.9. As they go around the circle (this time the amplitudes are equal) the displacement of P' will be at every instant equal in magnitude but opposite in sign to that of Q': the two motions are *in antiphase*, or of *opposite phase*.

A case of special interest is when the motions of P' and Q' are *in quadrature*. This is achieved by arranging P and Q as in Fig. 17.10. For this demonstration it is best to have the balls P and Q on different horizontal levels so that their shadows are easily identifiable on the screen. As P and Q go anticlockwise around the circle the shadow P' leads the shadow Q': and Q' gives the impression of lagging behind like a lazy dog on an elastic lead. Observing carefully, one notices that when P' has maximum displacement Q' has zero displacement, and vice

versa. Now, if the graph of Fig. 17.5(b) represents the motion of P', which of the other two graphs in Fig. 17.5 represents the motion of Q'? It is 17.5(a): you should check this carefully. Thus the motions of P' and Q' in this situation can be described by the equations

$$x_P = a_0 \cos \omega t,$$

$$x_Q = a_0 \sin \omega t.$$

(We write the amplitude, a_0, here because we arranged the amplitudes to be equal in the demonstration.) Readers having a good knowledge of trigonometrical functions will know that $\sin \theta = \cos (\theta - (\pi/2))$, and so the two equations could be written

$$x_P = a_0 \cos \omega t,$$

$$x_Q = a_0 \cos \left(\omega t - \frac{\pi}{2}\right).$$

But this can easily be seen by studying Fig. 17.10 in which the displacement x_Q equals $a_0 \cos Q\hat{O}B$, but $Q\hat{O}B$ is of course equal to $\omega t - (\pi/2)$. In the two equations above, the term $-(\pi/2)$ is the *phase difference* between the two motions. We can say that Q' has a *phase lag* of $\pi/2$ radians behind P', or that P' *leads* Q' by $\pi/2$ radians.

If we wanted Q' to lead P' with a phase-difference of $\pi/2$ radians, then, using the same expression as before to represent the motion of P', we would write the expression for Q' thus:

$$x_Q = a_0 \cos \left(\omega t + \frac{\pi}{2}\right).$$

In general, therefore, we can write the expressions

$$x_P = a_0 \cos \omega t,$$

$$x_Q = a_0 \cos (\omega t + \phi),$$

where ϕ is the phase difference between the two motions, and can be expressed in radians (or in degrees).

The concept of phase is vital in describing the relative motions of two or more sinusoidal oscillators. But it also has relevance when describing the motion of a single oscillator, as is shown by the example which follows.

Problem A particle is performing sinusoidal oscillation about a fixed point with amplitude 5 cm and frequency 8 Hz. A clock is started when the displacement of the particle is +2.5 cm. Find an expression for its displacement as a function of the time, t, registered by the clock.

Solution We can straight away write the expression in the form

$$x = a_0 \cos (\omega t + \phi)$$

and we have to find the appropriate values for a_0, ω, and ϕ. Now $a_0 = 0.05$ m, and $\omega = 2\pi f = 2\pi \times 8 = 16\pi$ radians per second. To find ϕ we reason as follows.

We require that when $t = 0, x = 0.025$. Therefore

$$0.025 = 0.05 \cos \phi$$

hence

$$\cos \phi = +0.5$$

and so

$$\phi = 60° \quad \text{or} \quad 300°.$$

But how does it come about that ϕ can have two possible values? You should be able to reason this out for yourself, perhaps with the help of sketch graphs, and bearing in mind that the displacement of the particle has the value +2.5 cm twice during every cycle.

The most general expression for sinusoidal motion is

$$x = a_0 \cos (\omega t + \phi), \qquad (17.9)$$

or, using the sine function, $x = a_0 \sin (\omega t + \phi)$. You will meet both the cosine and sine forms, in different contexts, and should be ready to handle both. From this point onwards in this chapter we shall keep to the cosine form, and, for simplicity's sake, usually assume when we are dealing with a single sinusoidal oscillator that our clock (usually an imaginary clock) has been started when the oscillator had its maximum positive displacement, so that the expression we use is $x = a_0 \cos \omega t$.

Chapter 18
THE MOTION OF AN OSCILLATOR

18.1 Velocity and acceleration of a sinusoidal oscillator
18.2 Computing the motion of a simple oscillator
18.3 The equation of motion of a simple harmonic oscillator
18.4 The simple pendulum
18.5 The energy of a simple harmonic oscillator

18.1 VELOCITY AND ACCELERATION OF A SINUSOIDAL OSCILLATOR

Just by looking at a pendulum swinging, or a weight bobbing up and down on the end of a spring, we can deduce certain facts about the velocity of a sinusoidal oscillator. When the oscillating object has maximum displacement, either positive or negative, its velocity is zero. The velocity is maximum when the object is passing through its equilibrium position, in either direction. It must be remembered that velocity is a vector quantity: it has direction as well as magnitude. If we think of a mass on a spring, when above its equilibrium position, as having positive displacement then we must also reckon its velocity as positive when it is moving upwards, and negative when downwards. Let us see how the velocity-time graph of a sinusoidal oscillator is related to the displacement-time graph. We can derive the velocity-time graph from the displacement-time graph using the knowledge that the velocity at any instant is equal to the gradient of the displacement-time curve at that instant. The upper curve in Fig. 18.1 shows the displacement, the middle curve the velocity, and the bottom curve the acceleration.

In Fig. 18.1 the time scale has been divided into arbitrary time units. The period of oscillation is twelve of these time units. Let us study the main features of the displacement and velocity curves. At times 0, 6, 12, 18, etc. the displacement has maximum positive or negative value: at these instants the gradient of the displacement-time graph is zero, and accordingly the velocity is zero. The velocity has its greatest value when the gradient of the displacement-time graph is greatest. At time = 3 the displacement curve has maximum negative gradient, and this is the instant when the velocity has its maximum negative value: similarly at time = 15, 27, etc. The maximum positive velocity occurs at time = 9, 21, etc. If the expression for the displacement is

$$x = a_0 \cos \omega t \qquad (18.1)$$

the expression for the velocity can be found by differentiating the above expression with respect to t:

$$\text{velocity } v = \frac{dx}{dt} = -a_0\omega \sin \omega t \qquad (18.2)$$

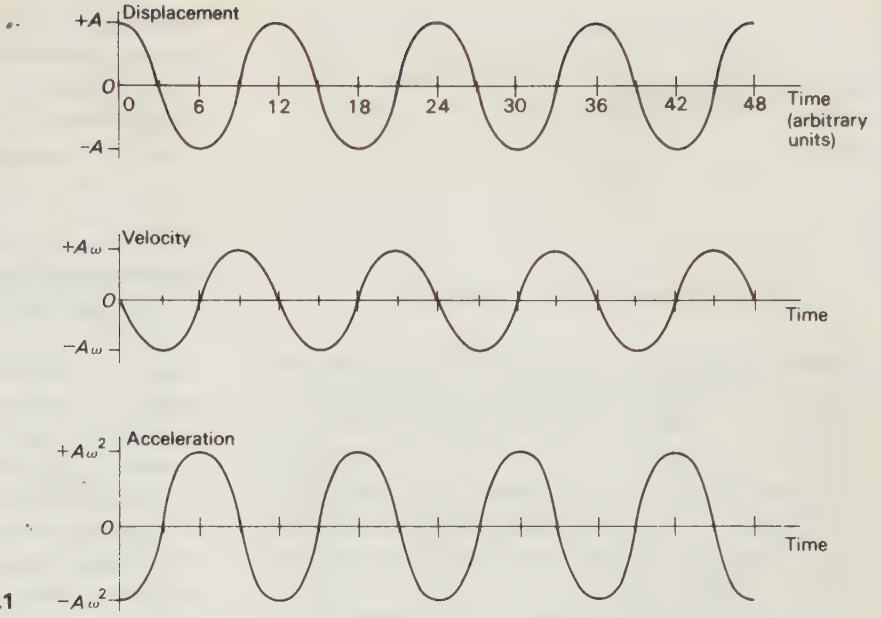


Fig. 18.1

and the middle curve of Fig. 18.1 is a graph of this expression. The maximum value of the velocity is $a_0\omega$, and since $\omega = 2\pi f$, this can also be written $2\pi f a_0$.

The acceleration-time graph can be derived from the velocity-time graph. Can you predict the shape of this graph? Pause and consider this for a while: just by watching a swinging pendulum it is very difficult to see how the acceleration is changing, although the velocity changes are easy to observe.

At any instant the acceleration is equal to the gradient of the velocity curve. Again, an expression for the acceleration can be obtained by differentiation, thus:

$$\text{Acceleration} = \frac{dv}{dt} = \frac{d}{dt}(-a_0\omega \sin \omega t)$$

$$= -a_0\omega^2 \cos \omega t \qquad (18.3)$$

and the greatest value of the acceleration is $a_0\omega^2$ which can also be written $4\pi^2 f^2 a_0$.

The expressions for velocity and acceleration can be derived, without calculus, from the circle diagram used earlier in explaining the concept of phase. In Fig. 18.2, once again, we have a ball travelling with uniform angular velocity ω in an anticlockwise direction around a circle whose radius is a_0. At time t the ball is at position P. As was explained previously, the point N moves with sinusoidal motion about O while P moves around the circle. The displacement of N was shown to be given by

$$x = a_0 \cos \omega t.$$

The linear velocity of the ball at P is $a_0\omega$, and the velocity of N along the line AOB must be the component of P's velocity parallel to AOB. This component is $-a_0\omega \sin \omega t$ (the negative sign appears because this component of velocity is in the negative x-direction). This result agrees with that derived using calculus above.

To find an expression for the acceleration of N we need to know, from the theory of uniform motion in a circle, that the acceleration of a particle travelling in a circular orbit with linear speed v and radius a_0 is v^2/a_0. (Unit 2). Since $v = a_0\omega$, this can be written as $\omega^2 a_0$. Also, this acceleration is directed

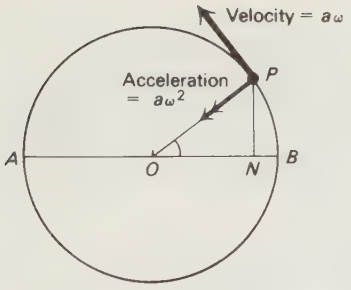


Fig. 18.2

towards the centre of the circle: it is therefore shown in Fig. 18.2 by the double arrowhead directed along PO. The acceleration of the point N is the component of P's acceleration parallel to the line AOB, and since this component of acceleration is in the negative x-direction, it is given by the expression $-a_0\omega^2 \cos \omega t$. This also agrees with the expression derived using calculus.

Comparing the expressions for displacement and acceleration we see that they both show the same pattern of time-variation because the time-dependent term, $\cos \omega t$, is the same for both. Thus the relationship between displacement and acceleration can be written

$$\text{acceleration} = -\omega^2 x. \qquad (18.4)$$

This reveals a very important property of sinusoidal motion: the acceleration of the oscillating object is directly proportional to its displacement, and is always directed in the reverse direction to the displacement, as the minus sign indicates. This means that the object is always accelerating towards its equilibrium position. The constant of proportionality, $-\omega^2$, can also be written $-4\pi^2 f^2$.

Now there arises a question of logic. We have shown that, if an object is moving with sinusoidal motion expressed either by the formula $x = a_0 \sin \omega t$ or by $x = a_0 \cos \omega t$, then it necessarily follows that the object's acceleration is directly proportional to its displacement. But is it true that, if an object moves in such a way that its acceleration is directly proportional to its displacement, then it *necessarily* follows that its motion will be of the sinusoidal kind? It will be shown later that the motion must be sinusoidal provided that the constant of proportionality relating acceleration and displacement is negative. Before moving on to the next section, you should consider carefully what kind of motion you would expect if an object moved in such a way that its acceleration was related to its displacement thus:

$$\text{acceleration} = Cx,$$

where C is a constant having positive value. For such an object, the greater its positive displacement from the equilibrium position, the greater its acceleration in the positive direction (that is, away from equilibrium). Thus its velocity in the positive direction, away from equilibrium, would get larger and larger the further away the object went. Clearly this means a "runaway" process, and we do not expect the object to turn around and come back to equilibrium ever!

The reasoning above should suggest that it is the *negative* value of the constant of proportionality relating acceleration and displacement that makes for periodic motion. This point will become clearer when we actually compute the motion of a simple mechanical oscillator.

Problem 1 Exercises on the description of simple harmonic motion.

The displacement of a certain particle moving with simple harmonic motion is described by the expression

$$x = 5 \cos 16\pi t \text{ cm}.$$

a) What is the amplitude of the motion?
b) What is the frequency of the motion?
c) If the particle starts its motion at time $t = 0$, how long after this does it first pass through its equilibrium position?
d) What is the velocity of the particle as it passes through its equilibrium position?
e) What is the displacement of the particle $\frac{1}{48}$ s after the start of its motion?
f) What is the maximum acceleration of the particle during its motion?

Solution a) In the expression $x = a_0 \cos \omega t$, a_0 is the amplitude of the motion: so here the amplitude is 5 cm.

b) The angular frequency is ω which equals $2\pi f$ where f is the frequency: so here the frequency is 8 Hz.

c) The period of the motion is the reciprocal of the frequency, so here the period is $\frac{1}{8}$ s. At time $t = 0$ the particle has maximum displacement, and it passes through the equilibrium point one quarter-cycle after this, that is, $\frac{1}{4} \times \frac{1}{8}$ s $= \frac{1}{32}$ s after.

d) The expression for the velocity of a particle whose displacement is described by $x = a_0 \cos \omega t$ is

$$\frac{dx}{dt} = \dot{x} = -a_0\omega \sin \omega t.$$

When the particle passes through the equilibrium position its velocity is maximum. The maximum value of $\dot{x}$ is $a_0\omega$, so in our example the velocity is $5 \times 16\pi$ cm s^{-1}, which equals 80π cm s^{-1}.

e) At a time $\frac{1}{48}$ s after the start of the motion, the displacement is given by

$$x = 5 \cos (16\pi \times \tfrac{1}{48})$$

$$= 5 \cos \frac{\pi}{3}$$

$$= 5 \times \tfrac{1}{2} = 2.5 \text{ cm}.$$

f) The general expression for acceleration is

$$\frac{d^2x}{dt^2} = \ddot{x} = -a_0\omega^2 \cos \omega t$$

and the maximum value of this is $a_0\omega^2$. In our example this is therefore $5 \times (16\pi)^2 = 1280\pi^2$ cm s^{-2}.

Problem 2 The piston of an automobile engine running at a steady speed moves with approximately simple harmonic motion. In a certain engine the stroke is 10 cm (this is the double-amplitude of the motion of the piston). The crankshaft of the engine is making 3600 revolutions per minute.

a) Write an expression for the displacement of the piston from its mid-stroke position as a function of time, t.

b) Calculate the maximum velocity of the piston during the stroke.

c) Calculate the maximum acceleration of the piston during the stroke.

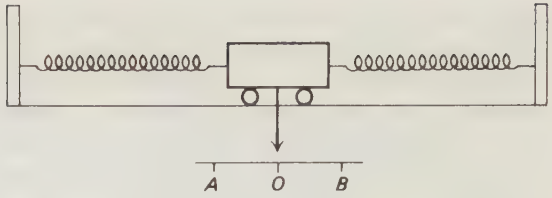


Fig. 18.3

d) The mass of the piston is 500 g. Calculate the maximum accelerating force acting upon the piston during the stroke.

Solution

a) The amplitude of the motion is 5 cm. The frequency is $\frac{3600}{60}$ cycles per second, that is, 60 Hz. Thus $\omega = 2\pi \times 60 = 120\pi$ s^{-1}. Thus the expression for displacement is $x = 5 \cos 120\pi t$.

b) The maximum velocity is $a_0\omega = 5 \times 120\pi = 600\pi = 1890$ cm s^{-1}.

c) The maximum acceleration is $a_0\omega^2 = 5 \times (120\pi)^2 = 710\,000$ cm s^{-2} $= 7100$ m s^{-2}, approximately $720 \times g$!

d) Using $F = ma$, the force is

$$0.5 \text{ kg} \times 7100 \text{ m s}^{-2} = 3550 \text{ N}.$$

18.2 COMPUTING THE MOTION OF A SIMPLE OSCILLATOR

A simple mechanical oscillator, ideal for an introductory study of oscillations is shown in Fig. 18.3. A trolley of the kind used in dynamics experiments is tethered by two long helical springs lying along the same axis. At equilibrium both springs are stretched to about twice their original unstretched length so that, when the trolley is pulled to one side and let go, it will oscillate back and forth without the springs ever slackening. The surface upon which the trolley runs needs to be smooth: a sheet of glass, tinplate, formica, or any similar hard, smooth material is suitable. This apparatus may be used for investigating how the period of oscillation depends on the mass of the object oscillating and the stiffness of the springs. But for

the present we are concerned simply with the *kind* of motion performed by this oscillator and the forces causing this motion. We will suppose that, for the time being, we can neglect any forces of friction. If there were no friction, the trolley, once displaced and let go, would continue to oscillate with constant amplitude. In practice the amplitude decays, and a typical trolley will perform 20 or 30 cycles before finally coming to rest.

At equilibrium the pointer attached to the trolley is opposite 0. Suppose now that it is displaced to *B*. There is now a resultant force on the trolley to the left, and if the trolley is being acted upon only by the forces of the springs, it must therefore (by Newton's second law of motion) have an acceleration to the left. Suppose that the trolley moves under the action of this force. Its velocity, to the left, will increase. When the trolley is at a point nearer to 0 than *B* the force on it is less, therefore the acceleration is less: but this still means that its velocity to the left is increasing although it is increasing at a slower rate than before. As the trolley passes through the equilibrium position, 0, the forces on it are balanced so its acceleration is zero. After passing 0 the force on the trolley is now to the right, and this force decelerates the trolley. The greater the trolley's displacement to the left of 0, the greater is this deceleration. At *A*, which is the same distance from 0 as *B*, the velocity of the trolley has been reduced to zero, and thereafter it accelerates back towards *B*. So we see that, at every point during the motion, the trolley is acted upon by a resultant force which is directed back towards the equilibrium position. Hence the acceleration (which, by Newton's second law, is directly proportional to the force) is always in the reverse direction to the displacement.

We will now compute the motion of a typical trolley-on-springs oscillator for one quarter-cycle of its motion: from *B* to 0 in Fig. 18.3. We will suppose that the trolley has a mass of 3 kg and that the springs, together, provide a force of 12 newtons per metre displacement of the trolley, either side of equilibrium.

Suppose that the trolley is displaced a distance x (measured in metres). Then the force on the trolley is $-12x$ (note the minus sign which indicates that the force acts in a direction opposite to the displacement). Hence the acceleration, which by Newton's second law is got by dividing the force by the mass, is $-4x$. Here we have, then, an acceleration which is directly proportional to the displacement and in a direction opposite to the displacement.

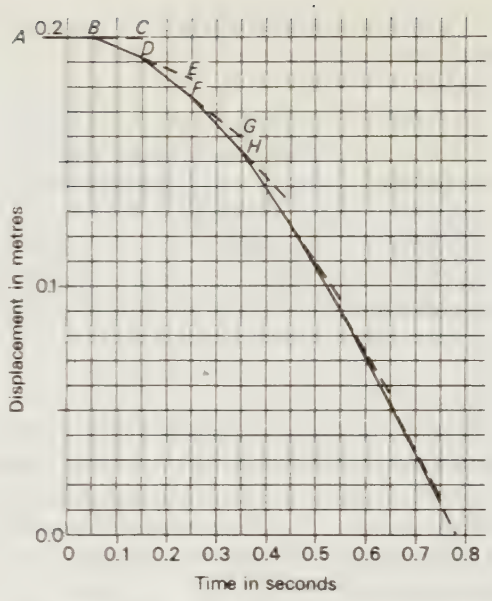


Fig. 18.4 Computing the graph of displacement against time for the simple harmonic oscillator.

The motion begins, we will assume, with the trolley pulled out to a distance 0.200 m and released from rest. We shall consider what happens in successive time intervals of 0.1 s from the start. Our computation will be only an approximation to the truth. How good an approximation it is we shall be able to see afterwards. In this method of computation we build up the displacement-time graph for the oscillator step by step. Figure 18.4 shows this graph completed. Note that the graph is made up of straight line segments. Each segment therefore represents motion with uniform velocity, and there is an abrupt change in velocity between segments. This of course is not the truth, only an approximation to it. Let us now study the principles of the computation. It may help to refer to an imaginary ticker-tape record of the

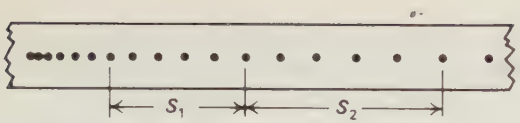


Fig. 18.5

motion, during the first quarter-cycle, as shown in Fig. 18.5.

Suppose the distances travelled in successive time-intervals Δt are s_1 and s_2. Then the average velocities for these two time intervals are $s_1/\Delta t$ and $s_2/\Delta t$. The acceleration, a, or rather the average acceleration, is therefore given by

$$a = \frac{\frac{s_2}{\Delta t} - \frac{s_1}{\Delta t}}{\Delta t} = \frac{s_2 - s_1}{\Delta t^2}.$$

If the object had been moving with uniform velocity it would have travelled equal distances in successive time intervals Δt. Because it is accelerating it has travelled an extra distance Δs given by

$$\Delta s = s_2 - s_1 = a\,\Delta t^2.$$

In building up the graph we calculate, and make use of, the extra distances travelled by the object, during successive intervals of time, as a result of its acceleration.

The acceleration is not constant, of course: it is proportional to the displacement and in our example is given by $a = -4x$. Our value for Δt is 0.1 s, so that the expression for the extra distance travelled during each time interval becomes

$$\Delta S = a(0.1)^2 = \frac{-4x}{100}.$$

We can now build up the displacement-time graph for our oscillator. During the first 0.1 s time interval the oscillator's velocity has increased from zero to some value in the negative x-direction: and we know that in fact this increase has taken place smoothly, but we assume, for our approximation, that there has been an abrupt increase in velocity halfway through the time interval. We know that the gradient of the displacement-time graph represents the velocity: this means that the segment AB (gradient = 0 and hence velocity = 0) represents the velocity at time $t = 0$. Segment BD represents the velocity at time $t = 0.1$ s. If the velocity of the oscillator had remained zero after the start, then it would be at point C on the graph after the second time interval: but its velocity has in fact changed and instead of travelling zero distance it has travelled an extra distance which we calculate using our formula derived above

$$\Delta S = \frac{-4x}{100}.$$

Here $x = 0.200$ m, so $\Delta S = -0.008$ m. Therefore the distance CD on the graph must be 0.008 m: this enables us to plot point D and then draw the line BD. Going on from point D, if the oscillator travelled from there with constant velocity it would arrive at point E: but since it is in fact accelerating it will arrive at point F and to locate this point on the graph we must compute the extra distance travelled in the 0.1 s time-interval around time 0.2 s. This extra distance is $-4/100 \times 0.192 = 0.008$ m (correct to the third decimal place). Thus the extra distance, EF on the graph, is 0.008 m so that point F can be plotted and the line DF drawn in.

You should continue the process and produce the graph until it cuts the time-axis. If this is done carefully it will be found that the line crosses the time axis very close to time 0.78 s. If the process were continued further it would be found that the line, having gone below the time-axis and curved upwards, cuts the time-axis again at 3×0.78, then again at 5×0.78, 7×0.78 s, and so forth. The period of oscillation is thus $4 \times 0.78 = 3.12$ s very nearly. The curve we have obtained by this method looks very much like the first part of a cosine graph. We can test this numerically: then we shall deal with the theory.

The equation to our graph of Fig. 18.4 can be written, we will assume, thus: $x = 0.200 \cos \theta$ since $a_0 = 0.2$ m. Now $\cos \theta = 0$ when $\theta = \pi/2$, $3(\pi/2)$, $5(\pi/2)$, etc. The first zero of our curve occurs when $t = 0.78$ which is very nearly $\pi/4$. So, in order for

our graph to be $\cos\theta$ plotted against t, θ must equal $2t$, so that when $t = \pi/4$, $\theta = \pi/2$ as we require. Let us try how good a fit the function

$$x = 0.200 \cos 2t$$

is to our curve. Take a trial value, say, time = 0.40 s. When $t = 0.40$, $\cos 2t = \cos 0.8$ (that is, 0.8 radians which equals 45.8°). Therefore $x = 0.200 \cos 45.8° = 0.200 \times 0.698 = 0.140$. This agrees very closely with the graph.

18.3 THE EQUATION OF MOTION OF A SIMPLE HARMONIC OSCILLATOR

We now need to know the exact form of the mathematical function which relates displacement to time for the motion of the mechanical oscillator which we have been discussing. We know that the oscillator is acted upon by a force which is directly proportional to its displacement, the force acting always in a direction opposite to the displacement. This fact can be tested by experiment. Hence, applying Newton's second law of motion, we know that the acceleration is always proportional to the displacement and is in a direction opposite to the displacement. We have seen that the computed graph for our example of an oscillator looks very much like a cosine function. Now we shall link these facts together.

Let the force per unit displacement provided by the springs be k. When the displacement is x, therefore, the force on the oscillating object is $-kx$ (we write a minus sign for reasons discussed earlier). Thus the acceleration, a, is given by

$$a = -\frac{k}{m}x,$$

where m is the mass of the object. In calculus notation,

$$\frac{d^2x}{dt^2} = -\frac{k}{m}x \qquad (18.5)$$

thus, whatever function x is of t, when it is differentiated twice with respect to t, it must equal $-(k/m)x$. Now we know (Eq. 18.4) that the function $x = a_0 \cos \omega t$, when differentiated twice with respect to t, equals $-\omega^2 x$. So let us try this function as a possible solution of the differential equation above.

$$\frac{dx}{dt} = -a_0\omega \sin \omega t$$

and

$$\frac{d^2x}{dt^2} = -a_0\omega^2 \cos \omega t = -\omega^2 x \qquad (18.6)$$

Comparing Eqs. (18.5) and (18.6) we see that $x = a_0 \cos \omega t$ is a solution to the differential equation if

$$\omega^2 = \frac{k}{m}. \qquad (18.7)$$

(In our numerical example discussed above the value of k/m was 4, so that $\omega = 2$ and the function should be $x = a_0 \cos 2t$. We found that this fitted very closely, with $a_0 = 0.200$.) Earlier, in Section 17.4, we showed that for sinusoidal motion the constant ω was related to the frequency and the period of oscillation. In particular, the period is given by

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

from which it follows that, since $\omega^2 = k/m$,

$$T = 2\pi\sqrt{\frac{m}{k}}. \qquad (18.8)$$

You should substitute the numerical values of the previous example in this formula to see whether or not it agrees with the results of the step-by-step computation.

Equation (18.8) above does not contain a term representing the *amplitude* of the oscillation. This implies that the period of oscillation is always the same regardless of amplitude (the isochronous property mentioned earlier). A rough experimental check, using the trolley-and-springs oscillator, will show that this is true within the limits of accuracy obtainable with a stop-watch used for timing the oscillations. However, we can argue that, for this type of motion, the period must be independent of the amplitude. We can develop the argument with the help of Fig. 18.4. Suppose we had displaced the trolley to twice the initial value used in our example, that is,

to a distance of 0.400 m from equilibrium. The initial acceleration, being proportional to the displacement, will be twice what it was in the original example: thus the extra distance ΔS (which was shown to be equal to $a\,(\Delta t)^2$) will be twice what it was, and the distance corresponding to CD on the graph will be 0.016 m instead of 0.008 m. Thus the displacement at point D will be just twice what it was in the original example. Similarly, the displacement at any subsequent instant of time will be exactly twice what it was in the original example, with the result that the curve will cut the time-axis at precisely the same instant as it did in the original example. Therefore doubling the amplitude of oscillation does not alter the period. Similarly it can be argued that the period will be the same for any amplitude of oscillation. Thus we see that the isochronism of this type of oscillator is a logical consequence of the proportionality of acceleration and displacement. The argument can be expressed in words thus: if we pull the trolley out to twice its original displacement the restoring force is twice as great as it was in the first example: therefore the acceleration is twice as great, and the velocity the trolley acquires in a given time interval is twice what it was in the original situation. Hence the average velocity during any given time interval is twice what it was in the corresponding time interval for the original example. Hence the distance it travels during any chosen time interval is twice what it was in that time interval for the original example. But since it started its motion with twice the displacement of the original, it takes the same time to reach the equilibrium position as the original did.

An oscillator for which the acceleration is proportional to the displacement and is always directed back towards the equilibrium position is often known as a *simple harmonic* oscillator. As we have argued, such an oscillator must be isochronous. We have also shown that the period of a simple harmonic oscillator is given by the expression

$$T = 2\pi \sqrt{\frac{\text{mass}}{\text{restoring force per unit displacement}}}.$$

This relationship can be easily put to experimental test using the trolley-on-springs apparatus. A quick test is done by doubling the mass of the oscillating body, simply by stacking an identical trolley on top of the first one. The period of oscillation is increased by a factor of $\sqrt{2}$, as the expression above predicts. The restoring force per unit displacement can be doubled by adding another pair of springs, similar to the first pair, alongside the original ones. The effect of this is found to be a reduction of period by a factor of $\sqrt{2}$, again as is predicted by the expression.

We now shall examine some other simple oscillatory systems which fulfill the conditions required for the performance of simple harmonic motion; and we shall thereby be able to find by theoretical analysis the factors which determine their periods of oscillation. It should be emphasised that very few oscillatory systems are *truly* simple harmonic, thus allowing easy analysis. However, many systems approximate well to being simple harmonic oscillators provided that the amplitude of oscillation is kept small. The reason for this is that, although it is rarely the case that the restoring force is directly proportional to the displacement (which is the condition for simple harmonic motion), the restoring force may be approximately proportional to the displacement over a limited range of displacements. A good example of an oscillator which is not truly simple harmonic is the simple pendulum.

18.4 THE SIMPLE PENDULUM

A small, massive object suspended by a thread of negligible mass from a rigid support-point, allowed to swing in a vertical plane, is termed a *simple* pendulum. This is to distinguish it from the commoner, *compound* pendulum in which the mass has considerable spatial extension and cannot be considered as all being concentrated at one point as in the simple pendulum.

If you have not already done so you should investigate experimentally two important properties of a simple pendulum. The first of these is the very nearly perfect isochronism for small amplitudes of swing. It is most impressively demonstrated by setting up side by side two pendulums consisting of lead balls,

one or two centimetres in diameter, suspended by threads about two metres long. Having made sure the threads are of equal length, set one pendulum swinging with a very small amplitude, say, about 5 cm, and the other with an amplitude of about ten times as much. It should be found that they do not become noticeably out of step over a time of ten cycles or so. The other important property is that the period of swing does not depend on the mass of the pendulum bob, but only on the length of the suspension. To demonstrate this, replace one of the pendulum bobs by a much more massive one, ten, fifty, or even a hundred times the mass of the first. Again, if the two pendulums are set swinging in phase with each other they will be found to keep in phase for several cycles. The analysis which follows will show how these two properties arise.

Figure 18.6 shows a simple pendulum displaced from its equilibrium position. The displacement can be measured either by the angle θ made by the thread with the vertical, or by the distance, x, through which the centre of mass of the pendulum bob has moved along the arc of its path. The length of the suspension, measured from the point of support to the centre of mass of the bob, is L. The mass of the bob is m, and the gravitational field strength (i.e. the gravitational force per unit mass) is g.

First of all we want an expression for the force acting on the bob in its direction of motion. The tension in the suspending thread is always acting in a direction perpendicular to the motion of the bob so that it has no effect on the bob's motion. The weight or force of gravity on the bob is mg, and the component of this parallel to the direction of the bob's motion is $mg \sin \theta$. Thus:

$$\text{restoring force at displacement } x = -mg \sin \theta.$$

Now θ (measured in radians) $= x/L$. If θ is small we can use the approximation $\sin \theta = \theta$. How good an approximation this is, we shall consider later. Using the approximation we can write

$$\text{restoring force at displacement } x = -mg \frac{x}{L}$$

and so we see that the restoring force is directly proportional to the displacement. Also, it is clear

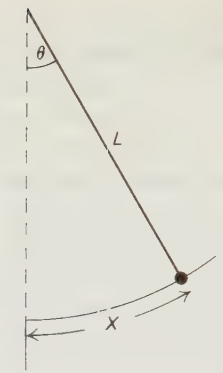


Fig. 18.6

that the force is always in the reverse direction to the displacement. The conditions for simple harmonic motion are therefore fulfilled, and so we should expect the oscillations to be isochronous, so long as the angle θ never becomes so large that the approximation $\sin \theta = \theta$ is invalid.

It was established earlier (Section 18.3) that, for a simple harmonic oscillator,

$$\text{period } T = 2\pi \sqrt{\frac{\text{mass}}{\text{force per unit displacement}}}$$

hence, for our simple pendulum,

$$\text{period } T = 2\pi \sqrt{\frac{m}{mg/L}}$$

hence

$$T = 2\pi \sqrt{\frac{L}{g}} \qquad (18.9)$$

and we see that the mass, m, does not appear in this expression, which predicts that the period is independent of the mass of the bob.

Problem 3 Oscillation of a vehicle on its sprung suspension.

In a certain automobile, when five passengers get in, their total mass being 360 kg, the springs of the suspension are compressed a distance 4 cm. The total mass including passengers supported by the

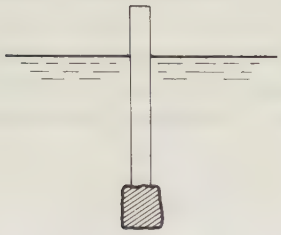


Fig. 18.7

suspension is 900 kg. Assuming that the displacement of the springs is directly proportional to the compressive force in them, calculate the period of oscillation of the loaded car on its suspension, neglecting any effects of damping (which of course is very heavy in a well-designed suspension system!). (Take $g = 10\ \text{N kg}^{-1}$.)

Solution First we find the force per unit displacement of the suspension springs. The weight of the passengers is $360 \times 10\ \text{N} = 3600\ \text{N}$, and since this causes a compression of 4 cm, which equals 0.04 m, the force per unit displacement is $3600/0.04\ \text{N m}^{-1} = 90\,000\ \text{N m}^{-1}$. Using the relation

$$\text{period} = 2\pi \sqrt{\frac{\text{mass}}{\text{force per unit displacement}}}$$

we get

$$T = 2\pi \sqrt{\frac{900}{90\,000}} = 2\pi \times \frac{1}{10} = 0.628\ \text{s}.$$

Problem 4 Oscillation of a climber dangling on a rope.

A typical mountaineer's rope, about 35 metres long, will "give" a distance of about 1.6 m under the weight of a climber hanging freely on the end. Assuming the climber's mass is 80 kg (about $12\frac{1}{2}$ stone), estimate the period of vertical oscillation of the climber when dangling freely on the end of the rope. How does this compare with the period of swing if the climber "pendules" (that is, swings on the rope like a pendulum-bob)?

Solution The solution goes on the same lines as the previous problem. The weight of the climber is $80 \times 10 = 800$ N, and so the force per unit displacement provided by the rope is $800/1.6\ \text{N m}^{-1} = 500\ \text{N m}^{-1}$. Hence the period is given by

$$T = 2\pi \sqrt{\frac{80}{500}} = 2\pi\sqrt{0.16} = 2\pi \times 0.4$$
$$= 0.8\pi = 2.5\ \text{s}.$$

The period of a simple pendulum is given by

$$T = 2\pi \sqrt{\frac{L}{g}}.$$

Substituting our numerical values we get

$$T = 2\pi\sqrt{\tfrac{35}{10}} = 2\pi \times 1.87 = 12\ \text{s}$$

very nearly.

Problem 5 Bobbing float in water. Any object floating in water, partly immersed, will, if displaced and let go, perform oscillations which quickly die away because of damping. What factors determine the period of such oscillation? One can get an appreciation of what they are by analysing a simplified situation. Figure 18.7 shows a uniform straight stick with a counterweight at its lower end to make it float upright. It has a total mass m and the area of cross-section of the stick is A. It is floating in a liquid of density ρ in a place where the gravitational field strength is g.

To find the period of small oscillations (vertical) we use the relationship

$$\text{period} = 2\pi \sqrt{\frac{\text{mass}}{\text{force per unit displacement}}}.$$

To get an expression for the force per unit displacement in this case we use Archimedes' principle: a body totally or partially immersed in a fluid experiences an upthrust equal to the weight of fluid displaced. When our stick is in equilibrium the pull of gravity on it is balanced by the upthrust of the liquid. If we

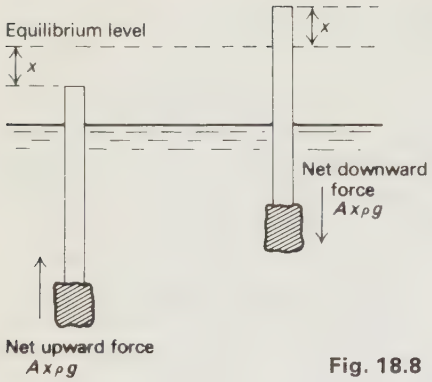


Fig. 18.8

displace it a distance x vertically downwards there is an increase in upthrust equal to the weight of extra liquid displaced. This increase in upthrust is the net upwards restoring force on the stick.

a) Consider the stick displaced vertically downwards a distance x. Write an expression for the net upwards force on it. Is this force directly proportional to x?

b) Consider the stick displaced vertically upwards a distance x, and write an expression for the net downwards force on the stick. Is *this* force directly proportional to x?

c) Write an expression for the force per unit displacement and hence derive an expression for the period of small vertical oscillation.

d) If the period in pure water is 1 second exactly, what is it in thin oil whose relative density is 0.81?

Solution

a) When the stick is displaced downwards a distance x it displaces an extra volume of liquid Ax. The weight of this extra displaced liquid is $Ax\rho g$. A, ρ, and g are all constant. By Archimedes' principle, this expression gives the net upwards force on the stick and clearly it is directly proportional to the displacement x (Fig. 18.8).

b) When displaced a distance x upwards, the upthrust on the stick is reduced by an amount $Ax\rho g$, following the same lines of argument as in (a). Since, before the stick was displaced from equilibrium, the upthrust balanced the stick's weight, this means that we have a net downwards force, tending to restore the stick to equilibrium, of amount $Ax\rho g$. Thus, once again, we have a restoring force proportional to the displacement.

c) Displaced either above or below equilibrium, the stick experiences a restoring force per unit displacement $A\rho g$. Thus the period of oscillation is given by

$$T = 2\pi \sqrt{\frac{m}{A\rho g}}.$$

d) The expression for the period shows that $T \propto (1/\sqrt{\rho})$. If T_1 is the period of oscillation in liquid whose density is ρ_1, and T_2 the period in liquid of density ρ_2, then

$$\frac{T_2}{T_1} = \sqrt{\frac{\rho_1}{\rho_2}}.$$

If T_2 is the period in the oil, then substitution of the numerical data yields $T_2 = 1.1$ second.

Problem 6 Pendulum clock in different gravity conditions.

a) If a pendulum clock which keeps correct time on earth were taken to the moon, how long would an hour indicated on the clock actually be, approximately? The moon's gravitational field strength is approximately one sixth of the earth's.

b) The earth's gravitational field strength weakens with increasing altitude above sea-level. For altitudes small compared with the earth's mean radius the formula

$$g = g_0 \left(1 - \frac{2h}{R}\right)$$

can be used, where g_0 is the field strength at sea level, and g is the field strength at an altitude h. R is the earth's mean radius, which is 6400 km, to two significant figures.

Estimate the number of seconds lost per day by a pendulum clock on top of a mountain 4000 m high, if the clock keeps correct time at sea-level. In making this estimate one can use approximations, and here are some guide-lines for the calculation:

i) Call the period of the pendulum at sea-level T_0, and at altitude h call it T. Remembering how the

period of any pendulum (of constant length) depends on g, write an expression for the ratio T/T_0 using the formula given above.

ii) Then write the expression you have just derived in the form: $T = T_0 + \varepsilon T_0$, where ε is the fractional change in period of the pendulum.

iii) The time lost in one day will be (in seconds): ε. (number of seconds in one day.)

Solution

a) The period of small oscillation of any pendulum (of constant length) is inversely proportional to $\sqrt{g}$. Calling the period on earth T_E and the period on the moon T_M, we have therefore

$$\frac{T_M}{T_E} = \sqrt{\frac{g_E}{g_M}}$$

and, putting in the numerical data, this ratio becomes equal to $\sqrt{6}$, which is 2.45. Thus an hour recorded by the clock on the moon would be nearly two and a half true hours.

b) As above,

$$\frac{T}{T_0} = \sqrt{\frac{g_0}{g}}$$

and since

$$g = g_0\left(1 - \frac{2h}{R}\right)$$

we can write

$$\frac{T}{T_0} = \sqrt{\frac{1}{\left(1 - \frac{2h}{R}\right)}} = \left(1 - \frac{2h}{R}\right)^{-1/2}.$$

Here we can make an approximation using the binomial theorem:

$$(1 + x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \text{etc.}$$

If $x \ll 1$, so that terms in x^2 and higher powers of x are negligible in comparison with the term in x, we can approximate thus:

$$(1 + x)^n \simeq 1 + nx$$

and so

$$(1 - x)^{-1/2} \simeq 1 - (-\tfrac{1}{2}x)$$

hence

$$\left(1 - \frac{2h}{R}\right)^{-1/2} \simeq 1 + \frac{h}{R}$$

and this is a valid approximation because, in our example, $h = 4000$ m, and $R = 6\,400\,000$ m, so that $h/R = \frac{1}{1600}$.

So we can write, to a very good approximation:

$$\frac{T}{T_0} = 1 + \frac{h}{R}$$

hence

$$T = T_0 + \frac{h}{R}T_0$$

(which is in the form $T = T_0 + \varepsilon T_0$, where ε is the fractional change in period of the pendulum). So the number of seconds lost in a day is

$$\varepsilon \text{ (number of seconds in a day)} = \tfrac{1}{1600}(24 \times 60 \times 60)$$
$$= 54 \text{ s.}$$

And this is certainly not a negligible error in a good clock!

18.5 THE ENERGY OF A SIMPLE HARMONIC OSCILLATOR

The important idea that will be developed in this section is that a simple harmonic oscillator's energy is in two forms, kinetic and potential, and that the energy is changing periodically from the one form into the other.

Consider the trolley-and-springs oscillator, as illustrated in Fig. 18.3. As was said earlier, if there is no energy loss as a result of friction or any other resistances to motion, the oscillator once started will continue to move back and forth with constant amplitude. This is an ideal situation never achieved in practice: but for the purposes of our argument we will suppose that we have such an ideal oscillator. At the ends of the motion, at maximum displacement, the velocity of the trolley is zero and therefore the kinetic energy is zero: the energy is entirely potential

in the stretched springs. We define the amount of potential energy as the work done in displacing the trolley from equilibrium to its extreme position. At the mid-point of the oscillation the potential energy is zero (by the above definition), but the velocity is a maximum and so the kinetic energy is a maximum, and the total energy of the oscillator is in kinetic form at this point. Thus we see that during one complete cycle of oscillation the energy of the oscillator changes thus:

$$\text{P.E.} \rightarrow \text{K.E.} \rightarrow \text{P.E.} \rightarrow \text{K.E.} \rightarrow \text{P.E.}$$

We shall now derive an expression for the energy of a simple harmonic oscillator. The mass of the oscillating body is m and the agency providing the restoring force gives a force per unit displacement k. The amplitude is a_0 and the angular frequency is ω. We shall need expressions for displacement and velocity at time t, thus:

$$x = a_0 \cos \omega t,$$
$$v = -a_0 \omega \sin \omega t.$$

The potential energy at a displacement x is given by the work done in displacing the object through a distance x from equilibrium. The force required to do this increases from zero at the equilibrium position according to some function f, of the displacement

$$\text{force} = f(s),$$

where s is the displacement from the origin.

Then the work done (which equals the potential energy gained)

$$= \int_0^x f(s)\, ds.$$

In our case

$$\text{force} = ks$$

and so

$$\text{potential energy} = \int_0^x k(s)\, ds$$
$$= \tfrac{1}{2}kx^2.$$

This result can be obtained without the use of the calculus for this simple case since the force is always

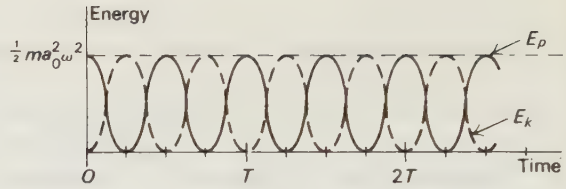


Fig. 18.9

directly proportional to the displacement and we may write

$$\text{potential energy} = \text{work done}$$
$$= \text{average force} \times \text{distance}$$
$$= \tfrac{1}{2}(0 + kx)x$$
$$= \tfrac{1}{2}kx^2.$$

Thus we can write an expression for the P.E. of the oscillator

$$E_p = \tfrac{1}{2}ka_0^2 \cos^2 \omega t. \qquad (18.10)$$

The K.E. is $\frac{1}{2}mv^2$, so the expression for K.E. becomes

$$E_k = \tfrac{1}{2}ma_0^2\omega^2 \sin^2 \omega t. \qquad (18.11)$$

It has already been shown earlier (Eq. 18.7) that

$$\omega^2 = \frac{k}{m},$$

hence $k = m\omega^2$ and using this substitution for k in Eq. (18.10) we can write

$$E_p = \tfrac{1}{2}ma_0^2\omega^2 \cos^2 \omega t. \qquad (18.12)$$

Figure 18.9 shows graphs of E_k and E_p as functions of time, and in this figure it can be seen that at any instant the sum of the two energy terms is a constant, $\frac{1}{2}ma_0^2\omega^2$. This fact can also be demonstrated algebraically, knowing that for any angle θ, $\cos^2 \theta + \sin^2 = 1$.

$$\text{Total energy} = E_k + E_p$$
$$= \tfrac{1}{2}ma_0^2\,\omega^2 \sin^2 \omega t + \tfrac{1}{2}ma_0^2\,\omega^2 \cos^2 \omega t$$
$$= \tfrac{1}{2}ma_0^2\,\omega^2 (\sin^2 \omega t + \cos^2 \omega t)$$
$$= \tfrac{1}{2}ma_0^2\,\omega^2 \qquad (18.13)$$

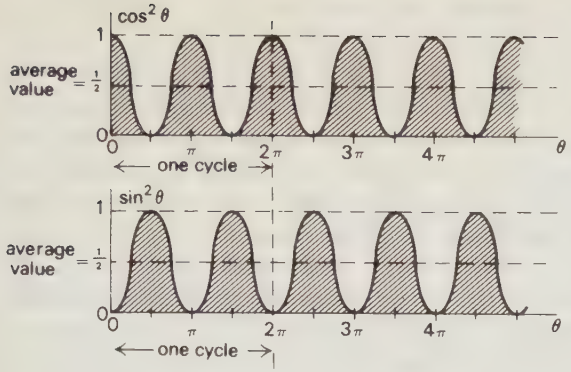


Fig. 18.10

As can be seen in Fig. 18.9, the period of the energy-cycle is half the period of oscillation. The total energy of a simple harmonic oscillator, as expression (18.13) shows, is proportional to the square of the amplitude, a_0, and to the square of the frequency, since $\omega = 2\pi f$ and so $\omega^2 = 4\pi^2 f^2$.

It now remains to show that the average kinetic energy and potential energy of a simple harmonic oscillator are equal. The averages are taken for one complete cycle of oscillation. To do this we simply need to prove that the average value of the function $\sin^2 \theta$ is equal to the average value of $\cos^2 \theta$ over the range $\theta = 0$ to $\theta = 2\pi$. Figure 18.10 shows these two functions plotted against θ. A graphical way of finding the average value of a varying quantity which is plotted on a graph is to find the area under the graph, then the average value of the quantity is represented by the straight line parallel to the horizontal axis which encloses an area beneath it equal to the area beneath the graph of the varying quantity. In Fig. 18.10 it is clear that the shaded areas beneath both graphs are equal, because the two graphs have the same shape and are simply "phase-shifted", as it were, one with respect to the other.

It is easily proved mathematically that the average value, taken over any whole number of cycles, of the functions $\cos^2 \theta$ and $\sin^2 \theta$, is $\frac{1}{2}$, as is clear from Fig. 18.10. Consider first the function $\cos^2 \theta$. The student familiar with trigonometrical functions will know that

$$\cos 2\theta = \cos^2 \theta - \sin^2 \theta$$

also that

$$\cos^2 \theta + \sin^2 \theta = 1$$

therefore

$$\sin^2 \theta = 1 - \cos^2 \theta$$

so we can write

$$\cos 2\theta = \cos^2 \theta - (1 - \cos^2 \theta)$$

therefore

$$\cos 2\theta = 2 \cos^2 \theta - 1$$

hence

$$\cos^2 \theta = \tfrac{1}{2} + \tfrac{1}{2} \cos 2\theta.$$

Now the average value of a cosine function, taken over a whole number of cycles, is zero. So the average value of the $\frac{1}{2} \cos 2\theta$ part of the expression is zero, and thus the average value of $\cos^2 \theta$ is simply $\frac{1}{2}$. Likewise it can be shown that

$$\sin^2 \theta = \tfrac{1}{2} - \tfrac{1}{2} \cos 2\theta$$

and therefore the average value of this is also $\frac{1}{2}$.

It follows from the preceding argument that the average kinetic energy and the average potential energy of a simple harmonic oscillator are both $\frac{1}{4} m a_0^2 \omega^2$. This fact has important consequences in the theory of thermal energy possessed by substances in the molecular state in which the atoms can vibrate and behave to some extent as simple harmonic oscillators.

Chapter 19
DAMPED OSCILLATIONS

19.1 Damped oscillations

19.1 DAMPED OSCILLATIONS

In all the oscillatory systems so far analysed it has been assumed that there is no energy loss due to friction or any other resistances to motion: thus the amplitude of oscillation has remained constant. In any real-life mechanical oscillatory system performing free oscillations, there *are* resistances to motion which cause a loss of energy and a decay of the amplitude of oscillation. Oscillations decaying in this way are known as *damped* oscillations, and the term *degree of damping* is generally used to signify how rapidly the amplitude decays. A heavily damped oscillation decays rapidly in amplitude, and a lightly damped oscillation decays slowly. Figure 19.1 shows the displacement-time for (a) an undamped, (b) a lightly damped, (c) a heavily damped, and (d) a very heavily damped oscillator.

There is a neat and simple demonstration of damped oscillations in which the oscillator draws its own displacement-time graph. Figure 19.2 shows the arrangement. The oscillator is simply a wooden lath about two metres long, about five centimetres wide and between half a centimetre and one centimetre thick. It is clamped firmly with G-cramps to two legs of a table or bench, and the free, vibrating end has a fine paint-brush attached to it which, wetted with ink, draws a trace on a sheet of paper pulled along the floor. The degree of damping can be increased by making use of air-drag, pinning a large sheet of card (about 30 cm × 20 cm should be large enough to produce an observable increase) in a vertical plane on to the free end of the lath.

We shall now consider, with reference to some simple laboratory experiments, the causes of damping in different situations, and the pattern of decay of amplitude in those situations.

The trolley-and-springs oscillator, already discussed in some detail, is in practice a fairly heavily damped oscillator. The energy loss causing damping is due mainly to friction in the bearings of the wheels and partly to the aerodynamic drag force (air drag) on the trolley when it is moving. The air drag is a force dependent on speed: the faster an object moves through the air the greater the air drag. The object

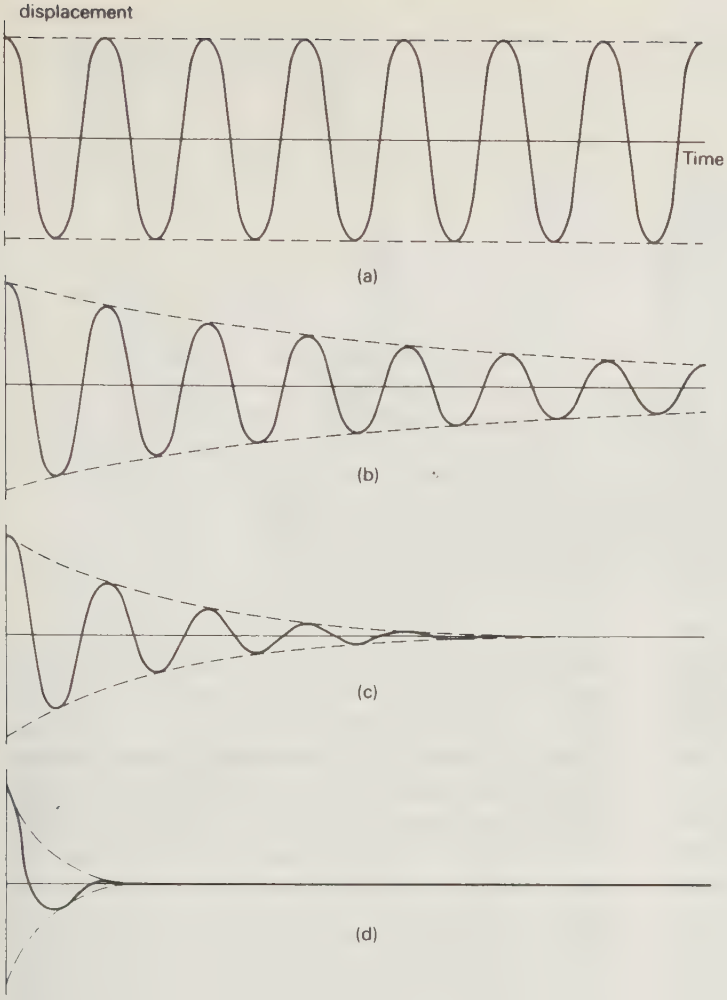


Fig. 19.1

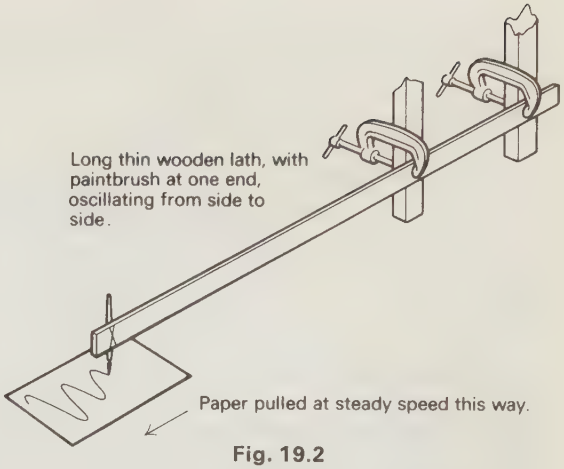


Fig. 19.2

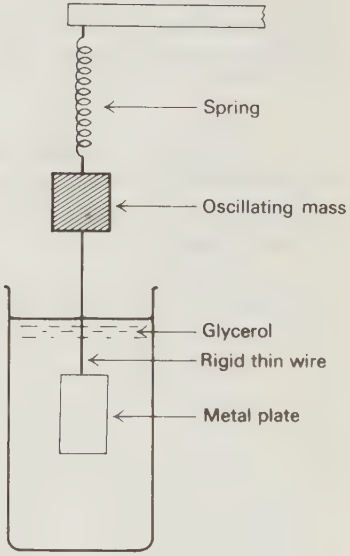


Fig. 19.3

"stirs up" the air and thus loses energy which becomes kinetic energy of the stirred up air—hotter air. The resisting force due to friction in the wheel-bearings is an uncertain factor: one cannot assume that it is a constant force, on the other hand one cannot say reliably how this force depends on speed. Indeed, damping forces are not generally easy to analyse mathematically. One certainly cannot rely on the frictional force between two surfaces in relative motion being constant; and the force of air drag, although it is approximately proportional to the square of the speed of an object travelling through the air, is difficult to control, and depends in a sensitive way on the nature of the surface (shape and texture) of the moving object. For experimental investigation it is useful to have a system in which the damping force is controllable. Figure 19.3 shows an oscillator with *viscous* damping. If an object (in this case it is

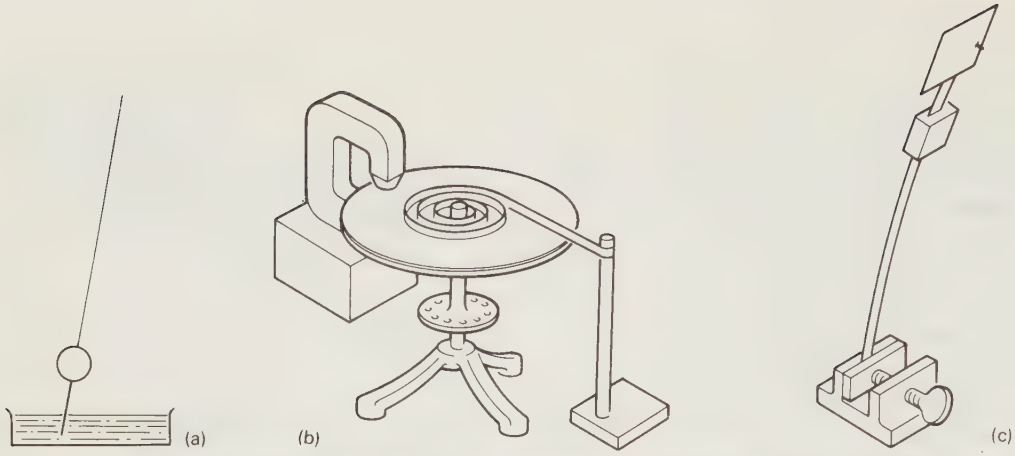


Fig. 19.4 Some systems for investigating damped oscillations. (a) A long pendulum with a heavy bob. The damping is provided by a rod attached to the bob which drags in an oil bath. (b) A turntable oscillator. The restoring force is provided by a spiral spring whilst damping is provided by a powerful magnet inducing eddy-currents in an aluminium disc attached to the turntable. (c) A strip of spring steel with a load at the top end. Damping is provided by the air drag on a card attached to the top of the strip. The lower end of the strip is very firmly clamped to a rigid support.

the flat metal plate) travels through a fluid in such a way that the fluid-flow relative to the object is streamline flow, then the viscous force resisting motion is found experimentally to be directly proportional to the speed of the object.

The magnitude of the viscous force also depends on the area of surface in contact with the fluid, and on the coefficient of viscosity of the fluid. Using glycerol, a highly viscous liquid, heavy damping can be produced with a metal plate of not too large an area. Lighter damping can be obtained in the same liquid by using plates of smaller area.

Several damped oscillatory systems with which it is possible to measure the amplitude of oscillation and observe the way in which it decays can quickly be set up. Each system can be arranged to have a sufficiently long period of oscillation so that a number of measurements of amplitude can be made during decay. In each case a pointer can be attached to the oscillating object and made to move along a centimetre scale. Suitable systems are shown in Fig. 19.4.

If a graph of amplitude against time is plotted for each of the systems of Fig. 19.4 it will be found that the graphs all have one thing in common: the amplitude decays rapidly at first, and then less and less rapidly as time goes on. But the precise decay-pattern is not the same for each system, as Fig. 19.5 shows. In this graph the amplitudes of the three different oscillators have been plotted on a common scale, the amplitude being expressed as a percentage of the initial amplitude. Curve (b), for the oscillator with viscous damping, shows a mathematically simple decay-pattern: note that it takes the same time for the amplitude to decay from 100% to 50% as it does to decay from 50% to 25%, or from 25% to 12.5%. In fact, choose any value of amplitude and you will find that it takes the same time for the amplitude to be halved: e.g. from 80% to 40% takes the same length of time. This simple mathematical property is not shown by curves (a) and (c).

The simple decay-pattern illustrated by curve (b) in Fig. 19.5 is known as *exponential* decay. We shall briefly investigate the mathematical properties of this particular form of decay.

Its distinguishing property, which we have already introduced, is that the amplitude decreases by the same ratio in successive equal intervals of time, that is,

$$\frac{a_0}{a_1} = \frac{a_1}{a_2} = \frac{a_2}{a_3} = \cdots = \text{a constant.}$$

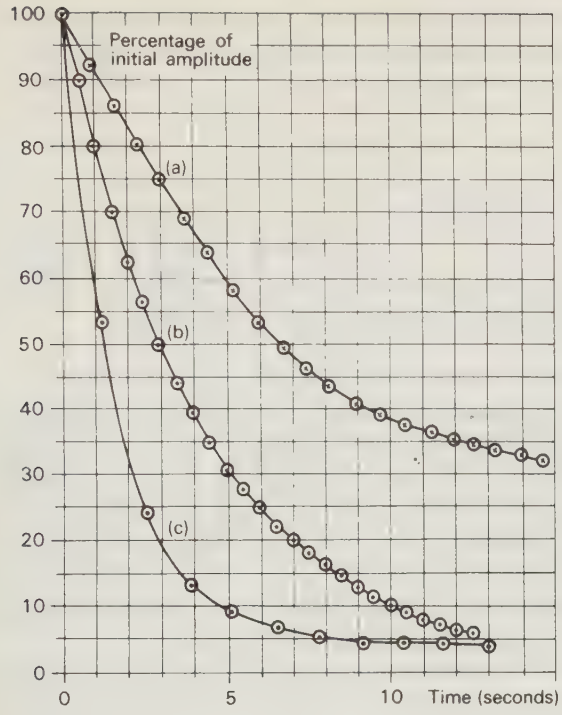


Fig. 19.5 Decay graphs for damped oscillators. Each plotted point corresponds to a measurement of the maximum displacement of the oscillator on one side of its equilibrium position. The general decay pattern is shown by a continuous curve made to fit the points.

Knowing this we can derive an expression for the amplitude as a function of time: that is to say, we can obtain the equation for the graph of Fig. 19.5(b). Taking logarithms (natural logarithms, to the base e) we can write:

$$\ln \frac{a_0}{a_1} = \ln \frac{a_1}{a_2} = \ln \frac{a_2}{a_3} \cdots = \Lambda,$$

where Λ is a constant called the logarithmic decrement. Hence

$$\ln \frac{a_0}{a_2} = \ln \left(\frac{a_0}{a_1} \times \frac{a_1}{a_2} \right) = \ln \frac{a_0}{a_1} + \ln \frac{a_1}{a_2} = 2\Lambda$$

and similarly

$$\ln \frac{a_0}{a_5} = 5\Lambda$$

so that in general, if a_t is the amplitude at time t, we can write

$$\ln \frac{a_0}{a_t} = \Lambda t.$$

Taking antilogarithms of both sides we obtain

$$\frac{a_0}{a_t} = e^{\Lambda t}$$

hence

$$a_t = a_0 e^{-\Lambda t}. \tag{19.2}$$

The greater the value of the logarithmic decrement, Λ, the more rapid is the decay.

In the physical world there are many situations where a measurable quantity decreases with time according to this exponential law. Examples are the charge on the plates of a capacitor discharging through an electrical resistance, and the activity of radioactive substance undergoing radioactive disintegration. A useful concept in such decay processes is that of *half-life*. The half-life of a quantity which is decaying exponentially is defined as the time taken for the quantity to halve its value. The half-life of the oscillatory system which gave curve (b) in Fig. 19.5 is about 3 seconds. Let us see how the half-life is related to the logarithmic decrement, Λ. The half-life is the time $t_{1/2}$ for which $a = \frac{1}{2}a_0$, that is,

$$\tfrac{1}{2}a_0 = a_0 e^{-\Lambda t_{\frac{1}{2}}}$$

hence

$$e^{\Lambda t_{\frac{1}{2}}} = 2.$$

Taking logarithms:

$$\Lambda t_{1/2} = \ln 2 = 0.693$$

therefore

$$\Lambda = \frac{0.693}{\text{half-life}}. \tag{19.3}$$

Up to this point our analysis has concerned only the decaying amplitude of oscillation: we have not concentrated attention on the variation of displacement with time in a decaying oscillation. Looking again at Fig. 19.1 we see that curve (a) is an undamped oscillation which, as we know, can be represented by the expression

$$x = a_0 \cos \omega t.$$

What is the form of the equation to the curves (b) and (c)? Instead of a constant amplitude a_0 in the expression above we must write a term for the decaying amplitude: this means that the expression will be of the form

$$x = a_0 e^{-\Lambda t} \cos \omega t. \qquad (19.4)$$

There now arises a question which might easily be overlooked. If we have an undamped oscillator, whose displacement can be represented by $x = a_0 \cos \omega t$, with angular frequency ω, and we then apply some damping to it, will its frequency be the same as that of the undamped oscillator? It will, in fact, *not* be the same, although in practice the effect of altering the degree of damping on an oscillator makes negligibly small changes in the frequency. We have already shown that, for an undamped simple harmonic oscillator,

$$\omega^2 = \frac{k}{m},$$

where m is the mass and k the restoring force per unit displacement. It can be shown mathematically that for an oscillator with viscous damping, for which the displacement is given by

$$x = a_0 e^{-\lambda t} \cos \omega t$$

the angular frequency is given by

$$\omega^2 = \frac{k}{m} - \Lambda^2 \qquad (19.5)$$

which shows that the effect of damping is to reduce the frequency (hence to increase the period) of oscillation. To appreciate how small this increase in period is, we can use numerical values appropriate to the experimental oscillator of Fig. 19.3 which gave an amplitude decay curve like curve (b) of Fig. 19.5. The mass of the oscillating load was 0.5 kg, and the stiffness constant (that is, the force per unit displacement) of the spring very nearly 80 N m^{-1}. Thus the angular frequency of undamped oscillation is given by

$$\omega^2 = \frac{80}{0.5} = 160$$

(whence f is very nearly 2 Hz, and so the period is very nearly half a second). The half-life of the oscillations is seen to be about six cycles, that is, about 3 s^{-1}. Using the relation

$$\Lambda = \frac{0.693}{\text{half-life}}$$

this means that

$$\Lambda = 0.23 \text{ s}^{-1}.$$

Using expression (19.5) above this gives, for the angular frequency of damped oscillations

$$\omega^2 = 160 - 0.23^2$$

which shows clearly that the effect of damping is to reduce the angular frequency only very slightly. $0.23^2 = 0.05$ approximately, and so the percentage change in ω^2 due to damping is about $0.05/160 \times 100 = 0.03\%$ only. This will mean a percentage change in frequency, and period, of half this value: 0.015%. This is far too small to be detectable using a stop-watch to time the oscillations.

It is interesting to predict how much damping would be needed to produce, say, a one percent change in frequency, which one would have a fair chance of detecting with a stopwatch. We require, then, that Λ^2 should be about 2 percent of k/m (since a change of 1 percent in ω implies a change of 2 percent in ω^2). If, referring to the same oscillator as before, $k/m = 160$, then Λ^2 must be approximately 3, since 3 is roughly 2 percent of 160. Hence

$$\Lambda = 1.7 \text{ s}^{-1} \qquad \text{approximately,}$$

and since

$$\Lambda = \frac{0.693}{\text{half-life}}$$

we must have

$$\text{half-life} = \frac{0.692}{1.7} \text{ s} = 0.4 \text{ s} \qquad \text{approximately.}$$

This means that the damping must be heavy enough to make the amplitude decay to half its initial value after a time a little less than the period (0.5 s) of undamped oscillations. Curve (c) of Fig. 19.1 has about this degree of damping, in which the amplitude is initially approximately halved every cycle.

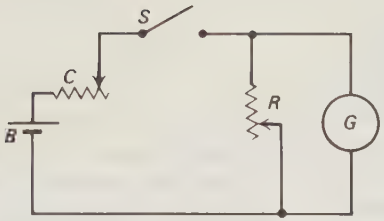


Fig. 19.6 Circuit for the investigation of damping in a ballistic galvanometer.

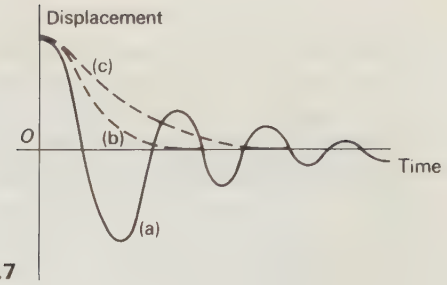


Fig. 19.7

Thus we see that, even with quite heavy damping, the period of oscillation differs only by a very small amount from the period of undamped oscillation being, in fact, very slightly greater. This means that, when we are performing laboratory experiments with not too heavily damped oscillators we can apply our simple theory of undamped oscillations (for example we can assume that $\omega^2 = k/m$) to a high degree of accuracy. But we must not finish this section of the work without considering the effects of *very* heavy damping on an oscillatory system.

To investigate a wide range of degrees of damping it is best to use a semi-electrical method, because it is difficult to get satisfactory results with a simple mechanical oscillator. The coil of a ballistic galvanometer is an oscillatory system having very little damping when the terminals of the galvanometer are on open circuit. Figure 19.6 shows the arrangement used for the investigation. Extra damping is provided by the resistance R which is variable. A high resistance produces light damping, and the damping can be made as heavy as desired by reducing the resistance. The other variable resistor C is used to select a suitable initial displacement of the galvanometer coil. When the switch S is closed the displacement is constant. The switch is then opened and then the galvanometer coil performs a damped oscillation. Typical results are shown in Fig. 19.7 for three different degrees of damping. Curve (a) shows that the galvanometer coil made a few cycles of oscillation before coming to rest. Curve (c) shows the effect of very heavy damping: the motion certainly cannot be described as an oscillation! Curve (b) shows the effect of an intermediate degree of damping where the damping is just sufficiently great to prevent any overshoot beyond the equilibrium position. This amount of damping is known as *critical damping*, and under critical damping an oscillatory system reaches equilibrium, after initial displacement, in the shortest possible time.

The damping in the experiment just described arises from loss of electrical energy. When the coil of the galvanometer rotates between the poles of its magnet an e.m.f. is induced in it, and this e.m.f. is proportional to the speed of rotation. The rate at which energy is converted from electrical form into heat in the circuit is E^2/R, where E is the e.m.f. and R is the resistance of the circuit containing the galvanometer coil. Thus the rate of energy loss is proportional to the square of the coil's speed, and this implies that the resulting force opposing the coil's motion is directly proportional to the speed of the coil. Thus we have damping of the viscous type. Also, since the rate of energy loss is inversely proportional to the circuit resistance it follows that the damping is increased by reducing the resistance.

A ballistic galvanometer is designed to have very little damping of its own. It is used for measuring electric charge which is passed through the coil in a time which must be very short compared with the natural period of oscillation of the coil: one then measures the "fling", that is, the first (and largest) displacement of the coil away from equilibrium. But a galvanometer designed to register steady electric currents has heavy damping built into its mechanism. If it had no such damping the taking of current measurements would be a laborious, even impossible, business: one would have to wait a very long time

after the current had been switched on until the pointer came finally to rest! The ideal situation is a critically damped galvanometer coil, because this will reach a steady reading in the shortest possible time. This damping is usually introduced by making the coil frame from a strip of light metal, e.g. aluminium, which behaves as a single-turn coil. Eddy currents are induced in the frame as the coil rotates and, as a result, electrical energy is dissipated in the frame itself. By arranging for the frame to have the correct electrical resistance, critical damping can be achieved (see Unit 6).

The loss of energy due to damping now needs to be considered briefly. We will restrict the discussion to viscous damping in which the amplitude obeys the exponential decay law. If a_0 is the initial amplitude, a_1 is the amplitude after one cycle, a_2 after two cycles, and so forth, then, as has already been established (Eq. 19.1)

$$\frac{a_0}{a_1} = \frac{a_1}{a_2} = \frac{a_2}{a_3} \quad \text{etc.}$$

The fraction of the oscillator's total energy lost during the first cycle is

$$\frac{a_0^2 - a_1^2}{a_0^2}$$

since the energy is proportional to the square of the amplitude. It follows that

$$\frac{a_0^2 - a_1^2}{a_0^2} = \frac{a_1^2 - a_2^2}{a_1^2} = \frac{a_2^2 - a_3^2}{a_2^2} \quad \text{etc.,}$$

so that, with this type of decay, the oscillator loses the same fraction of its energy per cycle.

We have assumed here that the energy of the oscillator is directly proportional to the square of the amplitude. This relationship was shown to be true, in Section 18.5, for an *undamped* oscillator. Is it also true for a damped oscillator? It is true, in fact, only at the instants when the displacement of the oscillator is a maximum and the velocity is zero, for then all the energy of the oscillator is in potential form, and the velocity-dependent damping force is zero; also, as was shown in Section 18.5, the maximum potential energy of a simple harmonic oscillator is $\frac{1}{2}ka_0^2$, that is, it is directly proportional to a_0^2. And, if you think about it, you will realise that the amplitude of a damped oscillator is indefinable except at the instants of maximum displacement, because it is changing and only measurable at those instants. Thus the curved lines joining the points in the graphs of Fig. 19.5 have really no meaning: only the individual points have physical significance.

Chapter 20
FORCED OSCILLATIONS

20.1 Forced oscillations
20.2 Resonance

20.1 FORCED OSCILLATIONS

A very simple experiment serves to introduce the idea of a *forced* oscillation. Set up a simple pendulum (any small heavy object on a thread will do) with your hand as the support. You will find that you can excite it into oscillation by very small rhythmic movements of your hand from side to side. If the movements of your hand are in rhythm with the natural, free vibrations of the pendulum, you can build up an amplitude which is very much greater than the amplitude of your hand's motion. If now you keep your hand perfectly still the amplitude of oscillation will of course decay as a result of damping until all motion ceases: but you can maintain oscillation at a constant amplitude by keeping your hand moving very slightly, all the time in rhythm with the natural swinging of the pendulum. The movement of your hand may be almost imperceptibly small. In a pendulum clock the amplitude of swing is kept constant by the escapement mechanism which delivers small impulses to the pendulum, in rhythm with its natural swing. The loss of energy due to damping is made good by energy fed from the wound-up clock spring (or the raised weights, as in a grandfather clock), *via* the gear trains and the escapement, to the pendulum.

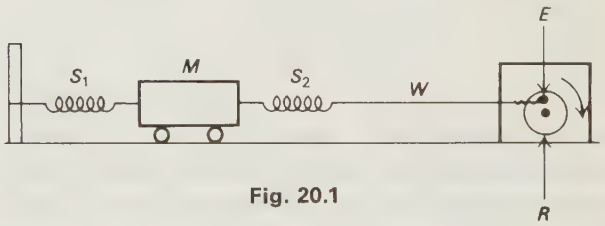


Fig. 20.1

Figure 20.1 shows a way of demonstrating forced oscillations in the laboratory, using the trolley-and-springs oscillator. The trolley, M, is attached by springs, S_1, to a rigid support. The other spring, S_2, is joined to a wire W whose other end is bent to form a loose-fitting ring which is slipped over a rod E mounted eccentrically on a small wheel R. The wheel is rotated by a motor which can be set to run

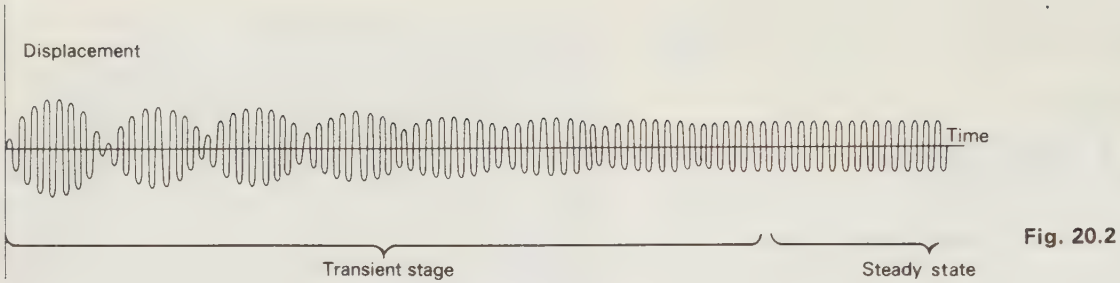


Fig. 20.2

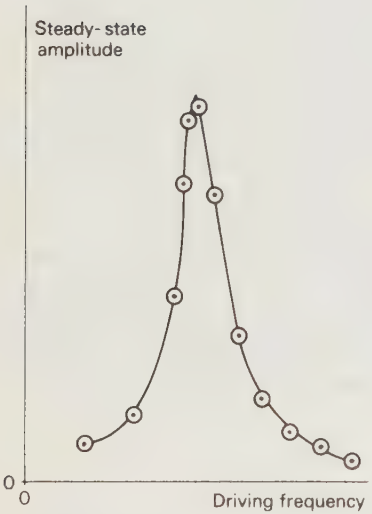


Fig. 20.3

at any desired speed. When the system is started from rest the motor quickly reaches a steady speed after being switched on. But the movement of M, which should be studied closely, is puzzling: it takes a fair length of time for the amplitude of oscillation to settle to steady value. The graph of Fig. 20.2 shows how the displacement varies with time at the start of the process. The initial stage of the motion during which the amplitude is not constant is known as the *transient* stage: after that we have the steady-state oscillation.

For the time being we shall concentrate attention on the steady-state oscillation. When one experiments with the effect of changing the speed, setting it at different speeds and then observing the steady-state oscillation which is attained at each motor speed, one finds that the steady-state amplitude depends on the motor speed in a sensitive way. The amplitude is greatest when the frequency of the motion of the rod E (which is the "driver" in this situation) is equal to the frequency of free oscillation of the trolley and springs arrangement (which we can call the "responder"). At any other driving frequency the amplitude of the responder is less, and the greater the difference between the driving frequency and the natural frequency of the responder, the smaller is the amplitude. Figure 20.3 shows a typical result from this experiment.

The terms *driver* and *responder* are engineering language and very useful in this context. "Driver" refers to whatever agency is providing a periodic force which is acting upon the responder: the "responder" being the complete oscillatory system which has its own characteristic frequency (or frequencies) of oscillation when allowed to oscillate freely (i.e. undriven).

The effect of damping on forced oscillations is of prime importance. An experiment in which one can vary the amount of damping applied to a forced oscillator is based on the apparatus used earlier to investigate damping (see Fig. 19.3). The arrangement is shown here in Fig. 20.4. The responder is the mass attached to the spring, and controllable damping is provided by the metal plate moving in the glycerol. The driver is a small wheel with an eccentrically mounted rod E (as in the apparatus of Fig. 19.3),

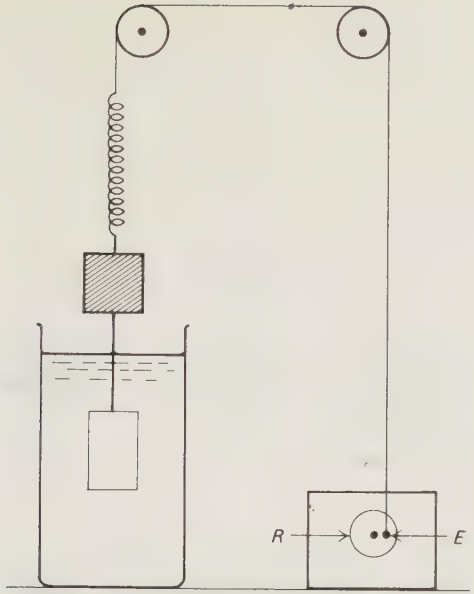


Fig. 20.4

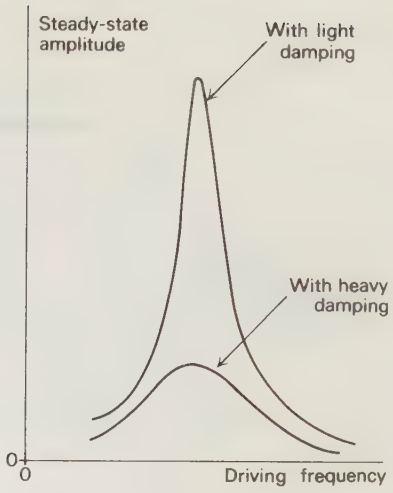


Fig. 20.5

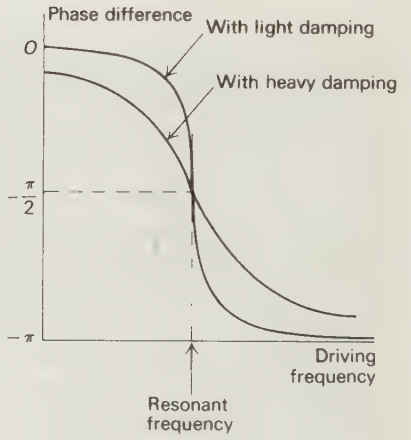


Fig. 20.6

the linkage from it to the responder being the cord which runs over the two pulleys. An experimental run can be done with light damping, using a metal plate of small area in the glycerol, and measuring the steady-state amplitude of oscillation for different settings of motor speed. The driver frequency is measured by having a revolution counter attached to the wheel R and counting the number of revolutions made over an interval of one minute or so. The experiment can be repeated with heavy damping provided by a plate of large area. Provided the plates are thin and have very small mass compared with the oscillating mass permanently attached to the spring, the extra mass of the larger plate will have negligible effect on the frequency of free oscillation of the system. Figure 20.5 shows how the steady-state amplitude depends on driver frequency for light and heavy damping.

When the steady-state amplitude has its maximum value, that is, when the driving frequency is equal to the natural frequency of free oscillation of the system, we have the condition known as *resonance*. The system can be said to *resonate* to the driving agency. With light damping a sharp resonance is obtained: the variation of amplitude with change in frequency is very marked when the driving frequency is close to the natural frequency of the responder. With heavy damping a broader resonance is obtained. In fact the resonant frequency itself is slightly altered when the degree of damping is varied, but the difference is scarcely perceptible in this practical example, and here we shall ignore it.

The degree of damping also affects the way in which the phase-difference between driver and responder varies with the driving frequency. The graph in Fig. 20.6 shows this. These phase relationships can

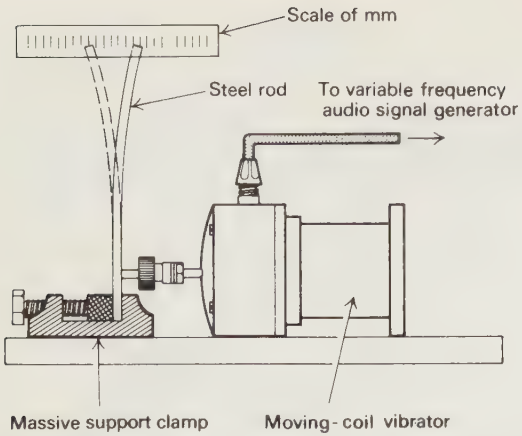


Fig. 20.7

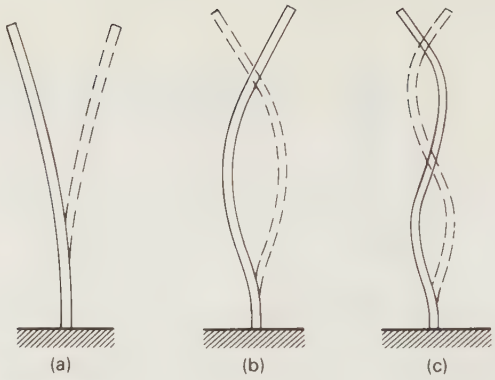


Fig. 20.8

be observed, using the apparatus of Fig. 20.4, by attaching an eye-catching horizontal pointer (e.g. a thin strip of brightly-coloured paper) to the load on the end of the spring, and a similar pointer to the part of the string between the top of the spring and the pulley above it. The relative motion of these two pointers shows clearly that driver and responder move very nearly in phase with each other when the driving frequency is much less than the resonant frequency, and with almost exactly opposite phase when the driving frequency is much higher than the resonant frequency. At resonance there is one quarter-cycle phase difference between driver and responder. Figure 20.6 shows the phase-difference between driver and responder as a function of driving frequency, for lightly and heavily damped forced oscillations.

20.2 RESONANCE

The phenomenon of *resonance* is one of far-reaching importance both in theoretical physical science and in the world of engineering, and in this section we shall discuss a few examples.

The oscillatory system shown in Fig. 20.4 has one and only one natural frequency of oscillation. It resonates at that frequency and only that frequency, when it is subjected to a periodic forcing agency of variable frequency. This implies that it is a very simple mechanical oscillatory system, and engineers very rarely deal with systems as elementary as this! Even a simple structure, such as a metal table with four legs and diagonal bracing struts, for supporting a piece of machinery, is capable of oscillating in different modes at many different frequencies. These frequencies can be calculated if one knows the dimensions, masses, and elastic properties of the various parts of the structure; but the calculation cannot be done without the use of a computer and even then the task of writing the computer program is long and intricate. In practice an engineer is more likely to make vibration tests on the structure itself, or on a scale model of the structure, by observing how it responds to being driven by a vibrator of variable frequency. Resonance in such a structure is generally highly undesirable, because the large amplitudes of oscillation of the structure at resonance may be sufficient to shake its fixings loose or even to damage the machinery supported by the table. In actual use the structure may be excited into resonant oscillations by the action of the machinery supported by the structure.

A very simple experiment illustrates well the existence of more than one resonant frequency in a

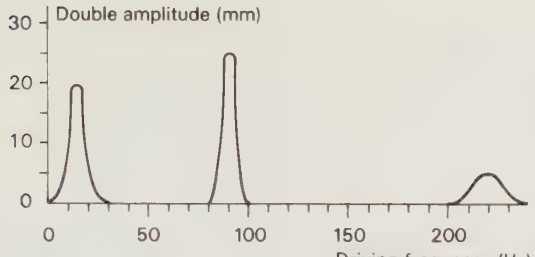


Fig. 20.9

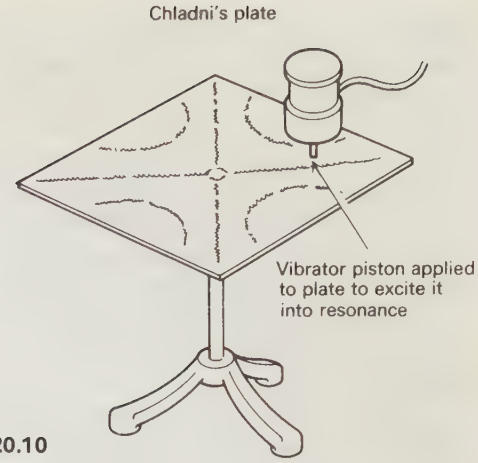


Fig. 20.10

mechanical structure, although the system in this example is so simple that it scarcely deserves the name *structure*, being no more than a uniform, springy, steel rod clamped firmly at one end with the other end free to vibrate. This rod is forced into vibration by a moving-coil vibrator which is driven by an audio-frequency signal generator. The arrangement is illustrated in Fig. 20.7. To start with, the frequency is set at its lowest value (say about 10 Hz). The rod is observed to vibrate, and after a while it attains a steady-state oscillation and the amplitude of vibration of the end of the rod can be measured on the scale of millimetres. Rough measurements of steady-state amplitude are made at different frequencies, and it is observed that there is a distinct resonant frequency at which the rod vibrates with very large amplitude in the mode (a) of Fig. 20.8. As the frequency is increased further, the amplitude decreases, but at a certain higher frequency the rod begins to oscillate in a different mode, (b) in Fig. 20.8. A third mode (c) can also be observed at a higher frequency still. If a rough graph of amplitude and driving frequency is made, it will look like the graph of Fig. 20.9.

In the experiment which yielded the results displayed in Fig. 20.9 the first resonance (fundamental mode) occurred at a frequency of about 14 Hz, and the second one at 90 Hz. The third resonance was of much smaller amplitude than the others, but nevertheless a distinct maximum of amplitude, at a frequency of about 220 Hz. Resonances at higher frequencies were detectable by ear as distinct peaks in the volume of sound emitted by the vibrating rod at certain frequencies, but the amplitude of vibration was too small to be observable with the naked eye. It is of interest to note here that the modes of vibration of a rod rigidly clamped at one end (what engineers call a *cantilever*) are the same modes as those of a tuning-fork's prongs. When a tuning-fork is struck the fundamental mode is the most prominent and long-lasting and is responsible for the characteristic note emitted by the fork; but the higher modes of oscillation are excited when the fork is struck and these are responsible for the "clang" which dies away rapidly, much more rapidly than the fundamental vibration.

A rod clamped rigidly at both ends has characteristic modes of vibration, each with its own natural frequency: these are different from the modes of vibration of a rod which is clamped only at one end. A uniform flat plate (a *lamina*) of elastic material—any reasonably springy metal will do—has characteristic modes of vibration which have a two-dimensional pattern. Fascinating experiments can be performed with Chladni's plate (Fig. 20.10). The

plate can be made to resonate in its many modes at different frequencies. It can be regarded as the two-dimensional counterpart of the vibrating rod of Fig. 20.7. Although the amplitude of vibration is generally too small to be observed, the pattern of the vibration is revealed beautifully by sprinkling a thin layer of fine dry sand evenly on the plate. When the plate is vibrating in one of its normal modes the sand is shaken away from the regions of maximum amplitude and collects in regions of minimum amplitude. Intriguing symmetrical patterns are revealed at different frequencies.

The reader will by now appreciate that a structure consisting of metal plates and metal rods is capable of a very large number of possible modes of oscillation, any one of which could be excited by an applied periodic force of the right frequency. What is more, there will exist new modes of vibration involving groups of rods or plates, extra to the simple modes characteristic of each rod and plate by itself! It is a sobering exercise for the would-be engineer to consider what composite modes of vibration can exist even in a very simple structure, and Problem 5.9 provides an example for consideration.

In certain conditions individual atoms and molecules behave like oscillatory systems having a number of natural modes of oscillation. The process in which a molecule can show resonant behaviour in response to a periodic excitation is outlined briefly, later in this chapter, and the absorption of energy from electromagnetic radiation, by atoms, at certain characteristic frequencies can be regarded as a situation of resonance (see Unit 8, on absorption spectra).

We return now to the simple forced oscillatory system with viscous damping of Fig. 20.4 to consider the shape of the resonance curves (see Fig. 20.5) which are characteristic of that system. We have observed that, the smaller the degree of damping, the sharper is the curve. It will now be shown that a quantitative measure of the width of a resonance curve is a useful concept. We define the width in terms of the graph of (amplitude)^2 against frequency, as in Fig. 20.11. The square of the amplitude is chosen, rather than the amplitude itself, because the energy of the oscillator is proportional to (amplitude)^2,

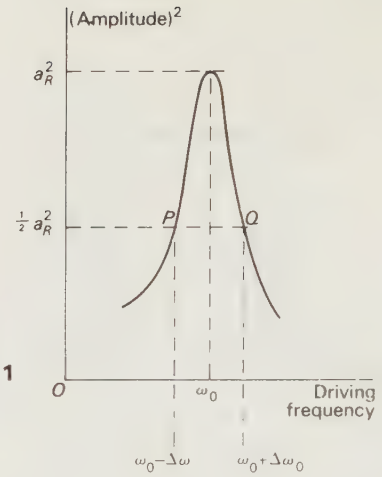


Fig. 20.11

and energy is generally the important parameter in such a system. The width of the resonance peak is defined as the difference in frequency between the points at which the squared amplitude is half the value of the squared amplitude at resonance, as can be seen in Fig. 20.11. If the damping is not too heavy, the shape of the peak between points P and Q is very nearly symmetrical about the resonant frequency, so that the frequency at point P can be written $\omega_0 - \Delta\omega$, and at point Q, $\omega_0 + \Delta\omega$, where ω_0 is the resonant (angular) frequency. Thus the width of the peak can be written $2\Delta\omega$. It is shown in Appendix 5 that $\Delta\omega = \Lambda$, where Λ is the decay constant for the oscillator when it undergoes free, damped oscillations, in the absence of any driving agency. Clearly, then, the smaller the damping, the smaller the value of Λ, and hence the narrower the width of the resonance peak.

It must be emphasised that the concept of width of a resonance peak is applicable only to situations in which the degree of damping is small enough not to invalidate the assumption (which is made in Appendix 5 before making the definition of width) that one is dealing only with frequencies which are very close to the resonant frequency, so that $\Delta\omega$

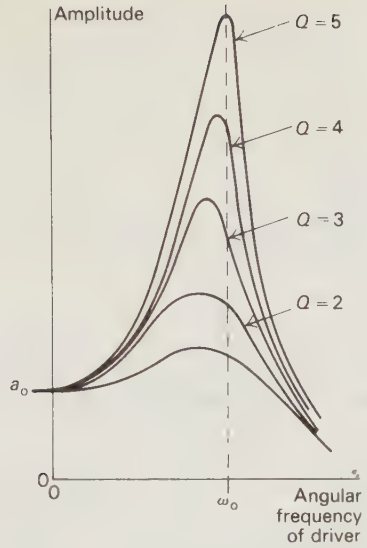


Fig. 20.12

is much smaller than ω_0 itself. In Fig. 20.12 you will see several resonance curves corresponding to different degrees of damping. For the lowest, and broadest, of this set of curves, which represents the response of a quite heavily damped oscillator, the concept of width is not applicable at all.

An alternative way of describing the sharpness of the resonance peak is to use the *Q-factor* (originally this meant quality-factor). The Q-factor is defined as the ratio

$$\frac{\text{resonant frequency}}{\text{width of resonance peak}}$$

or, in symbols,

$$Q = \frac{\omega_0}{2\,\Delta\omega} \qquad (20.1)$$

hence

$$Q = \frac{\omega_0}{2\Lambda}, \qquad (20.2)$$

in terms of the decay constant Λ. The resonance curves of Fig. 20.12 are labelled with their appropriate Q-factors.

This Q-factor can also be shown to relate the amplitude at resonance to the amplitude of the responder at very low frequencies. Figure 20.12 shows the amplitude function, which is given in Appendix 5, plotted against driver frequency for different values of Q-factor.

The Q-factor is a measure of the *selectivity* of the resonant system in that it compares the response of the system to being driven at its resonant frequency with its response to being driven at any frequency appreciably different from the resonant frequency. The concept of Q-factor is important in the tuned circuit of a radio receiver where it is used as a measure of the selectivity. Selectivity in a radio receiver means the ability of the receiver to separate a radio signal having the desired frequency from any other signals having different frequencies. In a good receiver the response of the tuned circuit to the unwanted signals should be so small that they are quite inaudible. In a medium-wave radio receiver the Q-factor of the tuned circuit is typically in the order of a few hundred. In ultra-high frequency (UHF) receivers of the type used in certain radars, resonant cavities are used, and these may have Q-factors of 30 000 or more.

Let us now consider the energy of a resonant system. If the responder is initially at rest and then a driving agency begins to act upon it at the resonant frequency, the amplitude of oscillation steadily grows. During this process energy is being "fed" from the driver to the responder to increase its amplitude: but one must not forget that, because damping is present, the responder is losing some of the energy fed to it, this lost energy being used to overcome damping. A steady state is reached when the responder oscillates with constant amplitude. The responder now possesses constant total energy, but the driver is still supplying energy to overcome damping, and this energy is being supplied at a steady average rate.

The total energy of a simple harmonic oscillator of mass m, angular frequency ω, and amplitude a_0, is $\frac{1}{2}ma_0^2\omega^2$, as was shown in Section 18.5. This expression gives the stored energy of a responder in the steady state, being driven at an angular frequency ω. Let us now derive an expression for the rate at which energy is being used to overcome damping when a responder is being driven at frequency ω.

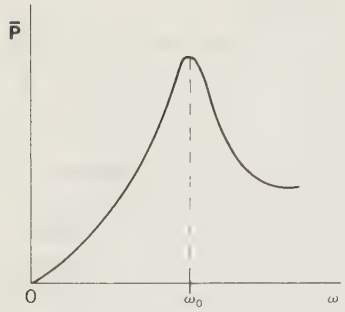


Fig. 20.13

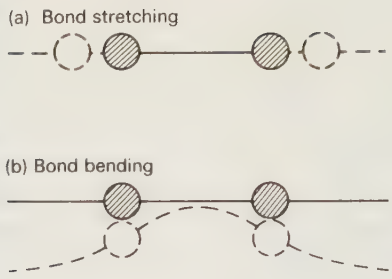


Fig. 20.14

Let the displacement of the responder at time t be given by

$$x = a_0 \cos \omega t$$

its velocity at time t is therefore given by

$$\frac{dx}{dt} = -a_0\omega \sin \omega t.$$

The viscous damping force, which is directly proportional to the velocity, is (see Appendix 5)

$$2m\Lambda \frac{dx}{dt},$$

and the rate P at which work is done against this force is the product of the force and the velocity, and is thus given by the expression

$$P = \left(2m\Lambda \frac{dx}{dt}\right)\frac{dx}{dt}$$

$$= 2m\Lambda \left(\frac{dx}{dt}\right)^2$$

$$= 2m\Lambda a_0^2\omega^2 \sin^2 \omega t.$$

As has been demonstrated earlier, the average value of $\sin^2 \omega t$ over any whole number of cycles is $\frac{1}{2}$: so the average rate of working ($\bar{P}$) against the viscous damping force is given by

$$\bar{P} = m\Lambda a_0^2\omega^2$$

$$= 2\Lambda \text{ (stored energy)}. \quad (20.3)$$

If the driving agency were suddenly "switched off", the oscillations would begin to decay according to the rule

$$\text{(average rate of energy loss)} = 2\Lambda \text{ (present stored energy)}$$

which, as has been discussed in Section 19.1 earlier, implies an exponential decay. If E_0 is the total energy of the oscillator at the instant when the driving agency stops acting, and E_t is the energy at a time t later, then

$$E_t = E_0 e^{-2\Lambda t}.$$

Since the energy is proportional to the square of the amplitude of oscillation, the amplitude decays according to the rule

$$a_t = a_0 e^{-\Lambda t}$$

as discussed in Section 19.1.

We consider now the question of the rate of energy absorption (or power absorption) of an oscillatory responder at different driving frequencies. Expression (20.3) gives the rate at which energy is fed from driver to responder, and we call this the rate of energy absorption by the responder.

Figure 20.13 shows how this rate of energy absorption by the responder varies with the driver frequency, ω. It should be noted that, although this curve has a shape similar to that of the resonance curve of Fig. 20.12 in which the amplitude tends to a *finite* limit as the driving frequency ω tends to zero, here the value of $\bar{P}$ tends to *zero* as the driving frequency ω tends to zero. Again, this is obvious, if you think about it. But the width of the resonance

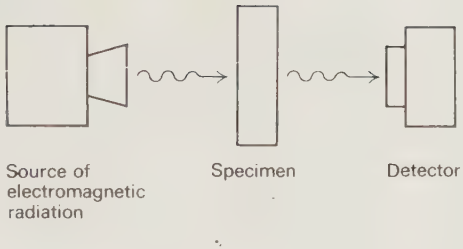


Fig. 20.15

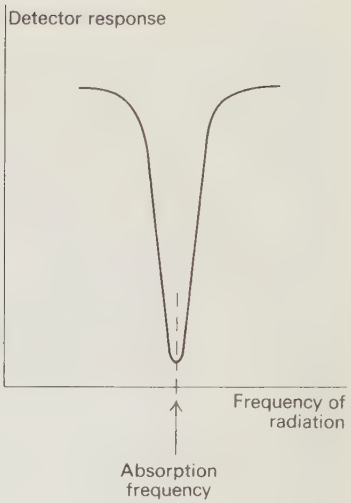


Fig. 20.16

peak is defined in a way exactly similar to the way depicted in Fig. 20.11, as the difference in frequency between the two points where $\bar{P}$ has half its resonant value. And, as before, it can be shown that for a not too heavily damped system, this width is equal to 2Λ. This width can be termed the *absorption bandwidth* of the oscillator, because it is a measure of the range or *band* of frequencies over which the oscillator is responsible for appreciable absorption of energy from whatever is driving it.

On the molecular scale of size the phenomenon of resonance is well known to the scientist who uses the technique known as *infrared spectroscopy*. When a beam of high-frequency electromagnetic radiation—infrared radiation—passes through a sample of a substance which is in the molecular state it is found that the sample absorbs energy from the radiation very strongly at certain precise frequencies. By measuring the frequencies at which the strong absorptions occur the spectroscopist can identify the types of chemical bond between atoms in the molecule. In this way infrared spectroscopy is a powerful tool for use in determining the structures of substances in the molecular state.

A molecule is built of atoms bonded together by electrical forces which behave to some extent like springy elastic links between the atoms. A pair of atoms can vibrate in such a way that the bond between them alternately stretches and shortens, as shown schematically in Fig. 20.14(a). (See also Unit 4 for potential energy/distance graphs.) The bond between a pair of atoms can bend back and forth, as shown in Fig. 20.14(b): the situation here is analogous to a pair of lead balls fixed to either end of a springy steel rod. With each of the two types of distortion described there is a vibratory motion and the characteristic frequency of the vibration depends upon the masses of the atoms at either end of the bond and upon the stiffness of the bond. The two types of motion described are not the only kinds with which a precise, characteristic frequency is associated but those two must serve to illustrate the principle in this extremely brief account.

The exact manner in which the electromagnetic radiation interacts with the molecules can only be satisfactorily described with the help of quantum theory. The basic principle of an absorption spectrometer (not merely an infrared spectrometer, which works in the infrared region of the electromagnetic spectrum) is shown in Fig. 20.15. We have already discussed how the rate at which energy is absorbed by an oscillator depends upon the driving frequency. The rate of energy absorption is a maximum at resonance. This means that, in the system shown in

Fig. 20.15, the energy of the electromagnetic wave reaching the detector will vary with the frequency of the source, and at resonance it will be a minimum because the energy flux reaching the detector is a minimum. A graph of detector response plotted against the frequency of the radiation will look like the curve shown in Fig. 20.16: this would be a very simple example of an absorption spectrum.

It is not possible in this book to discuss in detail any further examples of resonance. In concluding this section we briefly mention some other examples of importance in the study of the structure of matter. Subatomic particles which have electric charge (protons and electrons for instance) also have spin and thus they behave like small magnetic dipoles. Now a magnetic dipole which is free to move, when placed in a steady magnetic field, is an oscillatory system (compare a pivoted compass needle, how it flutters, when placed between the poles of a magnet), and it has a characteristic frequency of free oscillation. The particle can be forced into oscillation by means of a periodically varying magnetic field: for example, the magnetic field of an electromagnetic wave. If the frequency of the wave is the same as the natural frequency of the particle in the steady field, then the particle will be excited into resonance. If a very large number of particles are thus excited at the same frequency (there will be several billions of such particles in a milligram of matter), they may absorb energy at an appreciable rate from the electromagnetic radiation which falls upon them, and an absorption spectrum may be observed. *Electron spin resonance* is one example of this type of resonance, and *nuclear magnetic resonance* is another. Study of these phenomena can yield useful information about the subatomic structure of the substances being studied.

Chapter 21
COUPLED OSCILLATORS AND THE TRANSMISSION OF ENERGY

21.1 Coupled oscillators
21.2 A line of coupled oscillators: propagation of a longitudinal vibration
21.3 Wave machines, and the mathematical description of travelling sinusoidal waves
21.4 A string of particles: propagation of a transverse vibration

21.1 COUPLED OSCILLATORS

So far we have considered in detail, oscillatory systems which consist only of a single mass and a single agency providing a restoring force. Oscillatory systems in real life are rarely as simple as this: indeed a microscopic speck of any solid matter contains billions of oscillating masses, namely, all the atoms it contains! If all the atoms oscillated independently of each other, that is to say, if the movement of one atom did not have any effect on the forces acting upon any of the other atoms, then the situation might not be difficult to analyse. But atoms in a solid, and most complex oscillatory systems for that matter, consist of *coupled* oscillators. Oscillators are said to be coupled if the motion of one of the oscillating masses alters the forces acting upon any of the other oscillating masses. The principles of analysis of such systems can be shown by consideration of a few very elementary examples.

A simple laboratory experiment shows well the properties of a pair of identical coupled oscillators. Two similar trolleys, A and B, of equal mass, are attached by springs to each other, as shown in Fig. 21.1, and to rigid supports. If trolley B is held firmly while A is displaced and let go, then A oscillates in the familiar manner: likewise if A is held firmly and B displaced and let go, it will oscillate with the same frequency. If both A and B are free to move, do we regard A as the driver and B as the responder when they are in motion, or vice versa? How do we expect the system to oscillate? The situation is more complex than one might imagine. However, there are two quite distinct modes of oscillation which can be produced by this system. In one mode A and B move in phase with each other, moving as if A and B and the spring joining them were one single rigid object. This can be demonstrated by displacing A and B the same distance to one side and then letting them go at precisely the same instant. In the other mode of oscillation A and B move with opposite phase. This mode can be excited (*excited* meaning *set into motion*) by displacing A to the left and B to the right by equal amounts and releasing both simultaneously. In this mode the centre of mass of

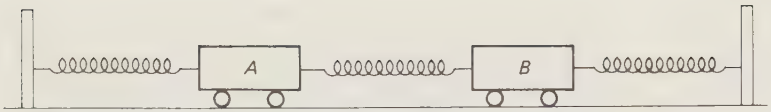


Fig. 21.1

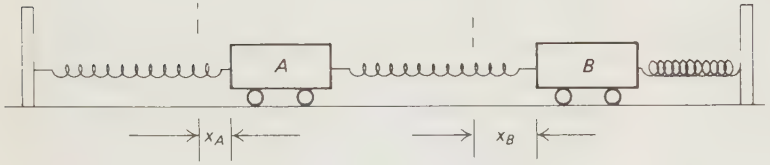


Fig. 21.2

the system does not shift its position. It is found that the common frequency of oscillation in the out-of-phase mode is greater than the frequency of the in-phase mode. These frequencies can be calculated as follows.

Suppose the springs are all of the same type and the force per unit displacement provided by them is k. We consider the forces acting upon A and B separately, when A has displacement x_A and B has displacement x_B (Fig. 21.2). The spring to the left of A has been lengthened by an amount x_A, and so the force due to this lengthening is $-kx_A$ (note the minus sign: this is because the force is in the negative x-direction). The spring between A and B has increased its length by an amount $(x_B - x_A)$, and the force on A due to this lengthening, which acts to the right, is $k(x_B - x_A)$. Thus the net force on A is given by

$$F_A = k(x_B - x_A) - kx_A = k(x_B - 2x_A).$$

The force on B due to the spring between A and B is $-k(x_B - x_A)$, and the force due to the spring on the right of B, which has shortened by an amount x_B, is $-kx_B$. Thus the net force on B is given by

$$F_B = -k(x_B - x_A) - kx_B = k(x_A - 2x_B).$$

We can now write the equations of motion for A and B, applying Newton's second law of motion, thus:

$$\text{for mass } A: \quad m\frac{d^2x_A}{dt^2} = k(x_B - 2x_A) \qquad (21.1)$$

$$\text{for mass } B: \quad m\frac{d^2x_B}{dt^2} = k(x_A - 2x_B). \qquad (21.2)$$

Our preliminary experimental investigation has shown that, in the two basic modes of oscillation, A oscillates with the same frequency as B and with the same amplitude. Let us assume, therefore, that the motions of A and B can be represented by expressions of the form

$$x_A = a_0 \cos \omega t,$$

$$x_B = a_0 \cos (\omega t + \phi),$$

where ϕ is the phase-difference between A and B. If we substitute these expressions for x_A and x_B into Eq. (21.2) we get:

$$-ma_0\omega^2 \cos (\omega t + \phi) = k[a_0 \cos \omega t - 2a_0 \cos (\omega t + \phi)],$$

and expanding the terms in $(\omega t + \phi)$ yields, after separating the terms in $\cos \omega t$ from the terms in $\sin \omega t$:

$$(-ma_0\omega^2 \cos \phi - ka_0 + 2ka_0 \cos \phi) \cos \omega t = (2ka_0 \sin \phi - ma_0\omega^2 \sin \phi) \sin \omega t.$$

Now the only way in which this equation can be satisfied, as it must be, for *any value* of t whatsoever, is for both sides of the equation to be equal to zero. Since neither $\cos \omega t$ or $\sin \omega t$ are equal to zero them-

selves, for all values of t, the expressions in brackets must be equal to zero. Hence we can write

$$-m\omega^2 \cos \phi - k + 2k \cos \phi = 0 \quad (21.3)$$

and

$$2k - m\omega^2 = 0.$$

If we carry out a similar process with Eq. (21.1), substituting in it the expressions for x_A and x_B, we finally obtain:

$$-m\omega^2 = k \cos \phi - 2k \quad (21.4)$$

and

$$0 = -k \sin \phi. \quad (21.5)$$

This last equation is satisfied by $\phi = 0$ and $\phi = \pi$. Putting $\phi = 0$ into (21.4) we get

$$-m\omega^2 = k - 2k$$

hence

$$\omega^2 = \frac{k}{m}$$

$\phi = \pi$ also satisfies Eq. (21.4), and this solution defines the out-of-phase mode (two oscillators with a phase-difference of π radians between them are oscillating with opposite phase). Substituting $\phi = \pi$ in Eq. (21.4) yields the result

$$\omega^2 = 3\frac{k}{m}.$$

Using the relation $T = 2\pi/\omega$ we can summarize the results as in Table 21.1.

The example we have treated is the simplest of all—the springs are all of the same stiffness and the masses are equal. A similar mathematical analysis could have been applied to a case in which the springs had unequal stiffness and the masses were unequal. It would then have been found that, once again, there are two distinct modes of oscillation, each with its characteristic frequency. A system comprising three equal masses connected by springs (or interacting by some other means which provides spring-like restoring forces) has three normal modes of vibration. These are illustrated in Fig. 21.3, the arrows indicating the relative motions of the masses.

Table 21.1

Mode of oscillation	Phase difference	Period
In-phase	0	$2\pi\sqrt{\frac{m}{k}}$
Out-of-phase	π	$2\pi\sqrt{\frac{m}{3k}}$

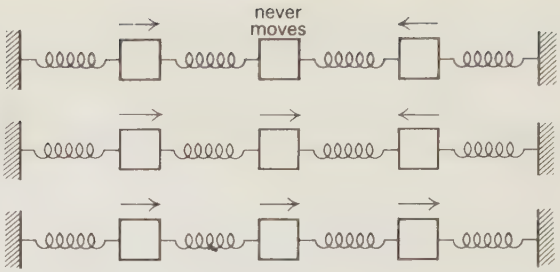


Fig. 21.3

By *normal modes* of oscillation we mean that each mass oscillates with simple harmonic motion of constant amplitude, so that its displacement is a sinusoidal function of time. You should not be misled into thinking that these modes are the only possible kinds of periodic motion possible in a system such as that of Fig. 21.3: an infinite variety of complex modes is possible. The particular mode of vibration depends upon precisely how the system was excited into vibration in the first place. For example, to excite the mode shown in the top sketch of Fig. 21.3 we would have to displace the two outer masses by equal and opposite amounts, and then let go. To excite the second mode would be more difficult: the two left-hand masses would have to be displaced by the appropriate different amounts one way and the right-hand mass by the appropriate amount the other way, and then all masses let go simultaneously. Any complex mode of vibration can be considered to be the result of combining the motions of two or more normal modes. This idea is developed a little further in the discussion of the modes of vibration of a stretched string (Section 23.2).

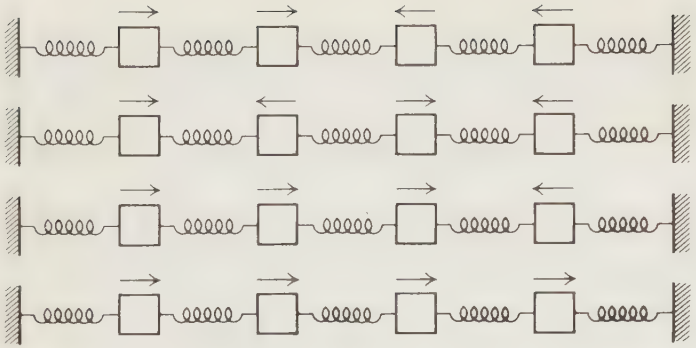


Fig. 21.4

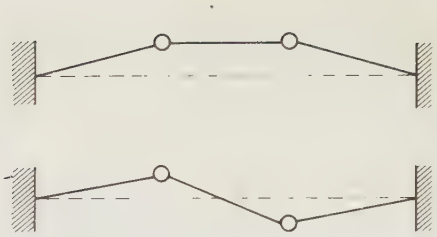




Fig. 21.5

The point of pursuing the study of coupled oscillators as far as this is to give you a basic conception of how a system of particles coupled together by elastic forces can be expected to behave. The particles might be ions in a crystal lattice, or individual atoms within a molecule. An important branch of solid-state physics concerns the ways in which atomic particles can vibrate when coupled to each other by forces, because it is these vibrations which are involved when mechanical impulses are transmitted through a solid material (for example, sound waves). Also, the energy associated with the vibrations is fundamental in determining the thermal properties of solids. Later in this chapter we consider the oscillatory behaviour of an infinite chain of identical coupled oscillators, for this enables us to analyse the way in which sound travels through solids.

Let us return briefly to a small number of coupled oscillators. Figure 21.4 shows the four possible normal modes of vibration when four equal masses are linked by identical springs. This suggests that with n equal masses linked by identical springs there would be n normal modes of oscillation. Although we cannot prove this here, this is in fact true.

The masses on springs which we have been describing have, we supposed, been performing only longitudinal oscillations, that is to say, with the vibrations in the same direction as the axis of the springs. But transverse vibrations are possible. Figure 21.5 shows the two transverse modes of oscillation of a two-mass system (the springs are represented here by straight lines). The three normal modes of a three-mass system are shown in the lower part of Fig. 21.5.

21.2 A LINE OF COUPLED OSCILLATORS: PROPAGATION OF A LONGITUDINAL VIBRATION

A simple experiment can be performed in the laboratory to investigate the behaviour of a large number of coupled oscillators, each oscillator having the same characteristics. A number of dynamics trolleys are linked together by springs, as in Fig. 21.6. The two end springs are tethered to rigid supports so that all springs are in tension. If trolley A is displaced

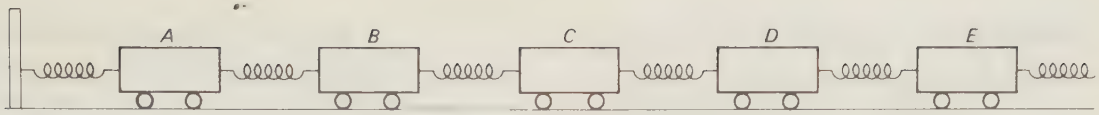


Fig. 21.6

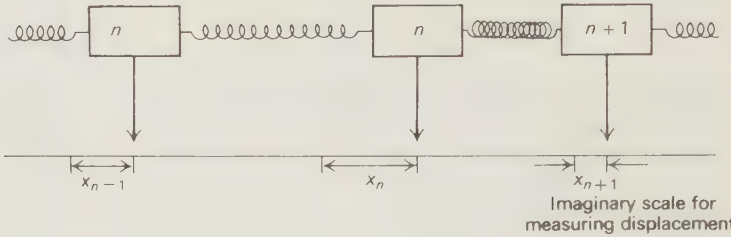


Fig. 21.7

suddenly a certain distance to the right and then held firm in its new position, it is observed that B moves to the right, then C, then D, and so on, the time-lag between the motions of successive trolleys being constant. A disturbance or *pulse* is thus observed to travel from left to right along the line of oscillators at a steady speed. If we analyse the dynamics of this system we should be able to find out what factors determine the speed of this pulse. If we can do this we shall be able to extend the idea to an understanding of how a compression wave travels through a continuous elastic medium.

You must not be misled into thinking that the way B moves is exactly the same as the way A was moved by the experimenter. The force causing the displacement of A was provided by the experimenter's hand, whereas the force acting on B to displace it was provided by the springs coupling it to A and C. Does the motion of C, however, exactly reproduce the motion of B? We cannot be certain of this, and rather than trying to decide the answer to this question we shall avoid it by supposing that we have a line of trolleys so long, that we have lost sight, as it were, of the place where the pulse was started, and that the motion of each trolley does exactly reproduce the motion of the trolley before it as the pulse travels along the line. Also we shall assume that (a) the disturbance has the form of a continuous oscillatory motion, with frequency f, and (b) there are no energy losses.

Now we shall carry out a mathematical analysis which will yield an expression for the speed at which the disturbance travels along the line. We consider the forces acting upon the nth trolley, which at the instant we are considering, has a displacement x_n from its equilibrium position (Fig. 21.7). The displacement at the same instant of the $(n-1)$th trolley is x_{n-1} and that of the $(n+1)$th trolley is x_{n+1}. Proceeding as in the analysis of two coupled oscillators we can write the equation for the nth trolley thus:

$$m\frac{\mathrm{d}^2x_n}{\mathrm{d}t^2} = -k[(x_n - x_{n-1}) + (x_n - x_{n+1})]$$

$$= -k(2x_n - x_{n-1} - x_{n+1}). \qquad (21.6)$$

We will assume that there is a phase-difference of ϕ between the motion of adjacent trolleys. Thus if we write, for the motion of the nth trolley,

$$x_n = a_0 \cos \omega t$$

then for the $(n-1)$th trolley

$$x_{n-1} = a_0 \cos (\omega t - \phi)$$

and for the $(n+1)$th trolley

$$x_{n+1} = a_0 \cos (\omega t + \phi).$$

Substituting these expressions in (21.6) we get

$$-ma_0\omega^2 \cos \omega t = -k[2a_0 \cos \omega t - a_0 \cos (\omega t - \phi) - a_0 \cos (\omega t + \phi)];$$

expanding, and dividing by $a_0 \cos \omega t$, we get

$$m\omega^2 = 2k - 2k \cos \phi$$

therefore $1 - \cos \phi = m\omega^2/2k$ which can be written alternatively

$$2 \sin^2 \frac{\phi}{2} = \frac{m\omega^2}{2k}$$

hence

$$\sin \frac{\phi}{2} = \frac{1}{2} \omega \sqrt{\frac{m}{k}}.$$

If we have so many trolleys, and the speed of the pulse is such, that ϕ is very small, we can make the approximation

$$\sin \frac{\phi}{2} = \frac{\phi}{2}$$

and then

$$\phi = \omega \sqrt{\frac{m}{k}}.$$

Now $\omega = 2\pi f$, where f is the frequency of oscillation of each trolley, so the time required for the phase of oscillation of one trolley to change by an amount ϕ is given by

$$t = \frac{\phi}{\omega} = \sqrt{\frac{m}{k}}.$$

After this time interval the $(n + 1)$th trolley has the displacement which the nth trolley had, the $(n + 2)$th trolley has the displacement which the $(n + 1)$th had, and so on. This is to say that the pattern of the disturbance has travelled onwards by a distance equal to the mean separation of the trolleys. We will call this distance s. Thus we can write, for the speed at which the disturbance travels along the line of trolleys:

$$v = \frac{s}{t} = s \sqrt{\frac{k}{m}}. \qquad (21.7)$$

Suppose that the coupling between the masses is provided by straight elastic wires, or rods, of cross-sectional area A and length s: we are now assuming that each mass is concentrated into a space which is small compared with the distance s. Then the force per unit displacement of the couplings is given by

$$k = \frac{EA}{s},$$

where E is Young's modulus of elasticity for the material of which the couplings are made. Also, one can describe the way in which the mass is distributed along the chain by referring to the mass per unit length, μ, where

$$\mu = \frac{m}{s},$$

therefore

$$m = \mu s.$$

Thus we can rewrite Eq. (21.7):

$$v = s \sqrt{\frac{EA/s}{\mu s}} = \sqrt{\frac{EA}{\mu}}.$$

Now we can apply this result to the situation in which the mass is uniformly distributed along the chain with a density ρ, so that

$$\mu = \rho A.$$

Then

$$v = \sqrt{\frac{EA}{\rho A}} = \sqrt{\frac{E}{\rho}}, \qquad (21.8)$$

which is the expression for the speed at which a compression wave travels along a rod or wire of elastic material.

Let us use the expression we have just derived to find the order of magnitude of the speed with which a longitudinal vibration might be expected to travel along a bar of solid rock. This will enable us to get some appreciation of the speed with which the tremors generated by earthquakes travel. The value of E, Young's modulus of elasticity, depends on the kind of rock, varying from about 2×10^{10} N m^{-2} for sandstone to about 8×10^{10} N m^{-2} for some kinds of granite. Thus, for a bar of hard granite whose

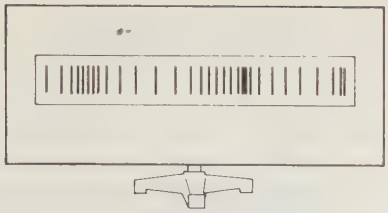


Fig. 21.8

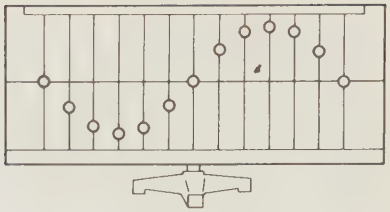


Fig. 21.9

density is about 3×10^3 kg m^{-3}, the speed of a compression wave will be given by

$$v = \sqrt{\frac{8 \times 10^{10}}{3 \times 10^3}} \text{ m s}^{-1} = 5000 \text{ m s}^{-1}$$

approximately, which is about 3 miles per second.

The waves caused by an earthquake are called *seismic* waves. They are not only compression waves: other kinds of wave-motion (for instance shear waves, and surface waves) are propagated through the earth at different characteristic speeds. Instruments, called seismographs, which detect the arrival of tremors from distant earthquakes can distinguish the tremors due to the different kinds of wave-motion, and analysis of the seismograph record can yield information about the interior structure of the earth. The compression waves going through the earth travel at speeds of about 3.7 miles per second near the surface, but about 6.5 miles per second at great depths, this increase in speed being due to the enormous increase in pressure at these depths. Such waves take about twenty minutes, therefore, to traverse the earth.

21.3 WAVE MACHINES, AND THE MATHEMATICAL DESCRIPTION OF TRAVELLING SINUSOIDAL WAVES

The mathematical argument of the previous section, which followed logically the discussions of the behaviour of coupled oscillators, has enabled us to predict the speed at which a travelling longitudinal, sinusoidal vibration moves along a line of identical coupled oscillators. You may find the process difficult to picture in the mind's eye: a wave machine of the kind shown in Fig. 21.8 shows the pattern of longitudinal wave motion very well. A number of brightly-painted rods represent the oscillators. They are driven by cranks linked to an axle which can be rotated by hand. The rods move back and forth, from left to right, with an approximately sinusoidal motion—the motion of each rod being out of phase with that of its neighbours in such a way that a longitudinal wave-pattern is seen to travel along horizontally. The wave-motion is of course artificial in the sense that the rods are not actually free oscillators like the trolleys of Fig. 21.6.

Another kind of wave machine, shown in Fig. 21.9, simulates a transverse wave. When the handle of the machine is turned at a steady speed the brightly-painted balls oscillate up and down in such a way that a sinusoidal wave-pattern appears to travel along horizontally at a steady speed. With reference to this model we can set up the equations which describe a travelling, transverse, sinusoidal wave. The particular feature of the model which concerns us, as it did in the discussion in the previous section, is the phase-difference between the motions of adjacent particles. Consider the motion of particle P in Fig. 21.10. Imagine that we start a clock at the instant when this particle has its maximum positive displacement: then we can write, for its displacement at any subsequent time t,

$$y = a_0 \cos \omega t,$$

where a_0 is the amplitude of its motion and $\omega = 2\pi f$, f being the frequency.

Now let us see how the motion of another particle, Q, can be represented. The wave-profile is moving from left to right with speed v, and so Q reaches its maximum positive displacement at a time s/v after P does, where s is the horizontal spacing of the particles.

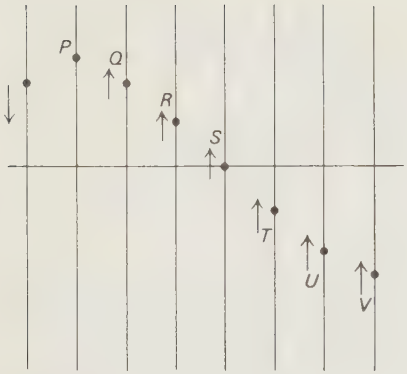


Fig. 21.10

It follows that the phase of Q lags behind the phase of P by an amount $\omega(s/v)$. Similarly, the phase of particle R lags behind that of Q by an amount $\omega(s/v)$, and thus lags behind that of P by $2\omega(s/v)$. Thus the motion of the particles is described by the set of equations

$$y_P = a_0 \cos \omega t$$

$$y_Q = a_0 \cos \omega \left(t - \frac{s}{v}\right)$$

$$y_R = a_0 \cos \omega \left(t - \frac{2s}{v}\right) \text{ etc.}$$

If we have a continuum of particles (that is to say, a particle can have any value of distance from P horizontally, not only a whole number multiple of s), then the motion of a particle whose horizontal distance from P is x is described by the expression

$$y = a_0 \cos \omega \left(t - \frac{x}{v}\right). \qquad (21.9)$$

Because this expression describes the motion of any particle in the medium through which the wave is travelling, we can say that this expression describes the wave itself.

If the wave is travelling from right to left, then particle Q will have a phase-*lead* (instead of a phase-lag) over P, and P will have a phase-lead over R, and so forth. The result of this is that the expression for this wave will be

$$y = a_0 \cos \omega \left(t + \frac{x}{v}\right). \qquad (21.10$$

The expressions we have derived describe pure sinusoidal waves which suffer no loss of amplitude as they travel along: in this sense they are "ideal" waves.

The relationship between the frequency, wavelength, and the speed of the wave-profile, for the original wave going from left to right, can be found as follows. The wavelength can be defined as the distance along the x-axis between successive particles which are moving with the same phase. We can write the expressions for the displacements of two such particles distant x_1 and x_2 from the origin,

$$y_1 = a_0 \cos \omega \left(t - \frac{x_1}{v}\right)$$

$$y_2 = a_0 \cos \omega \left(t - \frac{x_2}{v}\right).$$

But particle no. 2 has a phase-lag of precisely 2π radians behind particle no. 1, and since $y_1 = y_2$ at every instant, we can write

$$\omega \left(t - \frac{x_1}{v}\right) - 2\pi = \omega \left(t - \frac{x_2}{v}\right).$$

Then, writing $\omega = 2\pi f$, we simplify this to obtain

$$f\left(\frac{x_2 - x_1}{v}\right) = 1.$$

But we have already defined the wavelength, λ, as equal to $x_2 - x_1$, so it follows that

$$f\lambda = v. \qquad (21.11)$$

There is, as you probably know already, a much simpler verbal argument which leads to the same result. Imagine yourself to be standing still at a certain place with the wave travelling past you. The distance between successive wave-crests is the wavelength, λ. If f is the frequency, then the number of wave-crests which pass you in unit time is f. Hence the speed with which the wave is travelling, that is,

the distance travelled per unit time, must be the product of f and λ.

The expressions derived above for the displacement of particles of the wave-medium are not only applicable to transverse sinusoidal waves. They can just as well be used to describe the displacements of particles in a medium through which a longitudinal sinusoidal wave is travelling: a wave-motion such as the one discussed in Section 21.2.

There we analysed a longitudinal wave-motion dynamically, getting an expression for the speed of a sinusoidal wave in terms of the masses of the particles and the force per unit displacement provided by the coupling springs. In Section 21.4 we shall carry out a similar analysis of a medium along which a transverse sinusoidal wave is travelling, and shall obtain an expression for the speed of this wave. Perhaps it should be emphasized that, although the motion of the particles in the wave-machine of Fig. 21.9 exactly simulates the motion of the particles in the wave-medium to be discussed in Section 21.4, the dynamic situation responsible for their motion in each case is entirely different.

What use is there for the algebraic expressions we have just derived for the displacements of particles in a medium through which a sinusoidal wave is travelling? In this book there is space for only one particular application of the expressions, and this will be found in Section 22.1, where the expressions are used to explain how a stationary wave pattern is produced by the superposition of two travelling sinusoidal waves going in opposite directions.

21.4 A STRING OF PARTICLES: PROPAGATION OF A TRANSVERSE VIBRATION

Our aim now is to find out precisely how transverse waves travel along a stretched string. We use the word "string" here to mean a length of uniform solid material in a state of tension: thus, a stretched wire, or rope, or thin rod, or even a long helical spring, can be regarded as a string. Our string here is assumed to have no rigidity, that is to say, the only transverse

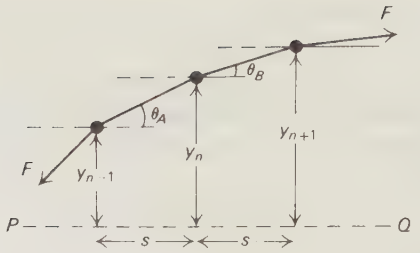


Fig. 21.11

forces on the particles of the string arise from the tension in the string when stretched. We know that all these things can be used to demonstrate wave motion: if we displace part of the string sideways (that is, if we make a *transverse* displacement) and let go, a restoring force acts tending to bring the string back to its equilibrium position. But, in general, the string overshoots the equilibrium position, and, also in general, the disturbance we initiated travels along the string. We would like to know, first of all, what determines the speed at which such a disturbance travels.

We begin by inventing a model of a stretched string which can be easily analysed. Our model consists of an infinitely long *mass-less* string in which the tension is F. Attached to it, and equally spaced at distances s from each other, are identical point-masses m, like beads threaded on the string. We shall assume that the transverse displacements of the particles are so small that any change in the tension can be neglected. Figure 21.11 shows a small section of the string with the particles displaced. The displacements are supposed to be very small in comparison with the spacing, s, but we have exaggerated the displacements enormously for the sake of clarity.

PQ is the line of the undisturbed string. We imagine that the particles are numbered, and we are looking at the $(n-1)$th, nth, and $(n+1)$th particles only. We shall consider the forces acting on the nth particle and the motion caused by these forces, using Newton's second law of motion: force = mass × acceleration.

The resultant force on the nth particle is given by

$$F_R = -F \sin \theta_A + F \sin \theta_B.$$

We assume that the angles θ_A and θ_B are small enough to justify the approximation: $\sin \theta \simeq \tan \theta$. Then we can write

$$\sin \theta_A = \frac{y_n - y_{n-1}}{s}$$

and

$$\sin \theta_B = \frac{y_{n+1} - y_n}{s}.$$

Hence we can write, for the resultant force,

$$F_R = \frac{F}{s}(y_{n-1} - 2y_n + y_{n+1}). \qquad (21.12)$$

The equation of motion of the nth particle is therefore

$$\frac{F}{s}(y_{n-1} - 2y_n + y_{n+1}) = m\frac{\mathrm{d}^2 y_n}{\mathrm{d}t^2}.$$

Let us assume that all the particles are already in a steady state of oscillation, with amplitude a_0 and frequency f, and that we have a wave-pattern travelling along from left to right at a steady speed c. As we established earlier in our discussion of the wave-machine (Fig. 21.8), the travelling wave-pattern arises from the fact that there is a phase-difference between the motion of adjacent oscillating particles. Calling this phase-difference ϕ, we can write for the motion of our three particles:

$$y_{n-1} = a_0 \cos(\omega t + \phi)$$

$$y_n = a_0 \cos \omega t$$

$$y_{n+1} = a_0 \cos(\omega t - \phi).$$

(This assumes that we start the clock which measures t at an instant when the nth particle has maximum positive displacement.)

We substitute these expressions in (21.12) and get

$$\frac{F}{s}[a_0 \cos(\omega t + \phi) - 2a_0 \cos \omega t + a_0 \cos(\omega t - \phi)] = -ma_0\omega^2 \cos \omega t.$$

Expanding the cosine terms, and cancelling a_0's, we get

$$\frac{F}{s}[\cos \omega t \cos \phi - \sin \omega t \sin \phi - 2 \cos \omega t + \cos \omega t \cos \phi + \sin \omega t \sin \phi] = -m\omega^2 \cos \omega t$$

so that

$$\frac{2F}{s}(1 - \cos \phi) = m\omega^2.$$

Since

$$1 - \cos \phi = 2 \sin^2 \frac{\phi}{2},$$

this becomes

$$\frac{4F}{s} \sin^2 \frac{\phi}{2} = m\omega^2.$$

Let us now assume that the spacing of the particles is very small compared with the wavelength of the waves: this will mean that ϕ is very small, and we can make the approximation

$$\sin \frac{\phi}{2} \simeq \frac{\phi}{2}.$$

Then we can write

$$\frac{F}{s}\phi^2 = m\omega^2$$

therefore

$$\phi = \omega\sqrt{\frac{ms}{F}}. \qquad (21.13)$$

We wanted originally to find the speed of the wave-pattern. To do this, we need to find the time (call it Δt) for the $(n + 1)$th particle to attain the same displacement as the nth particle: this will of course be the time taken for the phase to change by an amount ϕ, so

$$\Delta t = \frac{\phi}{\omega}.$$

And it is also the time taken for the wave-pattern to travel a distance s, so that the speed (v) of the wave-pattern

$$v = \frac{s}{\Delta t} = s\frac{\omega}{\phi}$$

combining this with (21.13) we get

$$v = \sqrt{\frac{Fs}{m}}.$$

Now, finally, we suppose that the particles are so close together that we can regard the mass as being continuously distributed along the string, so that we can specify the *mass per unit length*, μ. Then $\mu = m/s$ and hence

$$v = \sqrt{\frac{F}{\mu}}. \qquad (21.14)$$

Thus we have found that the speed of a periodic wave disturbance along a stretched string depends on the tension and mass per unit length, and does not depend on the frequency of the wave oscillation (provided, of course, that the frequency is not so great that the phase difference, ϕ, between adjacent particles becomes large enough to invalidate the approximation $\sin \phi/2 = \phi/2$ which we made earlier). It can be shown, by pursuing this piece of theoretical work further, that if the frequency *is* large enough to invalidate the approximation $\sin \phi/2 = \phi/2$, it is nevertheless still possible for a wave disturbance to be propagated along the string. Under these conditions, the speed of propagation is dependent on the frequency of oscillations. There is an upper limit to the frequency: oscillations having a higher frequency than this cannot be propagated along the string at all. The limiting frequency can be shown to be equal to

$$\frac{1}{\pi}\sqrt{\frac{F}{ms}},$$

where s is the spacing of the particles. At this frequency adjacent particles are oscillating in antiphase.

Now we may wonder, does the expression (21.14) work in practice? Let us put in numerical values appropriate to (a) a slinky spring stretched along the floor, (b) a piece of string stretched between your hands.

Example

a) If you weigh a "slinky" spring you may find (as we did) that it had a mass of about half a kilogramme. Stretch it out to a length of about 5 metres on the floor. Use a spring-balance to pull it out to this length, and you will find that the constant force needed is in the order of one newton. Hence the mass per unit length is about 0.1 kg m^{-1}, and we would expect the speed of a wave disturbance to be

$$\sqrt{\frac{1}{0.1}} = \sqrt{10} \simeq 3 \text{ m s}^{-1}.$$

And indeed, it takes about one or two seconds for a wave-pulse to go the length of the spring.

b) A piece of string, about 1 mm thick, 1 m long, has a mass of about a gram. If you stretch this in your hands with a force of, say, one Newton and someone twangs it near one end, it appears that the whole string is immediately affected, and if there is a pulse travelling along from the point where it was twanged, then it is certainly too fast to follow. Substitution in the formula gives, for the speed,

$$\sqrt{\frac{1}{10^{-3}}} \simeq 30 \text{ m s}^{-1}$$

(approx. 60 m.p.h.).

Chapter 22
WAVES ON WATER

22.1 Waves on water

22.1 WAVES ON WATER

Let us now look at some real waves. Figure 22.1 shows a glass tank (about 1.5 m × 0.3 m × 0.2 m) for studying the profile of waves on water. If dust is put in the water, some dust particles will remain suspended in the body of the water long enough to reveal the motion of the water itself. You should study this motion closely: you will notice that, in general, the particles circulate, their paths being approximately elliptical, sometimes circular. It is clear from these observations that, in general, the water particles have not only a vertical component of oscillatory motion, but also a horizontal component. Obviously, the particle-motion is more complex than in our nicely-behaved wave model.

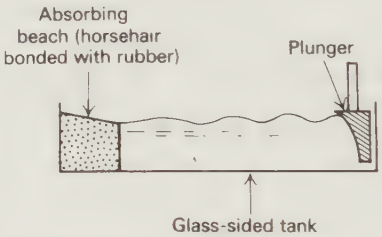


Fig. 22.1

If you are interested in a full analysis of water waves you will have to consult a more specialized treatise than this book (Barber, 1969).

Here we briefly quote the results of a theoretical analysis of water waves, and add some comments. The restoring forces which act upon the water when the surface is displaced from its horizontal equilibrium are partly gravitational, partly due to the surface tension of the water. If you are in a ship out at sea you can scarcely be unaware of the fact that water waves show enormous variety. At one extreme there is the large ocean swell in which the wavelength may be many metres long, even hundreds of metres; at the other extreme there are the tiny wrinkles on the water surface: small ripples whose wavelength may be only a few millimetres. In the ocean swell the restoring forces are almost entirely gravitational, and surface tension plays a negligible part. In the small ripples the restoring forces are very largely due to surface tension.

Table 22.1 Phase velocity of water waves: shallow water deep water

		Wavelength in metres									
		10^{-3}	10^{-2}	10^{-1}	1	10	10^2	10^3	10^4	10^5	10^6
Depth in metres	10^{-3}	0.67	0.18	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
	10^{-2}	0.67	0.25	0.30	0.31	0.31	0.31	0.31	0.31	0.31	0.31
	10^{-1}	0.67	0.25	0.40	0.93	0.99	0.99	0.99	0.99	0.99	0.99
	1	0.67	0.25	0.40	1.25	2.95	3.13	3.13	3.13	3.13	3.13
	10	0.67	0.25	0.40	1.25	3.95	0.33	9.91	9.91	9.91	9.91
	10^2	0.67	0.25	0.40	1.25	3.95	12.5	29.5	31.3	31.3	31.3
	10^3	0.67	0.25	0.40	1.25	3.95	12.5	39.5	93.3	99.1	99.1
	10^4	0.67	0.25	0.40	1.25	3.95	12.5	39.5	125	295	313
	10^5	0.67	0.25	0.40	1.25	3.95	12.5	39.5	125	395	991
Ripple-tank	5×10^{-3}	0.67	0.25	0.24	0.22	0.22	0.22				

The complete expression for the speed, V, of travelling sinusoidal waves in water is

$$V^2 = \left(\frac{g\lambda}{2\pi} + \frac{2\pi\gamma}{\rho\lambda}\right) \tanh\left(\frac{2\pi d}{\lambda}\right). \qquad (22.1)$$

where λ is the wavelength, ρ is the density of the water, γ is the surface tension, and d the water depth. Tanh is a trigonometrical function whose values can be found in standard tables. From this we can derive simpler, approximate expressions for particular classes of wave.

a) Deep water waves

These are defined as waves for which the wavelength is much smaller than the depth, but large enough to make surface tension a negligible factor. For such waves, since $\lambda \ll d$, therefore $2\pi d/\lambda \gg 1$, and so $\tanh(2\pi d/\lambda) \simeq 1$. Also it follows that $2\pi\gamma/\rho\lambda \ll g\lambda/2\pi$ and so the second term inside the brackets can be neglected, so that the expression reduces to $V^2 = \sqrt{g\lambda/2\pi}$. Thus we see that the velocity depends on the wavelength, longer waves travelling faster.

b) Shallow water waves

For these the wavelength is large compared with the depth, and large enough also to make surface tension a negligible factor. Thus $\lambda \gg d$, hence $2\pi d/\lambda \ll 1$, and so $\tanh 2\pi d/\lambda \simeq 2\pi d/\lambda$. And, as in (a), the second term inside the brackets of the original expression can be neglected. Thus the expression reduces to $V^2 = gd$. For this class of waves the velocity is independent of wavelength.

c) Surface ripples

For these, surface tension is the predominant factor and in this case the first term inside the brackets can be neglected. But, as in (a), $\lambda \ll d$ so we can make the approximation $\tanh 2\pi d/\lambda \simeq 1$. Thus the expression becomes: $V^2 = 2\pi\gamma/\rho\lambda$. Again, the velocity depends upon the wavelength, but in a different manner from (a). The shorter the wavelength of these ripples, the faster they go.

Table 22.1 shows the wave velocities for various depths of pure water and various wavelengths, in a

region where the strength of the gravitational field is 9.81 N kg^{-1}, as in Britain. We can refer to this figure in considering three particular types of wave which are of interest.

d) A tidal wave

This kind of wave has a very long wavelength, perhaps hundreds of metres. In a river whose depth is 10 m, as the table shows, the speed of such a wave would be about 10 m s^{-1} (approximately 20 mph).

e) A tsunami (pronounced "tsoo-nah-mee")

This ocean wave, caused by an earthquake, may have an effective wavelength of many kilometres. Taking a rough estimate of the average depth of the Pacific Ocean as 4000 m, the wavelength is much greater than this and once again we have shallow water waves. Table 22.1 shows that the speed should be in the range 100 to 300 m s^{-1} (200-600 mph approximately).

f) Ripples in a ripple-tank

Typically the depth of water is about 0.5 cm. A vibrator, running at moderate speed, may produce waves whose wavelength is in the order of 1 cm. The table shows that the speed should be around 0.2 m s^{-1}. This is slow enough to make the observation easy, and the frequency of vibration needed to produce waves of this length is not excessive ($f = v/\lambda \simeq 20/1 = 20$ Hz). Also (see following section), the problem of dispersion is not serious as can be seen from the extra line in the table, for water-depth 0.5 cm, the speed is very nearly constant for all waves whose wavelength is 1 cm or greater.

In our discussion of travelling waves so far we have talked confidently about the speed, or the velocity, of the waves. With waves which we can see, we can measure the speed by following the crest of the wave-pattern as it goes along. But the situation is not always as simple as this; indeed, with water waves it is very rarely simple, as we shall see.

When you are sitting by still water which is a foot or more deep, throw a small stone into the water to hit the surface a few feet from the water's edge. Then watch carefully how the ripples spread. The stone will produce a group of perhaps half a dozen distinct little wave crests. You may be puzzled to find that these ripples do not appear to travel in the same way as the ripples in a ripple-tank. They give a strange impression of travelling quickly and slowly at the same time. As we shall see, this is, in a sense, precisely what they are doing! If you fix your attention on one particular wave-crest, and try to follow it, you will find that you cannot do so for long because it disappears, although there is still a group of wave-crests spreading outwards. Although each individual crest disappears after a while, there always appears to be about the same number of crests in the complete group. This means that new crests must be appearing to make up for those which disappear. And this is exactly what is happening: you will see that the individual crests are in fact travelling faster than the group as a whole, so that new crests appear, apparently out of nothing, at the rear end of the group, travel along "through" the group, then diminish and disappear at its leading edge. Figure 22.2 shows the process in profile. In the pictures one ripple-crest has been drawn with a thickened line to make its progress easy to follow. You will see that the speed of this crest is twice the

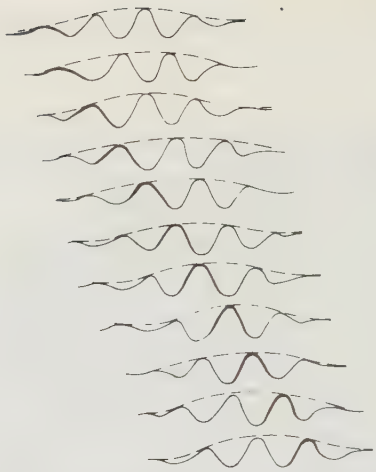


Fig. 22.2 The progress of a group of water waves.

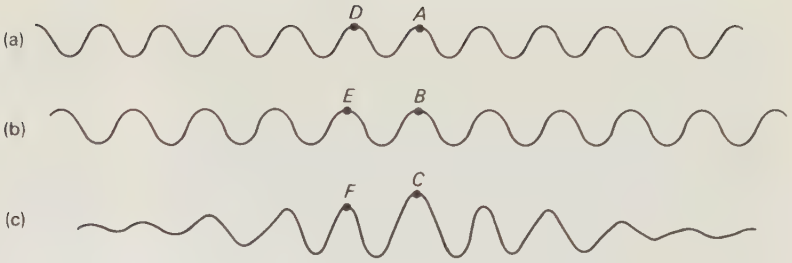


Fig. 22.3

speed of the summit of the group "envelope" (broken line). Notice one curious thing: if you were to fix a marker (a vertical stick, for example) in the water and count the number of crests which pass it as the group goes by, you would count perhaps ten crests (although the first and last might be too weak to observe with certainty): and yet there are never more than five distinct crests present in the group. In this particular situation the *group velocity* of the waves is just half the *phase velocity*. It is inadequate here to talk simply about the speed of the waves. We have to distinguish the phase velocity, which is the speed of individual crests, and corresponds to our old idea of wave speed, and the group velocity which is the speed of the wave envelope and also is the speed at which the total energy associated with the wave-group travels.

Looking again at Fig. 22.2 and imagining that the patterns were made of stiff bent wire, and not profiles of a water surface, we see that the pattern alters as the wave progresses: we could not use just one piece of bent wire, sliding along, to represent the progress of the waves, whereas one could do so for a pure sinusoidal wave. The fact that the wave-profile alters as it progresses is a consequence of the *dispersive* nature of the medium through which the wave is travelling. In a non-dispersive medium sinusoidal waves of different wavelengths all travel at the same speed. Up to now we have been careful to choose non-dispersive media for our waves so that there was one single speed for the waves, regardless of their wavelength, dependent only upon the properties of the medium.

It probably is far from clear to you that the changing profile of the wave-group of Fig. 22.2 is due to the fact that the velocity of a pure sinusoidal wave in the water depends upon its wavelength. The wave-group, we know, is not itself a pure sinusoidal wave: what relationship, then, does it have with a pure sinusoidal wave?

A wave-group whose profile is like those shown in Fig. 22.2 can be regarded as the result of superposing a number of pure sinusoidal wave-trains of different wavelengths. The full theory of this process of synthesizing a complex wave-profile by the superposition of pure sinsusoidal waves was developed by the French mathematician Fourier early in the nineteenth century. The theory cannot be treated here, but the germ of this idea of synthesis is embodied in the illustrative example which follows.

Suppose we have a pair of pure sinsusoidal wave-trains, each of constant wavelength and amplitude, represented by (a) and (b) in Fig. 22.3. Wave-train (b) has a slightly greater wavelength than wave (a). Imagine that these are the two component wave-trains which, when superposed, give the resultant complex wave-train whose motion we wish to study. At the instant represented in the diagram, part of the resultant wave-train will appear as (c). At this instant the crest A coincides exactly with the crest B and the resultant crest C is the largest crest in the group. Note that at a distance of $4\frac{1}{2}$ wavelengths to the left of A, a trough of (a) coincides exactly with a crest of (b), so that here there is exact cancellation: similarly to the right of A. Thus the profile of (c) is not unlike that of the wave-group in Fig. 22.2, but there is an important difference: the wave-group of Fig. 22.2 was an isolated group, whereas the profile of (c) is just one unit of an infinitely repeating pattern stretching to the right and left of it. Nevertheless, let us try to predict the progress

of (c). If the two wave-trains (a) and (b) travelled at exactly the same speed then the pattern (c) would also travel at the same speed and its shape would remain unaltered. This would be the situation in a non-dispersive medium. But we know that, for deep water waves, the speed depends upon the wavelength and, the greater the wavelength, the faster the speed. So, in deep water, the wave (b) will be going slightly faster than (a). This means that a little while after the instant shown in Fig. 22.3, B will be ahead of A, and E will coincide with D. Thus the largest crest of the group (c) will be crest F, crest C having diminished in size. Thus the centre of the wave-group has moved backwards relative to the component waves. Without attempting any calculation here we can see that it is possible for the speed of the wave-group to be different from the speed of the component pure sinusoidal waves: and in this example the speed of the wave-group is less than that of the component waves.

The situation just described is perhaps the simplest that can be devised to give a glimpse of the idea of Fourier analysis. In fact, an isolated group like that of Fig. 22.2, subjected to Fourier analysis, would be resolved into an infinite number of component sinusoidal wave-trains, each having a different wavelength and amplitude.

Finally it should be made clear that, in order to calculate theoretically the group velocity for a finite wave-train in a dispersive medium one does not have actually to carry out the process of Fourier analysis. A theoretical equation can be derived relating the group velocity, v_G with the phase velocity, v_P, and the wavelength, λ. This is

$$v_G = v_P - \frac{dv_P}{d\lambda}.$$

If you work out v_G for deep water waves you will find that it comes to exactly $v_P^{\frac{1}{2}}$. For surface ripples it is $v_P^{\frac{3}{2}}$.

To sum up: in a dispersive medium sinusoidal waves of different wavelengths (therefore different frequencies) travel at different speeds. In such a medium any *finite* wave-train is bound to change shape as it travels, and we need to distinguish the phase velocity of the waves in the train from the group velocity. Dispersion of light waves is discussed in a later chapter.

Chapter 23
WAVE BEHAVIOUR

23.1 Reflection of waves
23.2 Stationary waves
23.3 Wave phenomena in two dimensions
23.4 Particles, waves, and models
23.5 Some interference phenomena
23.6 Coherence

23.1 REFLECTION OF WAVES

We will start our investigation of this phenomenon by considering waves on a string. If a slinky spring is stretched taut on the floor with one end held rigidly, and then the far end is jerked rapidly to one side and back again, a pulse travels along and rebounds, inverted, from the clamped end. Figure 23.1 shows this.

A similar effect can be observed using a heavy rope, or long, floppy, rubber tube. We will now describe a way of looking at this reflection process which enables us to analyse it at each stage. We shall need to apply two concepts: the *principle of superposition*, and the mathematical concept of images formed by reflection. Indeed, the argument which follows is a nice application of this mathematical concept which is taught in many modern mathematics courses. But you must not be misled into thinking that the argument which follows is a modern one: it appeared in 1895 (Feather, 1961). In the upper part of Fig. 23.2 we see a pulse travelling from left to right along a string. If the string is attached to rigid supports at both ends the pulse will be reflected, without change of shape or amplitude, although it will be inverted, from either end. This assumes that we have an ideal string where there is no loss of energy. The pulse will always be upright when going from left to right, and upside down when going from right to left. In the lower part of Fig. 23.2 we see part of an infinite string with pulses going along it both ways. What we shall prove now is that the process taking place on the section PQ of the infinite string is identical to the process on the finite string PQ which is clamped at both ends. The upright pulses, Z, A, B, are identical, all travelling at the same speed from left to right. The string is marked off into equal lengths by the points N, O, P, Q, R, S. The upside-down pulses, Y', Z', A', are travelling at the same speed as the upright pulses, but in the opposite direction. Pulse Z' is related to pulse Z by geometrical reflection in the line of the undisturbed string, and by point reflection in the point O. The pulse-pairs A and A', B and B', are similarly related. Now concentrate attention upon what will happen at point Q. In a

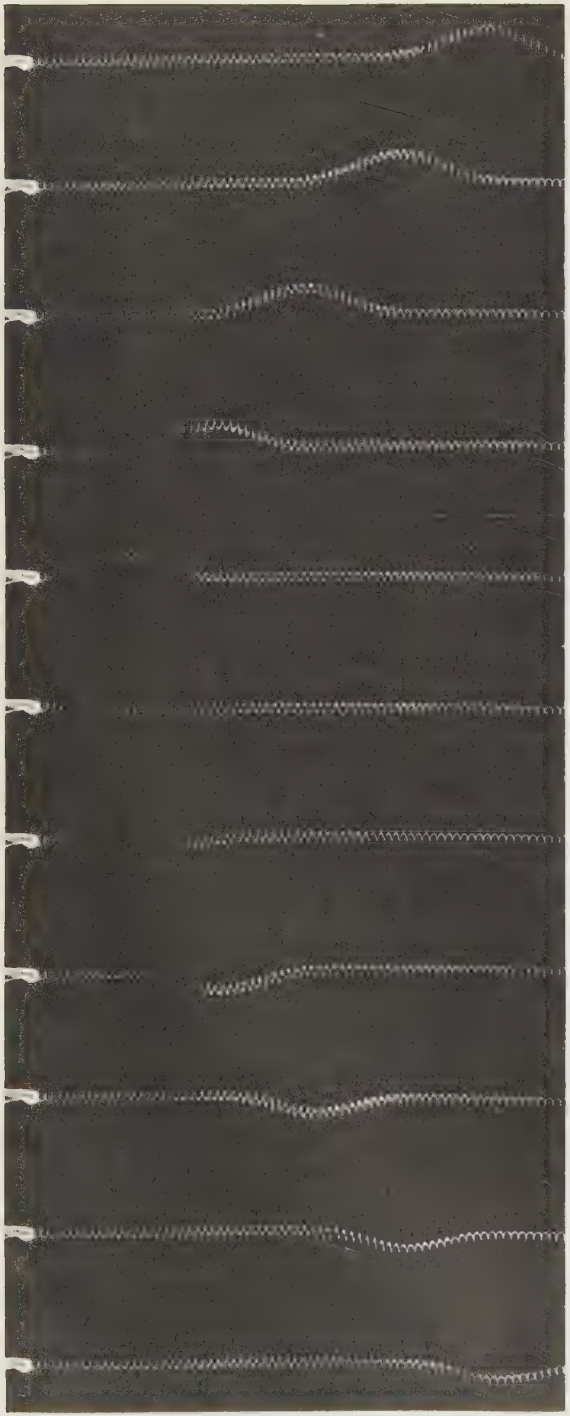


Fig. 23.1 Reflection of a pulse from the fixed end of a slinky spring. (From P.S.S.C. *Physics*, D. C. Heath, 1965, reproduced by permission of Education Development Center, Inc.)

moment the two pulses, A and A', will arrive, and it is clear that at point Q their effects will exactly cancel, so that the resultant displacement of the string at Q will be zero at every instant. A little later, pulse A will be on its way from Q to R, and pulse A on its journey from Q to P. When A arrives at P, it will meet Z and cause complete cancellation of P. Then, a little later still, Z will be on its way from P to Q. Thus it can be seen that, at points P and Q, there is zero displacement at every instant, and thus what goes on in the region PQ is identical to what happens on the clamped string. This result has far-reaching consequences because it implies that any situation in which a wave disturbance is travelling back and forth along a medium bounded by reflecting barriers can be considered as the superposition of two separate trains of waves travelling steadily in opposite directions through an infinite, unbounded, medium. In particular, we can now analyse closely what happens when the pulse of Fig. 23.2 is actually undergoing reflection.

Figure 23.3(a) shows the situation shortly after the leading edge of the pulse A has arrived at Q. The dotted lines represent the oncoming pulse and the reverse pulse as they would be on the infinite string. The resultant shape of this region of the infinite string would be as shown by the solid line, continued by the broken line: thus the solid line portion represents the actual profile of the finite string at this instant during the reflection of the pulse. Figure 23.3(b) shows the situation a little later.

As we have seen, a wave-pulse is reflected, *with inversion*, at a rigid boundary: this is called a hard reflection. But reflection occurs, in general, whenever there is a discontinuity in the wave medium: this does not have to be a rigid boundary. We can demonstrate a soft reflection by means of a slinky and a long thread, as shown in Fig. 23.4.

The thread keeps the spring in tension but offers negligible resistance to any transverse displacement of

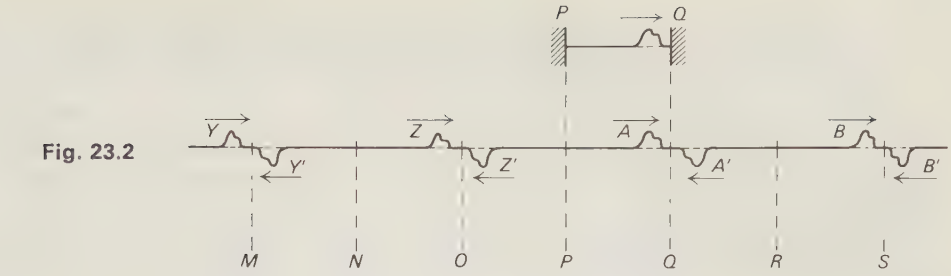


Fig. 23.2

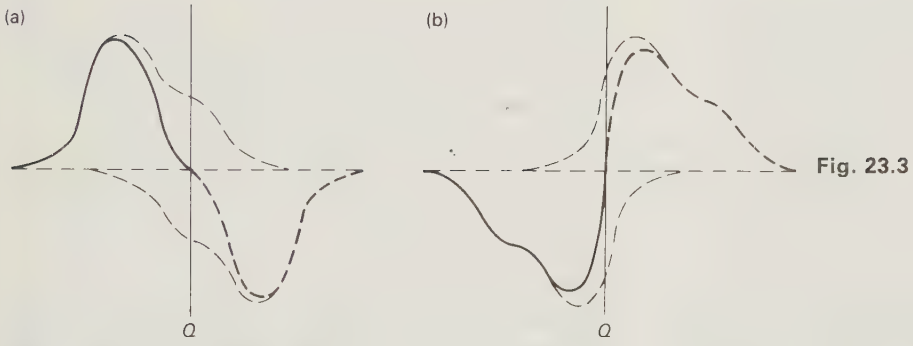


Fig. 23.3

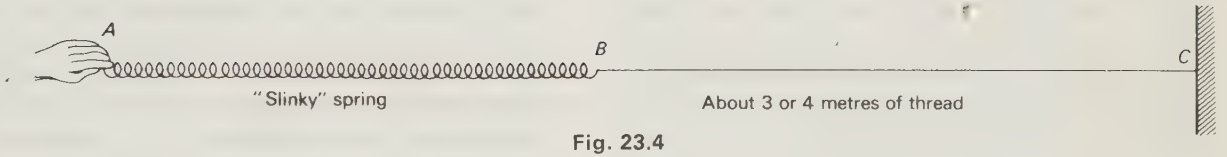


Fig. 23.4

the end J. When a pulse is sent along the spring from H to J it is reflected from J, but this time without inversion. If the thread has negligible mass, then it absorbs no energy and the reflected pulse carries the same energy as the incident pulse.

Both hard and soft reflections, in an ideal situation, involve 100% reflection of a wave's energy. Briefly here we must mention partial reflections in which only a fraction of the energy of the incident wave returns with the reflected wave. There are two principal situations in which we find waves are partially reflected. The first situation is the easier to demonstrate. Attach a small mass, say about 100 g of lead, to one coil about midway along a slinky spring. When you send a pulse along from one end you will find that there is a weak reflected pulse, and the pulse which travels onwards is slightly reduced in amplitude. In this situation a small, localized, inhomogeneity in an otherwise homogeneous wave-medium causes a partial reflection. It is intuitively obvious that, the larger this inhomogeneity, the stronger will be the reflection. In the limit, an inhomogeneity of infinite mass will cause 100% reflection of the wave's energy, acting exactly as a perfectly rigid boundary.

The second situation is the partial reflection which always occurs at a boundary between two wave-media having different characteristics. This is demonstrated with two slinky springs joined together, one of the springs having had its characteristics altered. For

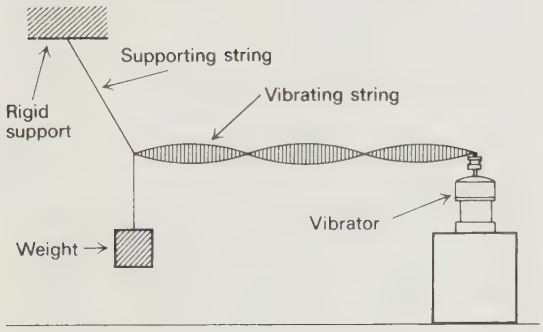


Fig. 23.5

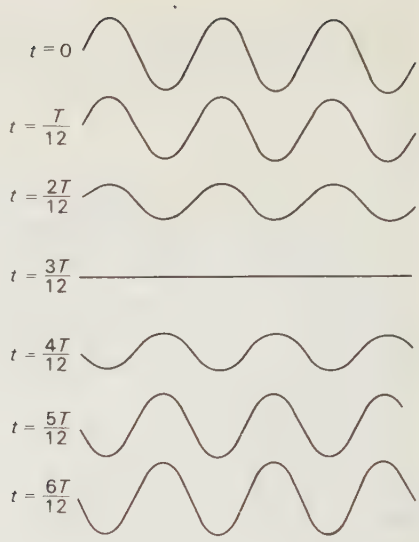


Fig. 23.6 These curves are "glimpses" of a stationary transverse sinusoidal wave, at time intervals of $T/12$, where T is the period of the vibration.

instance, one can alter the stiffness of the material by heating it strongly and letting it cool slowly (this will effectively ruin the spring except for the purposes of this demonstration!). Then, when a pulse is sent along from one end, it is partially reflected at the junction between the two springs. The characteristics of the two media which determine how strong the partial reflection will be are the impedances of the media. For a stretched string the *impedance* is μv, the product of the mass per unit length and the wave-speed along the string. If two strings having the same impedance (which does not mean they are necessarily identical) are joined, there will be no wave reflection at the junction. The greater the discrepancy between the two impedances, the stronger is the reflected wave. A rigid boundary acts as a medium with infinite impedance, a free boundary (as in Fig. 23.4) has zero impedance.

23.2 STATIONARY WAVES

It is best to begin by observing a stationary transverse wave on a stretched string. The arrangement shown in Fig. 23.5 is good. A suitable tension is applied by means of small weights hanging on the vertical string. The frequency of the vibrator is varied until the string vibrates with large amplitude. The pattern of vibration forms a number of fuzzy loops. The higher the frequency, the greater the number of loops. It is only at certain discrete frequencies that the string can be excited into vibration of appreciable amplitude. These are the string's resonant frequencies. With each different mode of vibration of the string, characterized by the number of loops in the pattern, one particular frequency is associated, called the *eigen frequency* of that mode ("eigen frequency" is of German-English coinage, "eigen" = "own"). To follow the motion of the string in detail we can use a stroboscopic lamp, adjusting the rate of flashing so that the string appears to move very slowly. We then observe that the string performs a periodic sequence of movements as illustrated in Fig. 23.6. It is a very special kind of wave motion: the crests do not move along the string, instead they shrink, turn into troughs, and back into crests again. Within any one loop of the pattern all the particles of the string are oscillating in phase with each other, and in antiphase with all particles in the two neighbouring loops. The amplitude of oscillation of any chosen particle is constant, and at certain places on the string this amplitude is a maximum. Such places are called

antinodes of the stationary wave pattern. At other places the amplitude is nearly zero. Points of zero amplitude are called *nodes*. (The word is derived probably from the Latin *nodus* a knot.)

Before attempting to analyse the process by which a stationary wave is generated, it is worth our returning to play with a slinky spring once more, and attempt to produce a stationary wave pattern with it. Shaking one end rhythmically from side to side, with gradually increasing frequency, one finds that at certain frequencies the wave-pattern ceases to travel along the spring: instead, a side-to-side wiggle begins and we have a stationary wave.

We know that any wave disturbance, excited at one end of, and travelling along, a stretched string (or spring) is reflected from the anchored end. This suggests that the stationary wave arises from the combined effect of an outgoing and a reflected wave. Let us try to predict what will be the result of superposing a continuous, sinusoidal wave of constant amplitude, travelling from left to right along a string, and the wave which results from this outgoing wave being reflected at the fixed end of the string. We can use the technique of Section 23.1 which uses the fact that our situation is equivalent to two continuous waves going in opposite directions along an infinite string. Figure 23.7 shows the profiles (broken lines) of the two waves, at successive instants in time. The solid line shows the result of their superposition. So here is our stationary wave, predicted by this geometrical argument.

Now we shall carry out the same process of superposition algebraically. We can represent the wave going from left to right by the expression

$$y_1 = a_0 \cos \omega \left(t - \frac{x}{v}\right).$$

This implies that the particle at $x = 0$ has maximum positive displacement ($+ a_0$) when we start our clock, that is, when $t = 0$. We choose the origin of the x-axis, in relation to the reflecting boundary, so that the reverse wave also has maximum positive displacement at $x = 0$ at the same instant as the forward wave (thus frame no. 1 in the sequence of Fig. 23.7

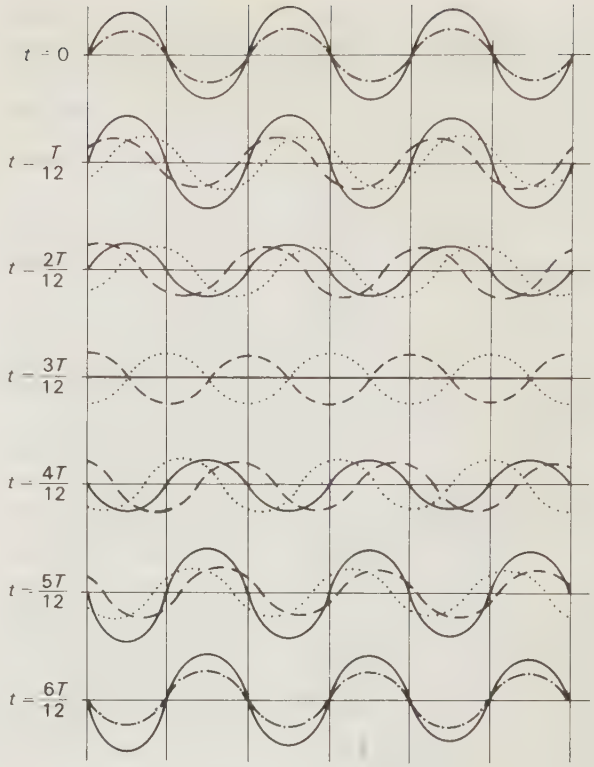


Fig. 23.7 The formation of a stationary wave. The dashed line represents a travelling sinusoidal wave going towards the left, and the dotted line represents a wave going to the right. The full line is the stationary wave pattern which results from the superposition of the two travelling waves (compare Fig. 23.6).

could represent the situation at the instant when $t = 0$). Thus we can write, for the reverse wave:

$$y_2 = a_0 \cos \omega \left(t + \frac{x}{v}\right).$$

The resultant displacement, y, for any value of x and t, is given by

$$y = y_1 + y_2,$$

which yields, after the cosine terms have been expanded,

$$y = 2a_0 \cos \omega t \cos \frac{\omega x}{v} \qquad (23.1)$$

and this expression represents the solid line in Fig. 23.7. Verify this for yourself. At time $t = 0$, $\cos \omega t = 1$, and then $y = 2a_0 \cos \omega x/v$. The maximum particle displacement occurs at the places where the value of x is such that

$$\cos \frac{\omega x}{v} = +1 \quad \text{or} \quad -1,$$

therefore

$$\frac{\omega x}{v} = n\pi, \quad \text{where} \quad n = 0, 1, 2, 3, \text{etc.}$$

Now from Eq. (17.5)

$$\omega = 2\pi f$$

and from Eq. (21.11)

$$\frac{f}{v} = \frac{1}{\lambda}.$$

So that

$$\frac{\omega}{v} = \frac{2\pi f}{v} = \frac{2\pi}{\lambda} \qquad \therefore x = \frac{n\lambda}{2}.$$

So antinodes occur at

$$x = 0, \quad x = \frac{\lambda}{2}, \quad x = \lambda, \quad x = \frac{3\lambda}{2}, \text{ etc.}$$

Nodes occur at values of x for which $\cos \omega x/v = 0$, that is, when

$$\frac{\omega x}{v} = (n + \tfrac{1}{2})\pi \qquad \text{and} \qquad \frac{2\pi x}{\lambda} = (n + \tfrac{1}{2})\pi.$$

Therefore

$$x = \frac{\lambda}{4}, \quad \frac{3\lambda}{4}, \quad \frac{5\lambda}{4}, \quad \frac{7\lambda}{4}, \text{ etc.}$$

These results can be summarized by saying that the loops of the stationary wave are one half-wavelength in width.

A stretched string, anchored firmly at both ends, is one example of a system which can support a stationary wave. A number of modes of vibration are possible. The boundary conditions which apply to this system are that the stationary wave pattern must have a node at each fixed end of the string: with this restriction, any number of loops is possible. If we assume that progressive waves of all frequencies travel with the same speed in this system, we can find an expression for the eigen frequencies of the system. The lowest eigen frequency will be that associated with the single-loop mode. If L is the length of the string,

$$\frac{\lambda_1}{2} = L, \qquad f_1 = \frac{v}{\lambda_1} = \frac{v}{2L},$$

where v is the speed of transverse waves in the system. The next eigen frequency is that of the two-loop mode, when $\lambda_2 = L$

$$f_2 = \frac{v}{L}.$$

Therefore the nth eigen frequency is given by

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

and thus the eigen frequencies go in whole number multiples of the lowest, fundamental, eigen frequency, f_1.

The modes of vibration described above are the normal modes, and each one has its characteristic eigen frequency. It is possible for more than one normal mode to be excited at the same time. The result is then a complex mode and it can be regarded as the superposition of the constituent normal modes. Figure 23.8 shows the profile of a complex mode formed by superposition of the two lowest modes of vibration of a stretched string.

The application of the foregoing principles to stringed musical instruments are of great importance, but there is no room to treat them in detail here. Problems 5.20–5.22 take the student into that realm, and consider also stationary *longitudinal* waves, with particular reference to wind instruments.

In the case of a stretched string the eigen frequencies of the normal modes of vibration are all whole number multiples of the fundamental frequency. This is by no means always the situation. The prong of a tuning fork, for instance, has overtones whose

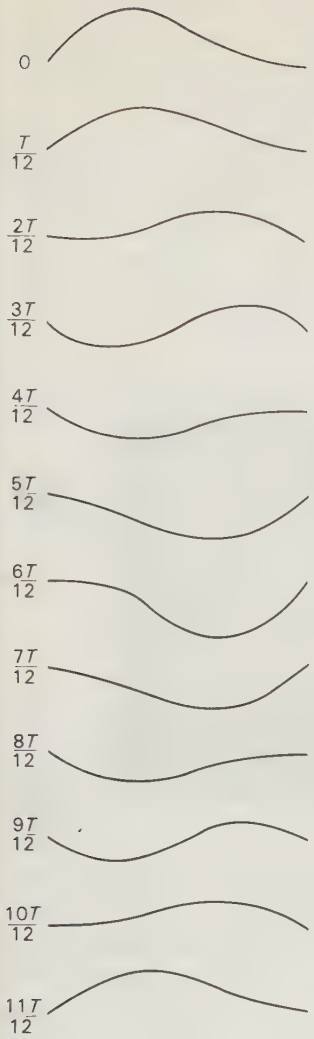


Fig. 23.8

frequencies are non-integral multiples of the fundamental. One prong vibrates in the modes which were shown earlier in the discussion of mechanical resonance: you should look back at Figs. 20.8 and 20.9, where you will see that the three lowest modes of vibration have eigen frequencies which are related in a fashion very different from that for a stretched string. A metal bell is a good example of a system in which the eigen frequencies are not generally integral multiples of the fundamental. Although the bell-makers of centuries ago could not perform a mathematical analysis of the oscillatory behaviour of a bell, they did know that certain overtones could have frequencies which caused a very discordant sound when heard in conjunction with the fundamental frequency. In a sweet-sounding bell the most discordant overtones are suppressed, by the skill of the bell-maker. A drumskin has many normal modes of vibration, and the overtones again have eigen frequencies which are not integral multiples of the fundamental—similarly, the wooden bars of a xylophone, or the metal bars of a glockenspiel, or metal chimes. Also, a prime example is shown in the delightful experiments with Chladni's plate, described in Section 20.2.

Stationary waves can often be seen in water. A cup of tea, when jiggled around, can show this, and stationary waves can be excited in a small tank, for instance your bath. Do you think that the eigen frequencies of deep water waves in a narrow rectangular tank, for instance, would be whole number multiples of a fundamental? The discussion in Chapter 22 can help you to answer this.

23.3 WAVE PHENOMENA IN TWO DIMENSIONS

It is very likely that you have already studied a variety of wave phenomena in a ripple-tank. Here, therefore we give only the briefest summary of these.

a) Propagation of straight and circular pulses

A short train of ripples which, seen from above, is straight for most of its length and which, because it is very short, we may call a pulse, can be produced by dipping a horizontal straight rod into the water surface. Alternatively a round straight rod can be rolled a short distance along the bottom of the ripple tank. The pulse remains straight and is observed to travel, with little dispersion (Section 22.1), at a constant speed, in a direction perpendicular to the line of the pulse, as in Fig. 23.9. Circular pulses are generated by dipping a fingertip into the water or by drops of water released from a liquid dropper.



Fig. 23.9

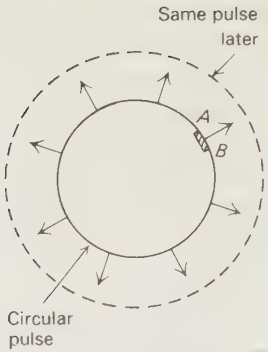
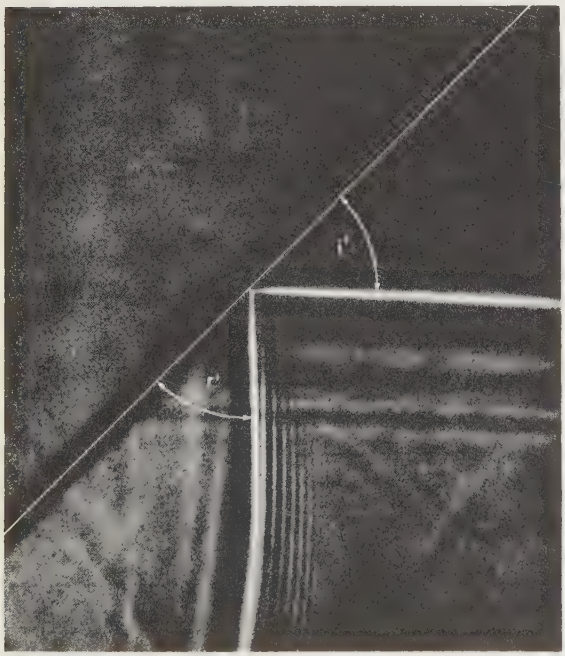
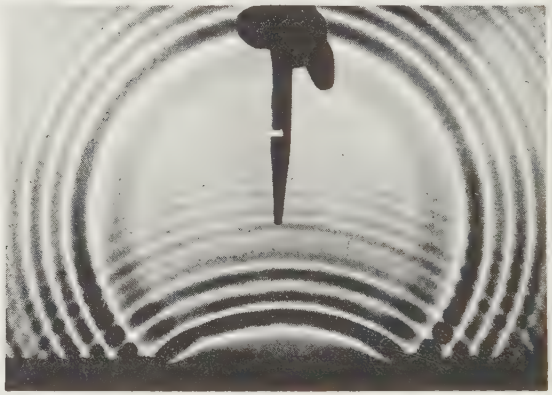


Fig. 23.10



(a)



(b)

Fig. 23.11 (a) A straight pulse moving up to a barrier and being reflected. (b) A circular pulse moving towards a straight barrier and being reflected. (From P.S.S.C. *Physics*, D. C. Heath, 1965. Reproduced by permission of Education Development Center, Inc.)

Provided the water has a uniform depth (you may recall that the speed of water waves depends on depth) the pulses keep their circular shape and appear to travel outwards from the source of the disturbance with a constant speed. Their progress is represented in Fig. 23.10. Here again, each portion of the pulse travels in a direction perpendicular to the line of the pulse (see the portion labelled AB in Fig. 23.10).

b) Reflection at barriers

Figure 23.11 shows the reflection of a straight and of a circular pulse at a straight barrier. In the case of the straight pulse it is always found that the angle between the incident pulse and the barrier is equal to the angle between the reflected pulse and the barrier. If we indicate in a diagram, as in Fig. 23.12, not the line

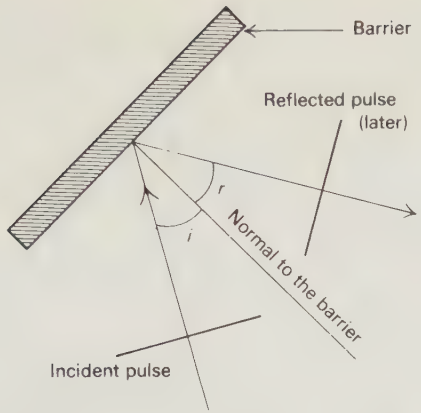


Fig. 23.12 Reflection of a straight pulse from a straight barrier. The angle of incidence is equal to the angle of reflection.

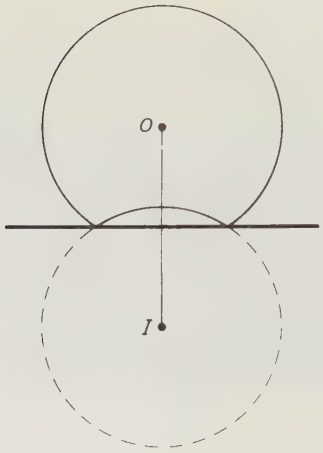
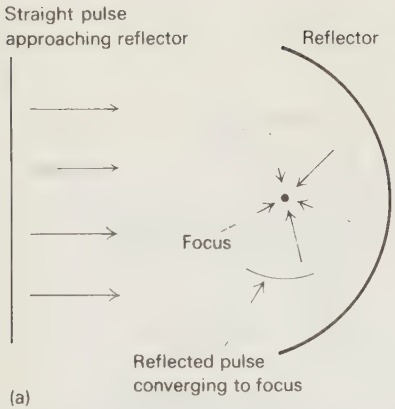


Fig. 23.13 Reflection of a circular pulse at a straight barrier. The pulse started at *O*. The reflected part has the same shape as a pulse which started from the "image point" *I* would have.

of the pulse itself but simply its direction of propagation (which of course is the direction of the energy flux associated with the wave disturbance), then we see that angles between the directions of propagation of the incident and reflected pulses and the barrier must also be equal. It is well known that this law of reflection is obeyed by thin beams of light at any polished reflecting surface. However, in optics, one is usually dealing with a three-dimensional situation and the geometrical relations between the light rays and a reflecting surface are more conveniently described in terms of an imaginary line drawn perpendicular to the reflecting surface. This line is called the *normal* to the surface, and is shown in Fig. 23.12. The angles of incidence and reflection are defined with reference to the normal, rather than to the plane of the surface itself.

The shape of the reflected ripples in Fig. 23.13 is seen to be circular, and the centre of their curvature is a point *I*, behind the reflecting barrier, such that *I* and *O* are equidistant from the reflecting surface and the line joining *I* and *O* is perpendicular to the surface. An analogous effect is observed with light and a plane reflector such as an ordinary mirror: if *O* is a point object in front of the mirror, then its image (a virtual image, in the language of optics) is at *I*.

Further close similarity between the behaviour of ripples in a ripple-tank and the behaviour of light, when reflected, can be demonstrated using curved barriers. Figure 23.14 shows just two instances. In (a) a straight pulse has been reflected from a concave barrier in the shape of a circular arc. The reflected pulse is observed to be circular, and moving inwards towards a point midway between the centre of curvature and the reflector. This is analogous to the effect of a concave mirror upon a parallel beam of light (or, for that matter, other forms of electromagnetic radiation which can be reflected in a similar way): the energy associated with the light is concentrated into a small space, the *focus* of the reflector. The effect is reversible, as can be shown by starting circular ripples from the focal point, when it will be found that the reflected ripples are nearly straight. Figure 23.14b shows the reflection of a straight pulse from a convex circular barrier. The centre of the reflected pulse is at the image point, *I*.



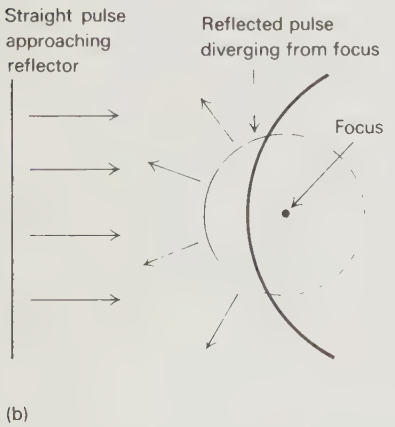


Fig. 23.14 Reflection of a straight pulse at a circular barrier. In each case the reflected ripples are approximately circular in shape.

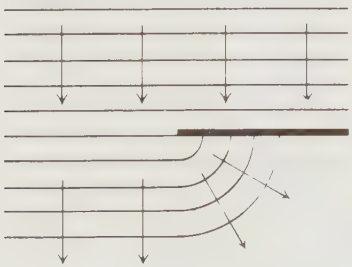


Fig. 23.15 Diffraction of straight ripples at the edge of a barrier.

c) Diffraction

A ripple-tank is admirably suited to demonstrating what happens to waves at the edges of barriers. Figure 23.15 shows a succession of ripples, produced by a straight rod vibrating up and down on the water surface with a constant frequency, passing the edge of a straight barrier. Whereas all the energy associated with the waves was originally travelling from the bottom to the top of the diagram after passing the barrier, some of the wave energy is being propagated in other directions.

Does a similar effect occur with light? Grimaldi (1665) observed that light could "bend into shadow" when shadows of very smooth-edged objects were cast on a wall by light from a very small, intense source, such as a small hole in a shutter with sunlight coming through it. It was he who gave the name *diffraction* to the phenomenon. Isaac Newton, in his treatise on *Opticks* which was published in 1704, described in detail similar experiments and tried to explain the phenomena using his corpusclar (or particle) model of light. He failed, and although his corpuscular model was accepted by natural philosophers throughout the eighteenth century and even well into the nineteenth it was Huygens' wave model as expounded in his *Traité de la lumière*, published at Leyden in 1690, which eventually was able to be developed sufficiently to explain diffraction phenomena. But this development did not occur until well into the nineteenth century, in the work of Thomas Young and Augustin Fresnel.

Diffraction at the edge of a barrier can be demonstrated with sound waves in air: in fact we experience it whenever we hear sounds coming from a source, out of sight, around the corner of a building. It can also be shown with radio waves, using a radio transmitter (for laboratory demonstration one uses a very high frequency (VHF) transmitter) and barrier made of electrically conducting material which is opaque to radio waves. The invisible waves are detected by suitable antenna connected to a radio receiver.

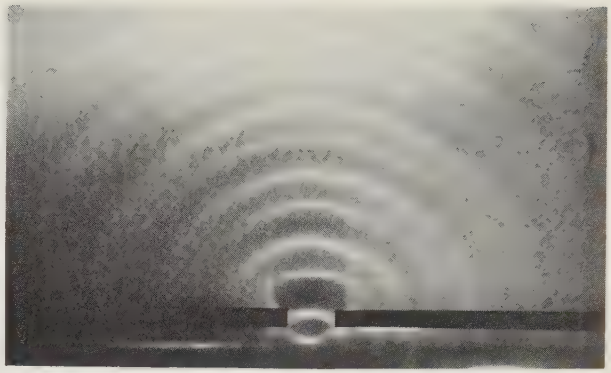
Diffraction by an aperture in a barrier is shown in the sequence of pictures in Fig. 23.16. Again, the source of waves is a vibrating rod. These pictures

show clearly that when the width of the aperture is small compared with the wavelength of the waves (the wavelength being here the distance between successive bright regions in the shadow photograph of the ripples) the pattern made by the diffracted waves is indistinguishable from that due to a single, weak, point source situated in the centre of the gap. The energy flux associated with the diffracted waves is equally distributed over a 180° sector.

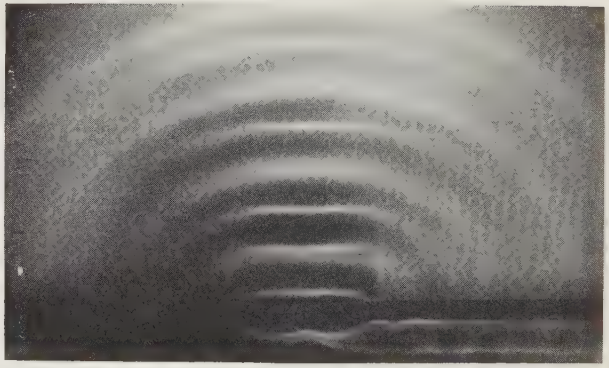
In Fig. 23.16(b), where the width of the aperture is some two or three times the wavelength of the waves we see that the energy flux distribution is not uniform, and the bulk of the energy flux is being propagated in directions similar to the direction of the incident waves. This is the interpretation of the observation that the ripples are strong and distinct only within a narrow sector emerging from the aperture. Also we see, if we look carefully at one particular diffracted ripple, moving our eye along it from left to right, that the amplitude is not simply increasing steadily towards the centre and then decreasing, but that it varies through a series of maxima and minima, although the central maximum is outstandingly the greatest. Figure 23.16(c) shows the pattern produced when the aperture is many wavelengths wide, and here we see that there is very little energy flux propagated in directions other than the original direction. The spatial fluctuations in amplitude of the diffracted ripples as seen in Fig. 23.16(b) are curious: these will be referred to later when diffraction of light by a small aperture is considered.

The diffraction of sound waves by an aperture is an interesting phenomenon to consider, and Problem (5.17) involves a discussion of this.

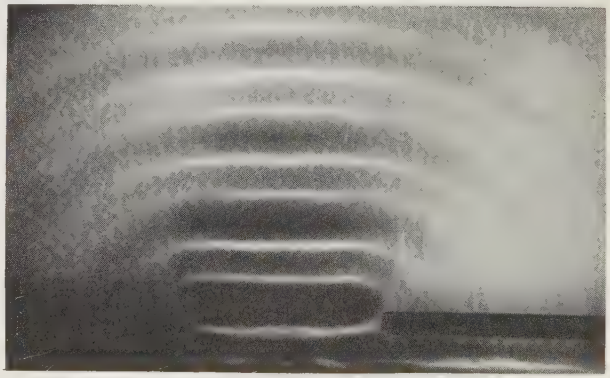
Another diffraction effect, of considerable importance in light and other electromagnetic waves, can be demonstrated in a ripple-tank. It is generally known as *scattering*, and Fig. 23.17 shows this. Figure 23.17(a) shows the effect of an obstacle whose width is large compared with the wavelength of the waves. The pattern produced by this is what one expects, if one knows about diffraction at the edge of a barrier. But the situation shown in Fig. 23.17(b) is perhaps surprising. The width of the obstacle here is small compared with the wavelength, and we get



(a)



(b)



(c)

Fig. 23.16 Diffraction of straight ripples passing through a series of apertures. (Photograph by E.J.W.)

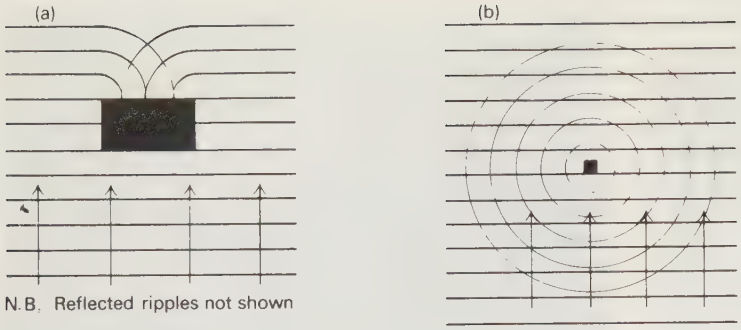


Fig. 23.17 Scattering of ripples by an obstacle.

precisely circular ripples produced. Clearly they are very weak in comparison with the incident ripples. Thus we see that, when the obstacle is very small, (a) the energy flux of the scattered waves is uniformly distributed into all directions, and (b) the total energy flux of the scattered waves is a very small fraction of the incident energy flux. We have made use of this phenomenon in Unit 4 when discussing the uses of X-ray diffraction.

The blue light from a clear sky is the result of sunlight scattered by particles of the atmosphere. Radio waves are scattered to some extent by the atmosphere, the amount of scattering depending both on the wavelength of the waves and the particular constitution of the region of atmosphere concerned.

d) Refraction

This phenomenon is the change in direction of propagation of waves resulting from a change in speed. In a ripple-tank this is demonstrated by arranging two regions of different depth: easily done by placing a flat sheet of glass or plastic in the tank so that there is a uniform region of shallow water above the sheet, while the water in the rest of the tank has a uniform greater depth. The effect upon a continuous wave-train going from deep into shallow water and thus undergoing a reduction in speed is shown in Fig. 23.18a. If the ripples are parallel to the straight boundary between deep and shallow water they do not change direction, but the wavelength is reduced. This is a special case of refraction: the more general case is shown in Fig. 23.18b. If the incident ripples are not parallel to the boundary, then their direction of propagation is altered. Not all the incident energy flux crosses the boundary: a fraction of it is always reflected.

Figure 23.19 shows how the direction of propagation of the incident waves, and the direction of the refracted waves, are related to the normal to the boundary. It always happens that when the speed is reduced the direction of propagation is deviated towards the normal, so that the angle of refraction, r, is smaller than the angle of incidence, i. The effect is reversible, so that if waves move into a region where they travel faster, their direction of propagation is deviated away from the normal. Rays of light suffer refraction in precisely the same fashion in passing from, say, air into water. If one accepts a wave model for light, this implies that the velocity of light in water is less than that in air. On the other hand, acceptance of the particle model requires that the light should be accelerated in passing from air to water. The wave model and the particle model thus come into direct conflict in this phenomenon. It was not until 1850, when Jean Foucault made direct measurements of the speed of light in air, water, and other media, that the simple particle model for refraction had to be discarded finally.

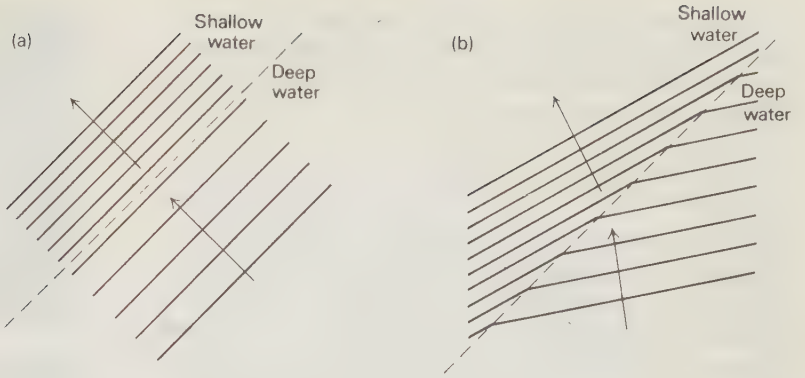


Fig. 23.18 Refraction of ripples at a straight boundary.

e) Dispersion

Referring again to Fig. 23.19, we see that this diagram can be used to represent the refraction of a light ray at the boundary between two transparent media. But it only truly represents the situation if the incident light is monochromatic (that is, is of a single, pure colour). If white light is used the refracted light is split up into the colours of the spectrum, the angle of refraction, r, being slightly different for the different colours. Dispersion, in light, is commonly taken to mean this splitting up. Refraction at the two successive faces of a triangular glass prism enhances the effect, and such prisms have been used since the time of Newton to produce spectra of light. But what exactly is dispersion, on the wave model of light? Yet another ripple-tank demonstration can help us although it is a difficult one to do convincingly. The tank and accessories are set up as for demonstrating refraction. Continuous waves of a suitably chosen long wavelength are refracted obliquely across a boundary, going from deep into shallow water. The direction of propagation of the refracted waves is carefully observed. It is found that, when the frequency of the vibrator is increased so that the wavelength is reduced, the direction of the refracted waves alters slightly. We interpret this as meaning that the change in speed at the boundary for long waves is slightly different from the change in speed for short waves, which in turn implies that the speed of the waves in the deep water, or in the shallow, or in both, depends upon the wavelength of the waves. This we know to be true: refer back, if you like, to Table 22.1.

In wave theory the term dispersion is used for the situation in which the speed of a sinusoidal wave depends upon the wavelength. Free space (empty space) is a non-dispersive medium for electromagnetic waves: that is to say, all electromagnetic waves travel (as far as is known) at precisely the same speed in free space. In any material medium, however, electromagnetic waves are subject to dispersion. For instance, in glass, the speed of light waves of the shortest wavelength (which give the sensation of violet colour to an observer) is about 2% less than the

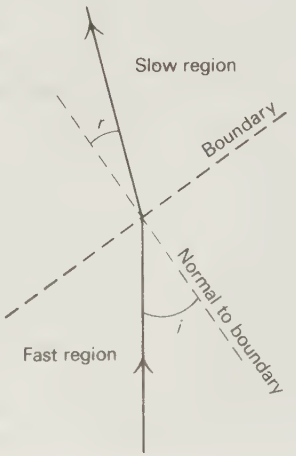


Fig. 23.19

speed of the longest wavelength waves (red). Dispersion, in connection with water waves, was discussed in Chapter 22, and the effect of a dispersive medium on a wave pulse was described.

f) Interference

We shall need to spend considerably more time on this phenomenon than on the others: but here we shall summarize, as we have done with the others, the elementary ripple-tank demonstrations and their relevance to the behaviour of other kinds of waves. Figure 23.20 shows the pattern produced by a pair of vibrating prongs, moving with the same frequency, amplitude, and in phase, on water in a ripple-tank. In the upper picture the frequency of vibration is high, the wavelength short. In the lower one the frequency is low, the wavelength longer. In some places the water surface is subject to a continuous large disturbance, in other places it remains almost perfectly calm as a result of the destructive interference of the two sets of waves. At point P in the upper picture there is constructive interference: a wave-crest from prong A arrives at P at the same instant as a wave-crest from B, and thus (by the principle of superposition), a double-size wave-crest results at P. Half a cycle later a trough from A arrives at P at the same instant as a trough from B, and we get a double-size trough. For point Q, in the calm water, we can follow a similar argument: its distance from A and B is such that when a crest arrives from A, a trough is arriving from B, and these two very nearly cancel to produce flat water. The cancellation is not complete because the amplitude at Q of the disturbance due to A is slightly less than that due to B, because Q is a little farther from A than from B. Half a cycle later it is a trough from A and a crest from B which combine, and again cancel. In this way the water at Q is continuously calm.

By comparing the two pictures in Fig. 23.20 one sees that the angular separation of the calm water regions depends on the wavelength of the waves. The precise dependence of the interference pattern upon parameters such as wavelength, and the distance between the sources of waves, will be discussed in

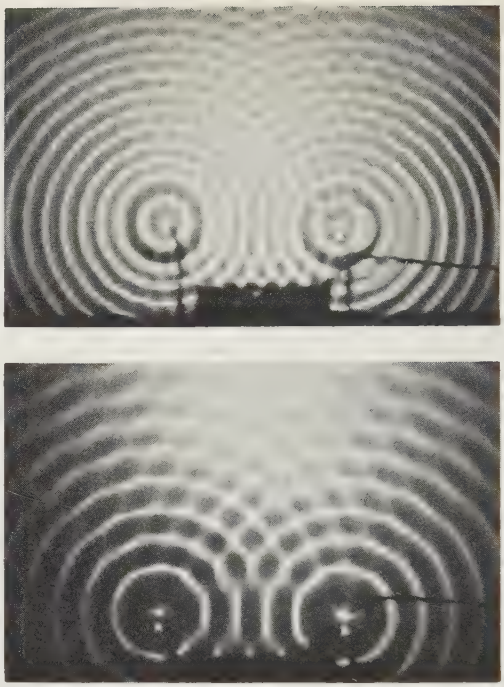


Fig. 23.20 Interference of ripples from two sources vibrating in phase. (Photograph by E.J.W.)

detail shortly, when the phenomenon in light is described and analysed. At this juncture we should note some correspondences between this twin-source interference, just described, and the corresponding one-dimensional situation described earlier, in Section 23.2. We saw there that the result of superposing two opposite-going sinusoidal waves, of the same frequency, amplitude, and speed, was a stationary wave. In the stationary wave there are places where there is a continuous vibration, and also places where there is no vibration. In this respect we have a situation similar to that in our twin-source interference pattern. The correspondence is almost exact if we imagine that, in the ripple-tank, we have a narrow channel between the two vibrating prongs. Along this channel there would travel two lots of waves, going in opposite directions, and so we would expect the motion of the water surface in this channel to be that of the station-

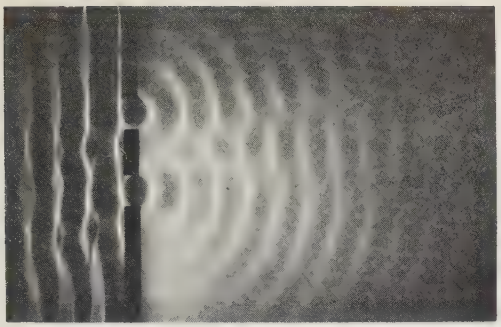


Fig. 23.21 Interference of the diffracted ripples from two apertures. (Photograph by E.J.W.)

ary wave. Just as we talked about nodes and antinodes in describing a stationary wave: points where there is no disturbance and points where the disturbance is a maximum, so we can talk about *nodal lines* and *antinodal lines* in the two-dimensional interference pattern in our ripple-tank.

Finally, before we leave ripple-tank demonstrations, we will mention a version of the twin-source interference experiment involving diffraction. Figure 23.21 shows the pattern obtained by having a barrier with two narrow apertures in it, with straight ripples approaching the barrier, and parallel to it, from one side. The pattern on the other side of the barrier is indistinguishable from the pattern produced by a pair of vibrating prongs: not surprising, because we have already seen that ripples diffracted by a very narrow aperture are indistinguishable from ripples produced by a single vibrating prong (except that they spread out over a 180° sector instead of a complete 360° circle).

23.4 PARTICLES, WAVES, AND MODELS

In examining the transmission of energy by a stream of particles we expect certain characteristic behaviour: in the absence of a field of force a stream of particles will travel in a straight line at constant velocity; such a stream of particles can be reflected and, when this occurs, the angles of incidence and reflection are equal: a stream of particles can be refracted and, where the velocity is found to be highest in the denser medium, the particle track is reminiscent of a ray of light being refracted as it passes from air into glass.

So long as this behaviour adequately describes a mode of energy transmission, a particle model for that mode is satisfactory.

In examining the transmission of energy by waves on water, for example, we find a further range of characteristic behaviour. Waves can travel in straight lines at a constant velocity; waves can be reflected and, when this occurs, the angle of incidence and reflection are equal: waves can be refracted and, when this occurs, the velocity of the wave is found to be lower in the denser medium; waves can suffer diffraction and the effects are most readily observed if the slit or barrier used have about the same dimensions as the wavelength involved; waves can interfere constructively and destructively under appropriate conditions.

It is not surprising then that, on the identification of these further properties in any mode of transmission of energy, the wave model should displace the particle model for that mode.

This has been precisely the case for light which displays all the typical behaviour of a wave. But so do sound, infra-radiation, ultra-violet radiation, X-radiation, radio waves and so on.

23.5 SOME INTERFERENCE PHENOMENA

We must now take a close look at some important examples of interference.

Thomas Young, a brilliant English scientist, trained initially in medicine, was the first person to devise a convincing demonstration of twin-source interference in light.

A modern, and also a very simple version of his experiment is depicted in Fig. 23.22. Lamp L is a clear glass, straight filament lamp (for example, a 12 V 24 W car headlamp-type bulb). It should be surrounded by a suitable opaque shield (not shown) to reduce stray light. S is an ordinary microscope slide blackened with "Aquadag," having a pair of parallel slits scored on it by a fairly sharp pointed instrument

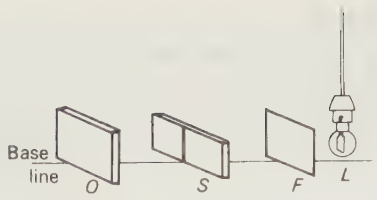


Fig. 23.22 A simple demonstration of Young's fringes.

such as a fine ballpoint pen. These slits are about 0.5 mm apart, and set parallel to the line of the lamp filament. The distance from the lamp to the screen is not critical; between half a metre and a metre is generally satisfactory. An observing screen, O, is placed about two metres from S. The best kind of screen is a ground- or etched-glass one, or one of translucent plastic or paper. In a well-darkened room, with precautions taken to prevent any appreciable stray light reaching the eye, one can see a number of parallel bright bands on the screen, tinged with spectral colours. If a colour filter, F (red is a good one to start with) is placed between L and S, the bands become more discrete. When the screen is moved closer to the twin slits, S, the bands become more closely spaced. The bright bands on the screen are regions of constructive interference between the diffracted waves emerging from the slits, and the dark spaces between them are regions of destructive interference. With the screen at a fixed distance from S you should try the effect of using a different colour filter, say green. This will be found to alter the spacing of the bright bands observed on the screen: they are more closely spaced with green light than with red. This we interpret as indicating that green light has a shorter wavelength than red. When a white light source is used the pattern on the screen is the combined effect of sets of bands having different spacings corresponding to the different colours present in white light. This explains the spectral colour tingeing of the bands when white light is used. You should also examine the effect of using different spacings between the two slits on the slide S. On a single slide one can rule three or four pairs of slits, with care, having spacings of about, say, 0.25 mm, 0.5 mm, 0.75 mm, and 1 mm. It is found that the smaller this spacing, the greater the spacing of the bright bands on the screen. These results are consistent with what is observed in a ripple-tank when one varies the relation between wavelength and source-spacing (see Fig. 23.20): the angular separation of the antinodal regions (regions of constructive interference) can be increased either by increasing the wavelength, or by reducing the distance between the sources.

We now proceed to find the precise relationship between the four variable parameters in this experiment:

d, the distance between the two slits,
x, the distance from the slits to the screen,
y, the distance between centres of adjacent bright bands on the screen,
λ, the wavelength of the light.

In Fig. 23.23 we show the geometrical essentials of the experiment. We imagine we are looking down on the experiment bench from above. The two slits are at A and B. O is a point midway between them and P is a point on the screen equidistant from the slits. The line OP is a useful axis for reference: we can call it the optical axis of the experiment. Now we imagine that the experiment has been set up in a symmetrical fashion, using monochromatic light, and with the source of light equidistant from A and B. In this way the wave vibrations at A and B can be assumed to be in phase with each other, and since the waves from these two secondary sources have had to travel equal distances to get to P, the vibrations at P must be in phase, so that there is constructive interference and hence a maximum of brightness at P. P therefore marks the central bright band of the pattern on the screen. Q is some other point on the screen. What is observed at Q depends on the distances AQ and BQ. If these distances differ by a whole number of wavelengths, then the wave vibrations arriving at Q will be in phase and there will be constructive interference: maximum brightness again. If the two distances differ by half a wavelength or three halves, or any odd number of half-wavelengths, then there will be destructive interference—darkness. The difference

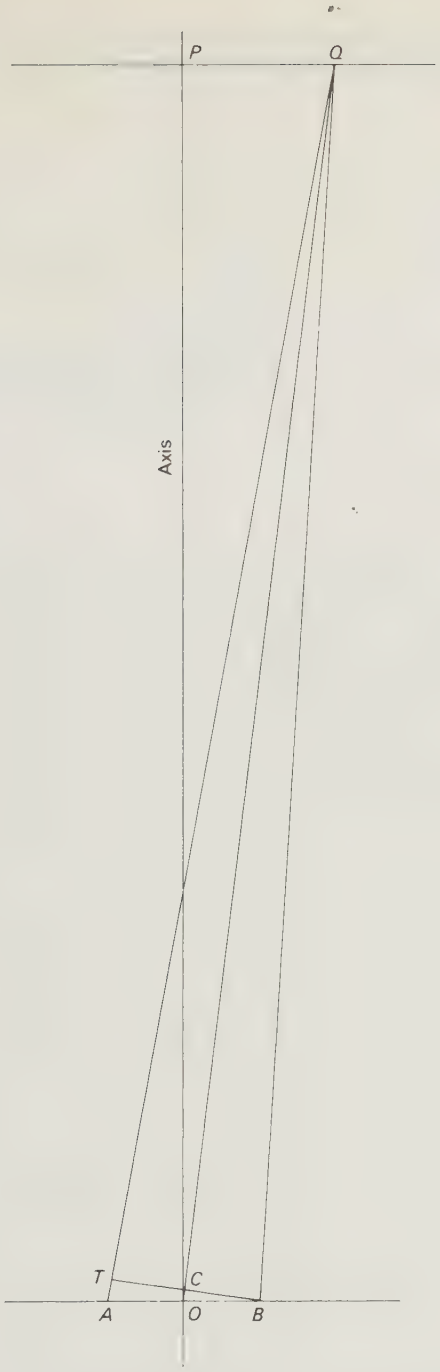


Fig. 23.23

between AQ and BQ is called the *optical path difference*, and if we place the point T on the line AQ so that $TQ = BQ$, then AT is the optical path difference. Thus the triangle TQB is an isosceles triangle and the construction line OQ is very nearly the bisector of the angle at Q, and hence the angle between OQ and BT, where they intersect, at C, can be taken to be a right angle. Thus the angle POQ equals the angle ABT (because in the right-angled triangle OCB, the angles BOC and CBO add up to a right angle, likewise the angles POQ and BOC add up to a right angle). Thus the right-angled triangles TBA and POQ are similar, since their smallest angles are also equal. Hence, in these triangles

$$\frac{AT}{AB} = \frac{PQ}{OQ}$$

but in practice the angle POQ is so very small that it will make no appreciable difference if instead of OQ we write OP in the equation above: they are both enormously long in comparison with PQ. So we write

$$\frac{AT}{AB} = \frac{PQ}{OP}$$

or, using the symbols, and remembering that AT is the optical path difference, we can write

$$\frac{\text{path difference}}{d} = \frac{y}{x}. \qquad (23.2)$$

Now imagine that we number the bright bands starting with the central bright band at P as zero. For the nth bright band the path difference is $n\lambda$, and for the $(n + 1)$th band it is $(n + 1)\lambda$. Thus we can write

$$\frac{n\lambda}{d} = \frac{y_n}{x} \quad \text{and} \quad \frac{(n + 1)\lambda}{d} = \frac{y_{n+1}}{x}$$

thus the distance between the adjacent bands, call it w, is given by

$$w = y_{n+1} - y_n = \frac{x\lambda}{d}.$$

In the experiment, therefore, by measuring d, x and w, we can calculate the wavelength of the light—at

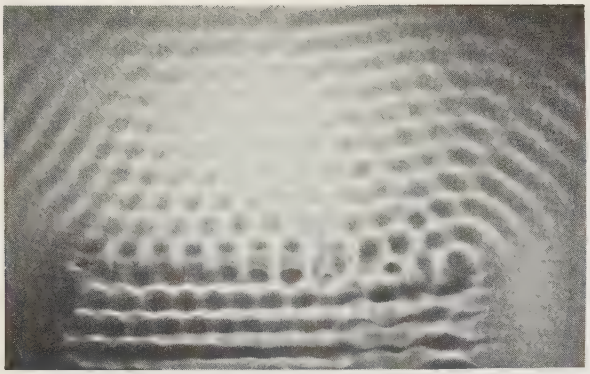


Fig. 23.24 Interference of diffracted ripples from an array of regularly spaced apertures. (Photograph by E.J.W.)

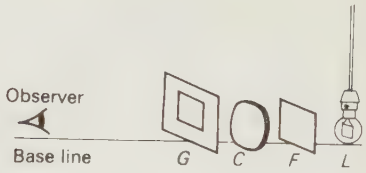


Fig. 23.25 A simple experiment for observing the spectra produced by a diffraction grating.

least approximately. Although it is possible to refine the experimental arrangement to improve the accuracy of such a measurement, one will always be limited by the fact that not many bands can be seen clearly, their spacing is small, and there is bound to be some uncertainty in locating the precise point of maximum brightness in a band. For greater precision one would use a *diffraction grating*.

There are two main classes of diffraction gratings —reflection gratings and transmission gratings. The action of a transmission grating, which is simply an array of equally spaced similar apertures in a barrier, is demonstrated in a ripple-tank. Figure 23.24 shows the result of placing a barrier containing a number of equally spaced slits in the path of a train of continuous straight ripples. The secondary waves emerging from the slits interfere in such a way as to produce not only a train of very nearly straight ripples travelling in the same direction as the original ones, but also trains of very nearly straight ripples travelling in other directions, on either side of the central train. Thus the energy flux associated with the waves is concentrated into certain specific directions including the original direction. If the number of slits is increased the emerging ripples become more nearly exactly straight, and thus the specific directions of energy flux become more sharply defined.

For a simple demonstration of the action of an optical diffraction grating one can use an inexpensive plastic replica grating, which is manufactured in much the same way as a gramophone record. In Fig. 23.25 the lamp L, a clear glass straight filament lamp like the one recommended earlier for the Young's slits experiment, is placed at the focus of a converging lens, C. F is a colour filter (red is good for a start). This arrangement produces a parallel beam of nearly monochromatic light. The beam passes through the grating at G. If you position your eye on the same horizontal level as the grating you will observe not only plenty of light coming straight through the grating, but also light emerging obliquely, on both sides of an axis perpendicular to the grating. On removing the colour filter F and allowing the full white light through the grating you see that light of different colours is channelled into different directions: thus spectra are produced by the grating. A good way to observe these spectra is to use a second converging lens, D, as shown in Fig. 23.26, to converge the diffracted beam and form an image on a small screen, S, placed in the focal plane of this lens. It is generally possible to see more than two spectra either side of the axis. The first pair of spectra appearing either side of the axis are known as the *first order spectra*, the next pair as the *second order*, and so on. It is not difficult to find the relationship between the wavelength of the light and the direction, defined by the angle between the axis and the direction of the diffracted beam, in which the light is channelled. Figure 23.27 shows a few slits in the

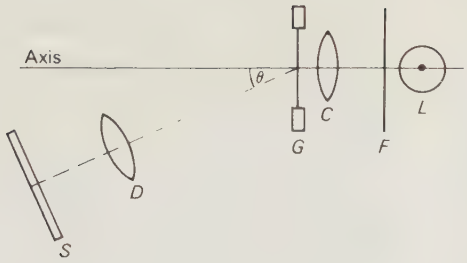


Fig. 23.26

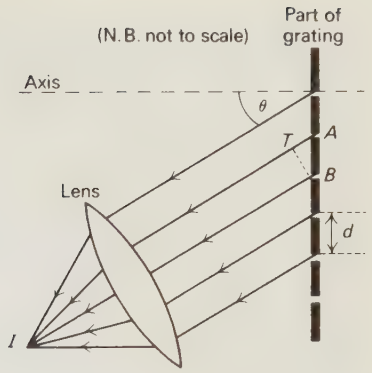


Fig. 23.27

grating, and the optical paths from these slits, through the converging lens, to the point I where an image is formed. For constructive interference to occur (at I) between the waves travelling along all these optical paths, the path difference between adjacent paths must be a whole number of wavelengths. The angle between these paths (and therefore between the direction of the diffracted beam) and the axis is θ, and is equal to the angle TBA in the diagram. In the right-angled triangle TBA therefore

$$AT = AB \sin \theta.$$

Now AB is the distance between adjacent slits in the grating, often called the grating element, and we use the symbol d for it. Since we must have AT, the path difference, equal to $n\lambda$, where n is a whole number, we can write

$$n\lambda = d \sin \theta. \qquad (23.3)$$

This equation defines the angles θ at which light of wavelength λ will emerge from the grating. With an accurately made grating, for which the value of d is known to a high order of accuracy, and a spectrometer to measure θ with high precision, one can make very exact determinations of the wavelength of light from a source.

We have said that, in general, you can observe more than one pair of spectra when using a grating. What determines how many spectra you see? We can answer this by analysing a typical practical situation. A certain type of plastic replica grating in common use at present has 15,400 rulings to the inch, as specified by the manufacturers. Since one inch equals 2.54 cm, to three significant figures, this means that the size of the grating element is given by

$$d = \frac{2.54}{15{,}400} \text{ cm} = 1.65 \times 10^{-4} \text{ cm}$$

$$= 1650 \text{ nm}.$$

Let us now find out what happens to monochromatic light of three particular colours when it goes through this grating—red light for which $\lambda = 650$ nm, yellow light for which $\lambda = 580$ nm, and violet light for which $\lambda = 450$ nm. (Red and violet light with the wavelengths quoted mark approximately the visible limits of the spectrum of white light obtained with a grating like the one described.) To find the values of the angle θ at which the light will appear in the different orders of spectra we use the equation, derived earlier,

$$n\lambda = d \sin \theta,$$

putting $n = 1$ for the first order, $n = 2$ for the second order spectrum, and so on. Table 23.1 below shows the result of doing this. The calculations predict that we should be able to see the complete first order and second order spectra, and part of the third order one. But it is not possible to see, for instance, the yellow in

the third order spectrum. Why not? The mathematically minded will say that this is because $\sin\theta$ cannot be greater than unity. But it is more convincing, perhaps, to see how this works out geometrically. Refer back to Fig. 23.26. The optical path difference between adjacent paths is AT. As the angle θ is increased, TBA remaining all the time a right angled triangle, AT becomes greater. When θ is very nearly 90° the length of the side AT is very nearly equal to AB which equals d, the grating element. Thus d is the upper theoretical limit for the optical path difference, and if this is less than three whole wavelengths for yellow light, then no yellow light will be observable in the third order spectrum.

Table 23.1

	Violet		Yellow		Red	
	$\sin\theta$	θ	$\sin\theta$	θ	$\sin\theta$	θ
$n = 1$	0.290	17°	0.374	22°	0.419	25°
$n = 2$	0.580	35°	0.748	48°	0.839	57°
$n = 3$	0.870	60°	1.12	–	1.26	–
	1.10	–				

For the most accurate spectroscopic work reflection gratings are generally used. These are made by very high-precision machines which rule minute parallel grooves on a highly polished metal mirror surface. The grooves are effectively non-reflecting, and the narrow strips of untouched metal between reflect the incident light, acting as sources of secondary waves, just as the adjacent apertures in a transmission grating act as sources of secondary waves.

The equally spaced layers of ions in a perfect crystal can act as the elements of a kind of reflection grating for very short wavelength electromagnetic waves, namely X-rays. This is the basis of X-ray crystallography and is described further in Unit 4.

23.6 COHERENCE

The relative ease with which we can observe interference effects with ripples on the surface of water or even with sound contrasts sharply with the care we have to take to observe analogous effects with light. This does not arise solely because the wavelength of light is so much shorter but because there is also a problem of phase.

The two dippers in a ripple-tank vibrate in phase and remain in phase. But suppose they were entirely independent. It would be unlikely that they would have precisely the same frequencies. If they started to vibrate in phase they would be continuously drifting in and out of phase. No stable interference pattern would then be seen although we might well observe a moving pattern.

Consider now two radio transmitters, also independent of one another and also starting their transmissions in phase. The same thing will occur and we shall find it very difficult to observe the moving pattern of interference. For any equipment we use to do this is likely to average the joint signal over an appreciable period of time and the average phase shift will be zero.

Light from any ordinary source, for example, a hot filament, is made up of a sequence of very short wave trains emitted first by one excited atom and then by another. Phase shifts will occur continuously and at random. We can expect the phase to remain fixed for no longer than it takes an atom to radiate a single wave train—say 10^{-8} s. We have no chance at all of observing interference patterns unless we take great care with the design of our experiment. Then we can obtain *coherent* beams in which the phase shifts occur simultaneously by allowing part of the beam of light to divide into two (e.g., by diffraction at two narrow slits placed side by side) and then to re-combine. Light sources (lasers) which do produce coherent beams are available and will be discussed in Unit 8.

PROBLEMS: Unit Five

5.1 In his famous treatise on motion, published in 1638, *Dialoghi delle Nuove Scienze*, Galileo considers how the period of a simple pendulum is related to its length. Study the following extract.

> Salviati: As to the times of vibration of bodies suspended by threads of different lengths, they bear to each other the same proportion as the square roots of the lengths of the thread; or one might say the lengths are to each other as the squares of the times; so that if one wishes to make the vibration-time of one pendulum twice that of another, he must make its suspension four times as long. . . .
>
> Sagredo: Then, if I understand you correctly, I can easily measure the length of a string whose upper end is attached at any height whatever even if this end were invisible and I could see only the lower extremity. For if I attach to the lower end of this string a rather heavy weight and give it a to-and-fro motion, and if I ask a friend to count a number of its vibrations, while I, during the same time-interval, count the number of vibrations of a pendulum which is exactly one cubit in length, then knowing the number of vibrations which each pendulum makes in the given interval of time one can determine the length of the string. Suppose, for example, that my friend counts 20 vibrations of the long cord during the same time in which I count 240 of my string which is one cubit in length; taking the squares of the two numbers, 20 and 240, namely 400 and 57 600, then, I say, the long string contains 57 600 units of such length that my pendulum will contain 400 of them; and since the length of my string is one cubit, I shall divide 57 600 by 400 and thus obtain 144. Accordingly I shall call the length of the string 144 cubits.
>
> Salviati: Nor will you miss it by as much as a hand's breadth, especially if you observe a large number of vibrations.

Now answer these questions.

a) Write your own, up-to-date version, in the most concise possible form, of Sagredo's instructions on how to find an unknown height by means of two pendulums.

b) Salviati's final remark can be understood as a statement of the accuracy obtainable by this method. If the error cannot exceed one hand's breadth in a total length of 144 cubits, what is the maximum percentage error implied by this (approximately!)? Do you think the claim is justified?

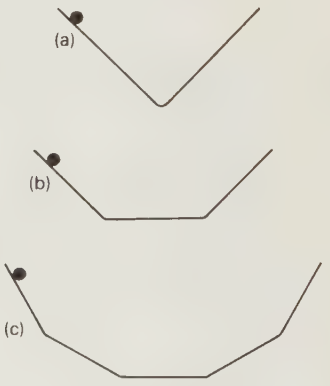


Fig. P5.1

5.2 Each of the sketches in Fig. P5.1 represents a frictionless track made up of straight sections, upon which a particle can slide. At the junctions between the straight

portions the angle can be imagined to be "radiused", but the curved bits can be assumed negligibly small compared with the straight bits.

Sketch the velocity-time graph for a particle released from the right-hand end of each track. Take velocities to the right as positive, and velocities to the left as negative.

It should become quite clear to you, after doing this, that only a track without discontinuities can give rise to a velocity-time graph without discontinuities.

5.3 Figure P5.2 shows one cycle of the motion of a simple harmonic oscillator. Using the arbitrary time-scale, answer the following questions.

a) At what times is the speed zero?
b) At what times is the speed greatest?
c) At what times is the acceleration zero?
d) At what times is the acceleration greatest?
e) At what times do you estimate the speed will be half its maximum value?

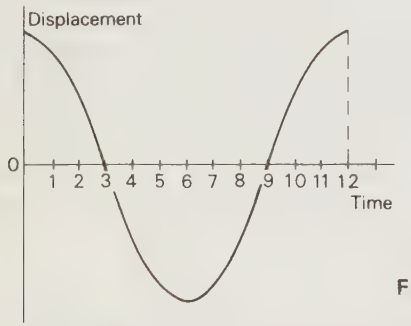


Fig. P5.2

5.4 If the amplitude of a simple harmonic oscillator is halved, what change, if any, occurs in:

a) the period,
b) the maximum velocity,
c) the total energy,
d) the maximum acceleration?

5.5 A simple harmonic oscillator moves from its central equilibrium position to its maximum displacement of 5 cm in a time of 1 second. Calculate:

a) the period,
b) the angular frequency, ω,
c) the maximum velocity,
d) the maximum acceleration.

5.6 An object whose mass is 2 kg is attached to the lower end of a tension spring and held at rest in such a position that the spring is straight but, as yet, unstretched. When released from this position the object oscillates vertically up and down with an amplitude of 30 cm, about the equilibrium position (in which it finally comes to rest). To answer the following questions, neglect damping, and use the principle of conservation of energy.

a) Calculate the amount of gravitational potential energy lost by the object when it descends from its initial position to the equilibrium position.
b) Calculate the increase in elastic potential energy stored in the spring during the same descent.
c) Account for the difference between the amounts of energy in (a) and (b).
d) What is the total energy of the oscillator?

5.7 Any experienced walker knows that the style of walking which uses the least energy concerns choosing the correct rhythm of leg movements, and thinking always of swinging one's leg from the hip. Sit on the edge of a table or other surface which is high enough for you to dangle one leg freely. Let it swing like a pendulum, from the hip: ideally get someone else to keep your leg swinging by repeated light pushes while you keep your leg completely relaxed. Note the approximate period of swing. Why, do you think, is it a good idea to move your legs with this rhythm when walking? Do you walk this way?

5.8 A simple pendulum can be forced into oscillation by moving its point of support very slightly back and forth horizontally with a regular rhythm.

Describe how you would expect the pendulum to behave when the frequency of the movement of the support is varied from a low to a high value. Consider, in your discussion, the frequency, amplitude, and relative phase, of the motions of the support and the pendulum bob. You can get a feel for the situation by tying a small heavy object to a length of cotton (about 1 ft, or 30 cm, is suitable) and using your hand as the support.

How it is possible, in this situation, for the pendulum to oscillate at frequencies which are different from its natural frequency?

5.9 Figure P5.3 shows a massive rectangular box, which can be assumed to be perfectly rigid, supported on four similar springs. How many modes of vibration are possible in this system? The mode that probably first comes to mind is the one in which the box simply bobs up and down, all springs being compressed simultaneously and relaxed simultaneously. But the box can also oscillate in a pitching mode, with springs A and B being compressed while C and D are being relaxed, and vice versa. What

other modes do you think are possible? Attempt to complete a table on the lines suggested below.

Description of mode	What the springs are doing at one instant			
	A	*B*	*C*	*D*
Bobbing	All being compressed			
Pitching	Compressed		Relaxed	

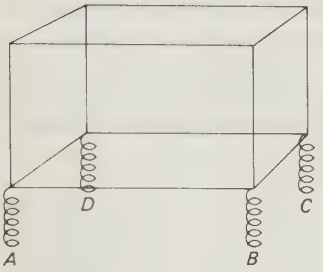


Fig. P5.3

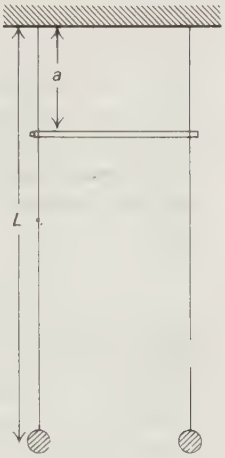


Fig. P5.4

5.10 This can be done as a home experiment or, alternatively, reasoned out in your imagination.

Two exactly similar simple pendulums are set up side by side and linked, not far from the top, by a light rigid strut (a straw, or thin strip of wood, for example). The system has two distinct normal modes of oscillation (i.e., modes in which each pendulum oscillates with simple harmonic motion of constant amplitude, if one neglects damping). What are these normal modes? How does the frequency of each normal mode depend upon the length L and the distance a?

If one of the pendulums is stationary and the other is drawn aside and released, what is the subsequent motion of the system? Illustrate your answer by sketch-graphs of displacement and time.

(*Note:* The motion of the two pendulums is assumed to be entirely in one plane: the plane of the diagram in Fig. P5.4.)

5.11 In Fig. 23.7 you see a sequence of shapes representing the vibration of a string in a complex mode. It was stated (Section 23.2) that this complex mode is the result of superposing the two lowest-frequency normal modes of vibration. How this superposition works can be understood by following the instructions here. In Fig. P5.5 the two faint curves represent the displacement of the string due to (a) vibration in the fundamental mode and (b) vibration in the second harmonic mode, whose frequency is twice the fundamental frequency. The amplitude of vibration in mode (a) has been chosen (arbitrarily) to be exactly twice the amplitude of (b). The first picture shows, as a thick line,

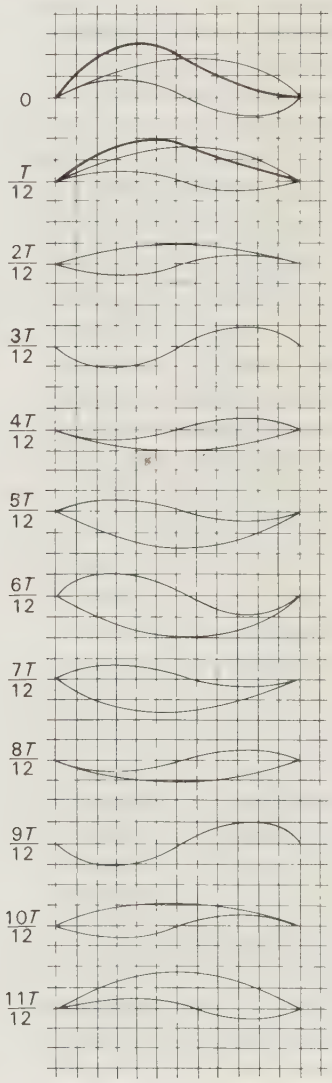


Fig. P5.5

the result of adding together the displacements due to vibrations (a) and (b). At time $T/12$ later, where T is the period of vibration of the fundamental mode (a), the curves are as in the second picture. In the third and subsequent pictures the result of superposing (a) and (b) is not shown. Study the pictures and sketch in pencil the resulting profile of the string in each picture, and check your final set of profiles with those shown in Fig. 23.7.

5.12 In this exercise you will plot a sequence of four wave-profiles, each with an amplitude of 1 cm, on a sheet of graph paper.

In Section 21.3 it was explained that a travelling sinusoidal wave can be described by the expression

$$y = a_0 \cos \omega\left(t - \frac{x}{v}\right).$$

Suppose we have such a wave whose amplitude is 1 cm, wavelength 2 cm, and frequency 0.25 s.

a) Calculate the transverse displacement, y, at time $t = 0$, for values of x in the range $x = 0$ to $x = 4$ cm, increasing by intervals of 0.25 cm. Plot these displacements against x near the top of a sheet of graph paper.
b) Repeat the calculations for the same values of x but at times $t = 0.5$ s, 1 s, and 2 s. Plot the three wave-profiles below the previous one.
c) Use the graphs to verify the relationship: frequency × wavelength = wave velocity.
d) Describe (but do not draw) in what way the results would be different if the expression for displacement

$$y = a_0 \cos \omega\left(t + \frac{x}{v}\right)$$

were used instead.

5.13 The expression for the speed of longitudinal waves along a rod or wire of elastic material, derived in Section 21.2, can be used to find the speed of longitudinal waves along a "column" of gas enclosed in a tube if, instead of Young's modulus of elasticity, E (which is applicable only to solid materials), one substitutes the bulk modulus of elasticity of the gas. Given that the bulk modulus for air is about 1.4×10^5 N m^{-2}, and the density about 1.25 kg m^{-3}, at ordinary room temperature, estimate the length of a pipe which, open at both ends, will have a fundamental frequency of 260 Hz (about middle C). (The fundamental mode of oscillation of the air-column in a pipe open at both ends occurs when the length of the pipe is approximately equal to half the wavelength of the stationary wave.) Problems 5.20 and 5.21 involve further discussion of the vibrations of air-columns, with particular reference to organ pipes.

5.14 A heavy rope, whose mass per unit length is 0.1 kg per metre, is stretched nearly horizontally with a tension of 40 N. What is the speed at which a transverse sinusoidal wave will travel along the rope?

If this rope is hanging freely and vertically from a fixed support, and a transverse wave-pulse is sent up the rope from the bottom end, will the pulse travel at constant speed? Explain your answer.

5.15 The top string (E) of a steel-string guitar is tuned to vibrate, in its fundamental mode, at a frequency of 330 Hz. The length of the string is 64 cm. If the mass per unit length of the string is 0.5 gm per metre, calculate the tension (in newtons) needed to make it vibrate at this pitch. You may need to refer to Section 23.2 (on stationary waves) and to Section 21.4 (speed of waves on a stretched string).

5.16 The speed of sound in air can be measured by means of the apparatus sketched in Fig. P5.6. The loudspeaker, L, is fed with the signal from an audio signal generator so that the diaphragm vibrates at a constant frequency. Fine dust (for instance, lycopodium powder) is put in the tube. This dust is whipped up in places where the air is vibrating strongly, and the small whirls of dust locate the antinodes of the longitudinal stationary wave which is set up when the tube is tuned to resonate at the frequency of the loudspeaker by adjusting the position of the piston, P.

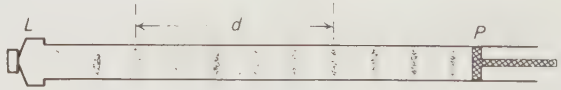


Fig. P5.6

Suppose that, in a particular experiment, the antinodes are as shown in the diagram below, labelled A. The distance d is found to be 1.65 m when the frequency is 500 Hz. Calculate the speed of travelling sound waves in air from this information.

5.17 Here are two phenomena you may have observed, or perhaps you will observe one day, which you should try to explain by applying your understanding of what happens when waves pass the edge of a barrier or go through an aperture.

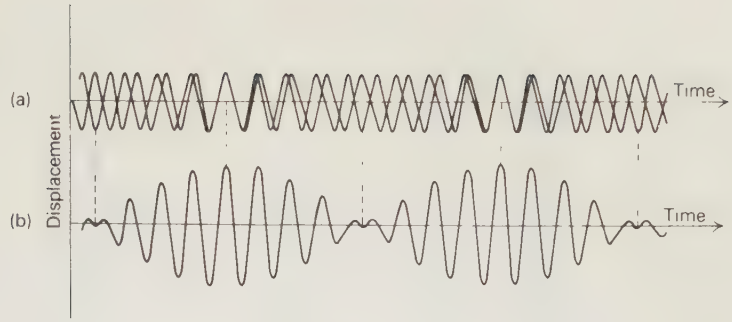


Fig. P5.7

a) Suppose a band of musicians is playing in the main street in a town and you are in a side street, so placed that you can hear the music but cannot see the band. The sound will seem muffled, and the most prominent sound is likely to be that of the bass drum, and other low-pitched instruments. You may not hear the high-pitched piccolos at all until you are actually in sight of the band.

b) When the doors of a concert hall are left open, a person outside and some distance away from the hall hears the bass notes much more clearly than the treble notes.

5.18 The interference of sound waves from two sources is effectively demonstrated by having two small loudspeakers connected in parallel to a suitable audio signal generator. These are set up in an open field, well away from any walls which might reflect the sound, on a still day, about a metre above the ground. With a spacing of about one metre between the two speakers, and the signal generator set at a frequency of about 1 kHz, one can easily locate the places where there is destructive interference by walking about in the region in front of the two speakers: these are places where the sound vanishes almost entirely. If a large group of people is prepared to take part in the experiment they can be told to walk around and stop at a place where they hear minimum sound. If they are well spaced out from each other, they will be observed to form ranks along the lines of nodes, and the interference pattern is thus convincingly made visible. Better still, a photograph of the crowd from a high vantage-point makes a permanent record of the pattern.

You can predict the form of this pattern as follows. Taking the speed of sound in air as about $330\ \mathrm{m\ s^{-1}}$ calculate the wavelength of the sound waves from a source whose frequency is 1 kHz. Using a suitable scale (say 10 cm on paper represents 1 m), draw on a sheet of paper, near the centre of one long edge, two points to represent the sources set 1 m apart. Represent the successive compressions of the sound wave coming from the sources by a series of concentric semi-circles, drawn to scale. The points where the circles intersect are places where constructive interference occurs: join these up by free-hand drawn, full lines, "fanning out" from between the sources. These are the lines of antinodes. Estimate by eye the lines of nodes (destructive interference regions) between the lines of antinodes, and draw these as broken lines.

5.19 When two musical notes of nearly, but not quite, the same frequency are sounded simultaneously, the listener is aware of a kind of throbbing, a periodic rise and fall in loudness of the sound. The sensation, particularly to a musical ear, can be very unpleasant. The effect is best demonstrated using two audio signal generators, each with its own loudspeaker. The closer the frequencies of the two notes, the slower is the beat-rate, and when the two notes are exactly in unison, no beats are heard: but a perfect unison is almost impossible to achieve with two independent audio signal generators. If the wave-form of the sound, when beats are occurring, is displayed on an oscilloscope via a microphone, the trace will appear as shown in Fig. P5.7. The rapid periodic variation is the frequency of the note heard, and the slowly varying "envelope" of the trace (shown by a broken line) is the periodic rise and fall in amplitude which gives the throbbing effect.

The phenomenon can be explained mathematically. Suppose the displacement of the microphone diaphragm due to the wave arriving from one of the sound sources is represented by $x_1 = a_0 \cos 2\pi f_1 t$, where f_1 is the frequency of the source. Write an expression for the displacement (same amplitude) due to the other source, frequency f_2. The resultant displacement will be given by: $x = x_1 + x_2$.

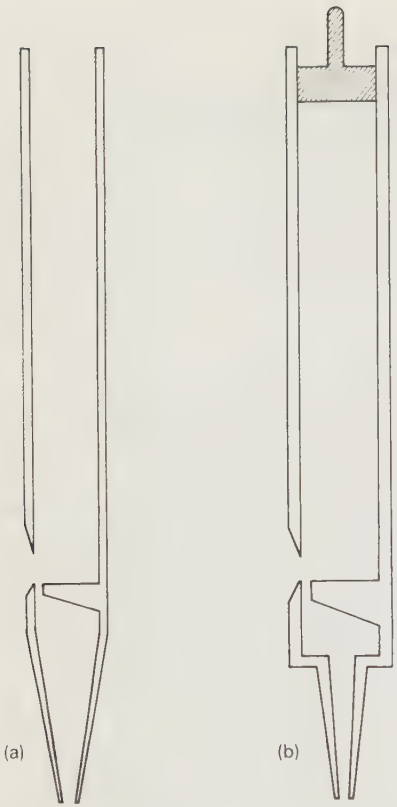


Fig. P5.8

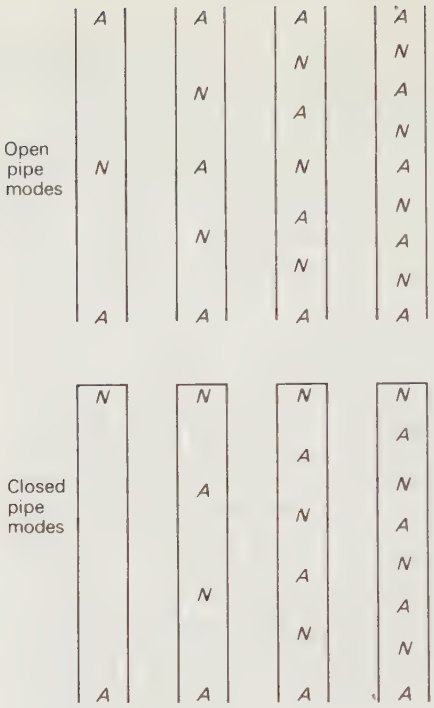


Fig. P5.9

Write an expression for x, and simplify it using the trigonometrical relation:

$$\cos\theta + \cos\phi = 2\cos\left(\frac{\theta + \phi}{2}\right)\cos\left(\frac{\theta - \phi}{2}\right).$$

The expression now contains the product of two cosine terms: one of the terms represents a rapidly varying quantity and the other a slowly varying quantity. If you made a sketch-graph of this expression it would appear as in the figure.

The rapidly varying term represents the frequency of the note heard by the listener: what is this frequency, in terms of f_1 and f_2? The slowly varying term gives the beat frequency. This frequency is the number of times per second that the appropriate cosine term has the value zero. Knowing this, express the beat frequency in terms of f_1 and f_2.

Three problems for musicians

5.20 One class of organ pipes, called flue pipes, are basically whistles. A jet of air emerging from a narrow slit impinges on a knife-edge setting up eddies in the air-stream which excite longitudinal vibrations in the column of air enclosed in the pipe. A precise description of the motion of the air is very complex, but a knowledge of stationary waves enables one to understand how the length of the pipe is related to the pitch of the musical note it produces.

In the pipe (a) shown in Fig. P5.8, an open flue pipe, the fundamental mode of vibration of the air-column is one in which there is an antinode at the bottom end and an antinode just outside the open end at the top with a node in the middle. This can be regarded as a stationary wave one half-wavelength long. In the pipe (b), a closed pipe, the fundamental mode has a node at the top and an antinode at the bottom end.

The lowest note on an early organ (sixteenth and seventeenth centuries) was generally the C two octaves below middle C. Now, if two notes are an octave apart, this means that their frequencies are in the ratio 2:1. Given that the pitch of middle C is about 256 Hz, estimate

Table P5.1

Note on piano	Frequency (concert pitch), to nearest Hz	Note on piano	Frequency (concert pitch), to nearest Hz
C_4 (middle C)	262 Hz	C_6	1046
$C_4\#$	277	$C_6\#$	1109
D_4	294	D_6	1175
$D_4\#$	311	$D_6\#$	1245
E_4	330	E_6	1319
F_4	349	F_6	1397
$F_4\#$	370	$F_6\#$	1480
G_4	392	G_6	1568
$G_4\#$	415	$G_6\#$	1661
A_4	440	A_6	1760
$A_4\#$	466	$A_6\#$	1865
B_4	494	B_6	1976
C_5 (octave above middle C)	523	C_7	2093
$C_5\#$	554	$C_7\#$	2218
D_5	587	D_7	2349
$D_5\#$	622	$D_7\#$	2489
E_5	659	E_7	2637
F_5	698	F_7	2794
$F_5\#$	740	$F_7\#$	2960
G_5	784	G_7	3136
$G_5\#$	831	$G_7\#$	3322
A_5	880	A_7	3520
$A_5\#$	932	$A_7\#$	3729
B_5	988	B_7	3951
C_6	1046	C_8	4186

the length of the longest open and closed pipes on such an organ. Take the speed of sound in air as $320\ m\ s^{-1}$. Would such an organ fit into the room of a house?

5.21 In Problem 5.20 the fundamental *modes* of oscillation of the column of air in an organ pipe were described. Higher frequency modes are possible and they can usually be excited in a flue-type organ pipe, and in wind instruments (for example, the recorder, or flute à bec) which work on a similar principle, by overblowing, that is, by blowing harder than normal. The higher modes of oscillation in open and closed pipes are represented symbolically in Fig. P5.9, where N signifies a node and A an antinode. If the frequency of the fundamental mode is f_0 in each case, what are the frequencies, in terms of f_0, of the higher modes, in the two classes of pipe?

If you are a musician you may be interested to know what musical notes are produced when an air-column vibrates in these higher modes. Suppose each type of pipe has a fundamental frequency corresponding to middle C (262 Hz for concert pitch). Use Table P5.1 to find the notes of the overtones of the pipes, and, if you can, play the *nearest* notes on a piano and hear what they sound like.

You will probably be wondering why the frequencies which you have calculated do not agree exactly with those for the notes C_4, C_5, G_5, C_6, E_6, etc. listed in the table. There are two reasons: firstly, the frequency of middle C (concert pitch) is not exactly 262 Hz, it is 261.63 Hz, because the standard of pitch is taken as A_4, with a frequency of precisely 440.00 Hz. The second reason is that the piano has what is called a *well-tempered scale*, so that music played on it will sound equally agreeable whatever key the music is in. There is no room for further explanation here. If you are interested, refer to one of the many books written on the science of music, for instance, *The Physics of Musical Sound* by J. J. Josephs (Van Nostrand, Momentum Books No. 13: 1967). Chapter 5 of this book is particularly relevant—"Musical scales, temperament, and tonality."

5.22 This is more of a practical exercise than a problem. It is instructive to repeat an experiment which has been performed countless times since the era of the ancient Greek philosophers, several of whom were interested in the harmonious sounds produced by plucked strings.

Any stringed instrument can be used for this experiment. Pluck a string at about its mid-point and listen to the sound. Then, with a finger of one hand, keep the finger-tip resting very lightly at the precise mid-point of the string and at the same time pluck the string with the other hand at a point about one quarter of the length along the string. As soon as the string has been plucked, remove the finger which was resting lightly at the mid-point. The string should continue to sound, and you will find that the note is exactly one octave above the previous note sounded. This is the first overtone. It may take some practice to get a clear, ringing note. Then touch the string lightly at a point precisely one third of the way along it, and pluck the shorter part of the string about halfway between the resting finger and the end of the string (that is, about one sixth of the way along the string). Now you should hear a note of higher pitch still, at a musical interval of exactly one major fifth above the previous note. This is the second overtone. Higher overtones can be excited in a similar manner, but they become successively more difficult to sound clearly. If you have an oscilloscope and a microphone, and a room free from stray noises, you can show that the frequencies of the fundamental and the overtones go in the ratio 1:2:3:4 and so forth.

The sequence of notes you have produced is known as the harmonic series. But note that the overtones of other musical instruments do not necessarily correspond to the harmonic series (see Section 23.2).

Referring back to Problem 5.21 you will see that an open pipe can produce all the notes of the harmonic series, whereas a closed pipe can give only the odd-numbered harmonics.

Finally, here is something to think about. Why are the strings of a stringed instrument (guitar, violin, piano, harp, for instance) not all of the same thickness? One of the clues to this lies in the formula for the speed of a transverse wave on a stretched string (Eq. 21.14). It is not a simple question to answer.

5.23 Light and X-rays are both said to be electro-magnetic waves.

a) What kind of experimental results suggest that both are wave motions?
b) What arguments, based on experimental results, can you give to suggest that they have very different wavelengths?
c) Mention any evidence that might suggest that light and X-rays are the same kind of radiation.

Oxford and Cambridge Examination Board.
Nuffield A-level Physics, (1970).

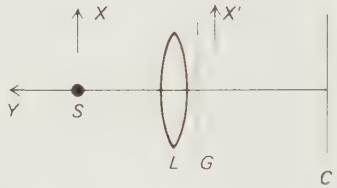


Fig. P5.10

5.24 A point source of light (S), a lens (L), and a grating (G) are arranged as shown (Fig. P5.10) to give a set of spectra on the screen (C). Explain what you think would happen to the appearance of the pattern on C in each of the following three experiments:

a) if S moved in the direction X,
b) if (with S in its original position) G moved in direction X_1,
c) if (with G in its original position) S moved in the direction Y.

Oxford and Cambridge Examination Board.
Nuffield S-level, Physics (1970.)

BIBLIOGRAPHY: Unit Five

Barber, N. (1969). *Water Waves*, Wykeham Publishing (London) Ltd.
A lively and thorough treatment of the subject which avoids any difficult mathematics.

Bishop, R. E. D. (1965). *Vibration*, Cambridge University Press.
A simple account from an engineering standpoint.

Compton and Doan. (1925). *X-ray Spectra from a Ruled Reflection Grating*, National Academy of Sciences, 598–601.
This article is reprinted in:
Classical Scientific Papers—Physics, Mills and Boon, starting on p.304.
This brief account of what must be the ultimate achievable with a reflection grating is well worth reading.

Feather, N. (1961). *Vibrations and Waves*, Edinburgh University Press.
This work treats the whole subject with a strong emphasis on the historical development of the concepts.

Holden, A. (1965). *The Nature of Solids*, Columbia University Press.
The behaviour of coupled oscillators is described, with reference to the behaviour of atoms in solid materials in Ch. 2 "Heat" and Ch. 3 "Heat Capacity".

Huygens, C. (1962). *Treatise on Light*, Dover Pub. Inc.
This translation is by Sylvanus P. Thompson, and the following are relevant to Ch. 5–21:
Ch. 1 On Light.
Ch. 2 On Reflexion.
Ch. 3 On Refraction.

Jaffe, B. (1961). *Michelson and the Speed of Light*, Heinemann.
This small book is written for the general reader, and includes brief accounts of the work of Fizeau and Foucault on the speed of light, as well as Michelson's own.

Josephs, J. J. (1967). *The Physics of Musical Sounds*, Momentum Books No. 13, Van Nostrand.
A good introduction for the reader interested in the science of musical instruments. A detailed description of the vibrations of Chladni's plate will be found here.

McKenzie, A. E. E. (1960). *The Major Achievements of Science*, Vols. 1 and 2, Cambridge University Press.
Chapter 12 of Volume 2 is concerned with the wave theory of light, with quotations from the works of Huygens, Young, and others.

Magie, W. F. (1963). *A Source Book in Physics*, Harvard University Press.
Excerpts from original writings. The very first section of the book contains selections from the writings of Galileo, in particular on the isochronism of the pendulum, pp. 17–19. In the section entitled 'Sound' are to be found chapters on Mersenne, Wallis, Sauveur, and Helmholtz; and in the section 'Light' quotations from Descartes, Huygens, Grimaldi, Newton, Young, Fresnel, Fizeau, Foucault, and many others, including Rowland who developed machines for constructing diffraction gratings.

Newton, I. (1962). *Opticks*, Dover Pub. Inc.
This is a modern printing of Sir Isaac Newton's original work. The 3rd Book, Part I (starting p. 317 in the Dover

edition) deals with Grimaldi's discovery of the diffraction of light by obstacles, and Newton's own investigations of the phenomenon.

(PSSC) (1968). *Physics*, Heath.
This is the textbook for the Physical Sciences Study Committee course in physics, and a clear, elementary, well-illustrated treatment of vibrations and waves will be found in the following chapters:

Ch. 14 Particle Model of Light.
Ch. 15 Introduction to Waves.
Ch. 16 Waves and Light.
Ch. 17 Interference.
Ch. 18 Light Waves.

Rogers, E. M. (1960). *Physics for the Inquiring Mind*, Oxford.
Chapter 10 "Vibrations and Waves" is a lively and lucid treatment of the subject.

Tricker, R. A. R. (1964). *Bores, Breakers, and Waves*, Mills and Boon.
A thorough and entertaining work on water waves which also avoids any difficult mathematics.

van Heel and Velzel. (1968). *What is Light?* Weidenfeld and Nicolson.
A survey of theories of the nature of light, attractively illustrated, suited to the non-specialist reader.

UNIT SIX
Electricity

24 Introduction
25 Steady Direct Currents
26 Power, Energy, and Electromotive Force
27 Some Useful Circuits
28 Electric Charge
29 Using Circuit Elements
30 Electromagnetism
31 Electromagnetic Induction
32 Alternating Currents

Chapter 24
INTRODUCTION

We can distinguish two kinds of reason for wanting to understand about electricity—theoretical and practical. Firstly, an understanding of basic electrical theory is necessary for understanding the structure of matter, because the model of matter which is introduced in this book consists of particles held together by electrical forces. The emphasis is on electrical forces, although other kinds of force will be met in Unit 7. Secondly, if we want to understand the uses to which electricity is put—heating, lighting, motive power, radio, computers, and a host of others, we must first understand the behaviour of electric currents in conductors and circuits. In our opinion the question of the electrical nature of matter is better dealt with after a practical introduction to electricity. This will be found in Unit 8 (Electricity and Matter).

In this chapter we shall meet the principles which are necessary for understanding how electricity is made to serve useful ends. As an introduction we shall raise some questions which are relevant to the present-day and the future technological world by considering two particular areas of development in electrical technology which have radically changed the pattern of life in Britain, and in other developed countries, in the twentieth century. We shall discuss very briefly the growth of electrical power in Britain, and the miniaturization of electronic circuits.

Michael Faraday discovered in 1831 that an electric current could be generated by the relative motion of a magnet and a conductor. In the years that followed, this effect, *electromagnetic induction*, was explored by many people with a view to designing a machine which could generate electrical energy in useful quantities. In 1886 John Hopkinson, engineer physicist, wrote the first treatise on dynamo design. By about 1900 electrical generators were in common use: multipolar generators driven by reciprocating steam engines, many producing several thousand kilowatts of electrical power. Electrical energy was all generated, at this time, by private companies, mostly for the sole purpose of supplying street lighting. In 1906 the total capacity of all electrical generating plants in the United Kingdom was about 1000 MW, and in that year the number of kilowatt-hour units sold was 530 million, but the average cost of generation and

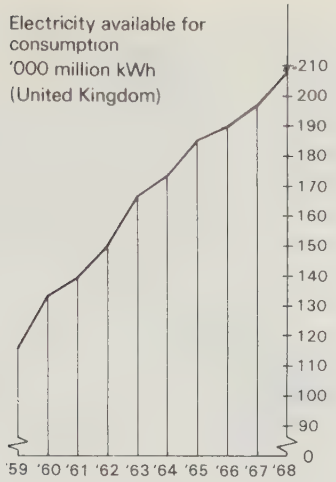


Fig. 24.1 Electricity available for consumption in the United Kingdom in 1970. (From *Britain 1970*, HMSO, 1970. Reproduced by permission of the publisher.)

distribution was high. In 1910, S. Z. de Ferranti, addressing the Institution of Electrical Engineers, expounded the idea of making an abundant and cheap supply of electrical energy available all over the country, mainly for the purposes of conservation of coal resources, to encourage production of home-grown food, and for better utilization of labour.

The demands imposed upon industry by the First World War spurred the government on to consider ways of interconnecting generating stations to economize on plant, coal, and other costs. In 1926 the Central Electricity Board was constituted. The Board set about getting the private electrical generating companies to co-operate in the establishing of a national network of transmission mains, originally called the Gridiron, but now known as the National Grid. At that time the total capacity in the country was about 3000 MW, with about 5000 million units being sold per year, at an average unit price of about £0.01. The grid was essential if generation and distribution of electrical energy was to grow in an economically viable fashion. This is for two main reasons: firstly, it is not practicable to store appreciable quantities of electrical energy—it must, nearly all, be generated as and when needed; secondly, it is an expensive and slow business to start up and shut down a generating plant in response to a rapidly fluctuating demand. (See Chapter 7.)

In the 1950s, the maximum capacity of a single generator was 60 MW: today it is 660 MW. Figure 24.1 shows the growth of electrical energy output from all generating plants in the United Kingdom. From this it can be seen that the consumption in 1968 was more than 200 000 million units, forty times what it was when the National Grid was proposed.

The present output of the electricity industry is summarized in Table 24.1, and the figures in it should be compared with the ones quoted earlier. (Table 24.1 and Fig. 24.1 are from *Britain 1970*, H.M.S.O., 1970. Crown Copyright. Reproduced by permission of the publishers.)

Table 24.1

Year: 1968	Total output capacity (MW)	Electrical energy sent out (million kWh)
Central Electricity Generating Board	44 372	168 817
Southern Scotland Electricity Board	3837	13 405
North Scottish Highlands Electricity Board	1851	3 659
Total	50 060	185 881
Steam and oil	44 574	161 273
Hydroelectric	1288	2 423
Pumped storage	760	848
Nuclear	3438	21 337

Now let us pose ourselves some questions which relate to the foregoing extremely brief summary.

a) Why is it not practicable to store electrical energy? (If it were practicable, then a national supply could be a much simpler system than it is now.)
b) How efficient are the processes of converting energy from steam and oil (the chief sources in the U.K.) into electrical form?
c) How much electrical energy is wasted in transmission from generator to consumer?

d) Why are nuclear power stations at present able to supply only about 12% of the electrical power in the U.K. (although, note, they have so far produced more energy than all the nuclear power stations in the rest of the world)? Will this percentage increase soon?

The answers to these questions largely determine the economics of the electrical power industry. We cannot answer them now, but what is presented in this book should enable you to see how to answer them, although you may need to go to other sources of information for facts and figures (see Central Office of Information, 1970; Tomalin, 1968).

At the other extreme from heavy current electrical engineering is the electronics industry which, in its brief life, has already been through several revolutionary stages of development.

"Micro-miniaturization" is a term which briefly sums up the revolutionary change that is happening in electronics. At the Stanford Research Institute in California there is talk of being able, in the next few years, to manufacture electronic circuits so small that some hundred thousand million components can be fitted into one cubic inch of space. At the time of writing there are integrated circuits readily available to the amateur in electronics in which several dozen components occupy scarcely more than a square millimetre on a flat chip of substrate material. The connecting leads to such a circuit, and the protective casing, occupy several hundred times as much volume of space as the circuit itself.

The period of rapid progress in making electronic circuits smaller, and more rugged, began in 1939, at the outbreak of the Second World War. Military electronic equipment, especially airborne radio and radar, had to be compact and highly reliable. Electronic vacuum tubes (valves) were made as small as possible, some of the smallest being known as *acorn valves*, which gives an idea of their size. After that war the change to transistors and other semiconductor devices marked a real turning point in the electronic revolution. The first transistors were manufactured in the early 1950s, and large sums of money were invested in the USA, as part of the "Defense Program," in developing the transistor and

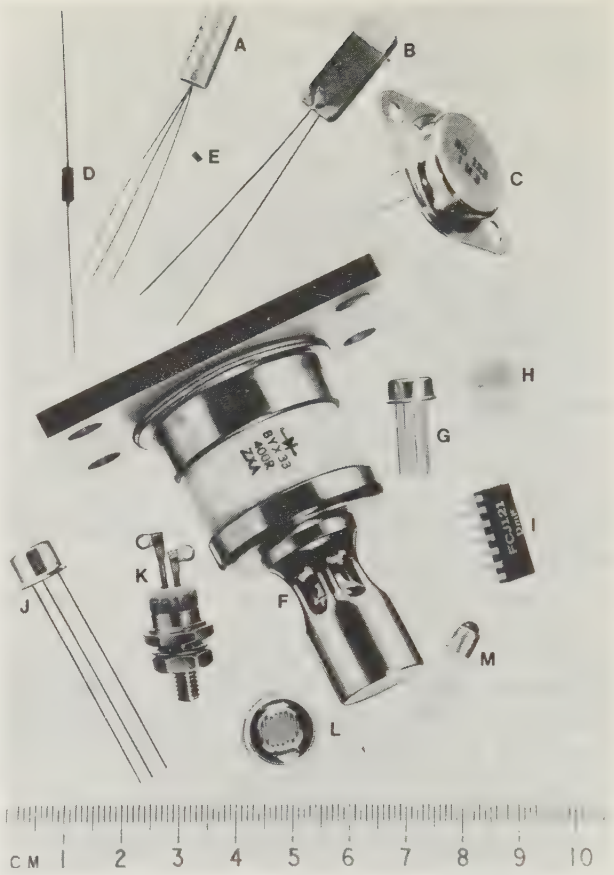


Fig. 24.2 A selection of semi-conductor devices. A, J, low power transistors; B, plate thermistor; C, power transistor; D, low power diode; E, leadless inverted device; F, 250 amp silicon rectifier diode; G, H, I, integrated circuits; K, thyristor; L, photoconductive cell; M, silicon phototransistor. (Courtesy of Mullard Ltd.)

other semiconductor devices. The advantages of transistors over vacuum tubes can be summed up briefly as follows:

a) They do not deteriorate with time, whereas vacuum tubes do.
b) They waste much less electrical power than vacuum tubes.
c) There is no warm-up period after switching on.
d) They are physically much more robust than tubes.
e) They can be made very much smaller than vacuum tubes which perform a similar function.

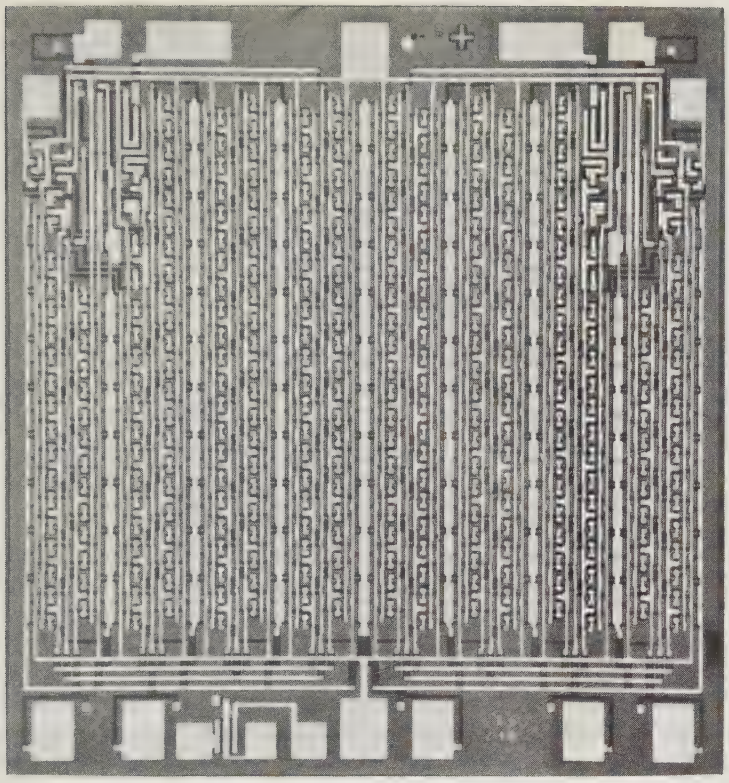


Fig. 24.3 Integrated circuits; this dual 80 bit static shift register is used in a computer. It is about 2.3 mm square. (Courtesy of Plessey Co., Ltd.)

This last advantage is perhaps the most siginficant in the electronic revolution. Whereas a vacuum tube comprises an evacuated glass envelope containing electrodes which must have a certain amount of space between them, a transistor consists of a minute chip of semiconductor material such as silicon with tiny blobs of other semiconductor material in contact with it. Figure 24.2 shows a collection of semiconductor devices.

If the decade starting in 1950 was the era of transistor development, then the following decade was the era of the printed circuit and latterly the integrated circuit. Figure 24.3 shows a much magnified photograph of an integrated circuit device used in a computer. The technological skill required to produce devices like this is of a very high order: materials of extremely high purity have to be used, and the utmost in microscopic accuracy in positioning the elements of the circuit. The circuits are etched by suitable acids, the parts which are not to be removed being protected by a lacquer. The process is known as photolithography. The protective lacquer is sensitive to ultraviolet light; if it is exposed to this light it "hardens" and resists the action of the special photographic developing solution which removes the lacquer which has not been exposed to light. A microscopic mask is prepared and placed in contact with the lacquered chip: this mask allows ultraviolet light to get at those areas which are to resist the action of the etching acid. The process of lacquering, masking, exposing to light, developing, etching, may be repeated several times to etch away successive layers of different semiconductor material which have previously been deposited on the substrate (Fig. 24.4).

Without the advent of miniature semiconductor devices, like transistors, it would have been impossible to develop computers beyond a very rudi-

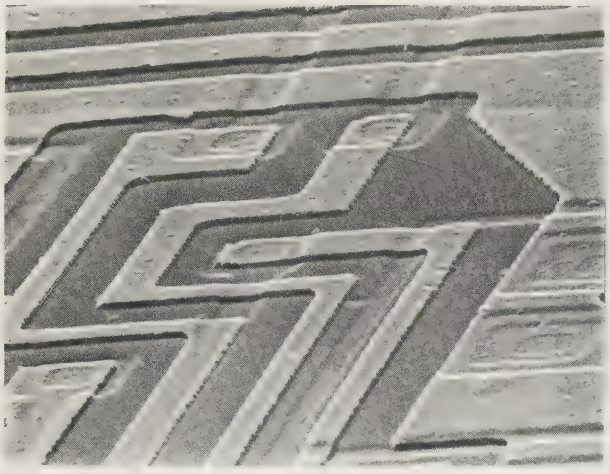


Fig. 24.4 A stereo-scan photograph at high magnification of a bi-polar chip. (Courtesy of Plessey Co., Ltd.)

mentary stage. In the early 1950s the primitive electronic computer in Britain known as ENIAC contained 18 000 vacuum tubes and the failure rate of these was such that it spent more time out of commission than it spent working. It was, by modern standards, a very limited machine. It filled a whole room: its modern counterpart can be made small enough to fit in your pocket. The effect of the computer revolution upon our lives need not be dwelt upon here. (See Handel, 1967.)

There is no doubt that electronic circuits will continue to get smaller and smaller, without losing reliability. Some present students of physics will have one day to use, or even design, such circuits. It is for this reason that the treatment of electric circuit theory in this chapter is planned in such a way as to lead the student to concern himself more with the function of an electric circuit element rather than its internal structure, and to think of ways of combining circuit elements to perform a desired function which cannot be performed by a single element alone.

Now to pose some particular questions which may arise from thinking about the miniaturizing of electric circuits.

a) How can ordinary electrical components, such as resistors and capacitors, be made microscopically small?
b) How does one deal with the problem of the heat which is produced when electric currents flow in these microscopically small circuits?
c) Some electrical components are bound to fail at some unpredictable stage during use: how reliable, therefore, can one expect a computer containing several million components to be?

This book, although it does not provide the full, detailed answers to these questions, nevertheless aims at giving you the basic ideas needed to tackle such questions intelligently.

In this unit we introduce the concepts of electric current, potential difference (voltage), resistance, power, charge, energy, capacitance, and the ideas needed to understand electromagnetism and alternating current theory. The problems at the end of the unit not only provide exercises in using the concepts, but describe important modern applications and thus aim at widening the range of your knowledge, not simply at testing and reinforcing your understanding.

Chapter 25
STEADY DIRECT CURRENTS

25.1 The simple circuit
25.2 Potential difference
25.3 Resistance
25.4 The resistance of wires

25.1 THE SIMPLE CIRCUIT

It is not possible to tell whether a metal wire is carrying an electric current merely by looking at it. There are, however, two detectable effects which we ascribe as being due to an electric current. Firstly, a heating effect. One of the main practical uses of electricity is to heat a wire. A nichrome alloy wire, heated by an electric current, forms the basis of all electric room heaters, immersion heaters and other heating elements. A tungsten wire, kept white hot by an electric current, in a low-pressure atmosphere of inert gas, is the filament of an electric lamp. The other effect, the magnetic effect, is less readily detected. Oersted announced in 1820 his discovery of the effect. He used a battery made of copper and zinc electrodes in sulphuric acid, a wire, and a pivoted magnetic compass needle (Magie, 1963).

Both effects have been employed in instruments for measuring the strength of an electric current. The hot-wire ammeter, in which the expansion of a wire when heated by the current was made to move a pointer over a scale, is now obsolete. The moving-coil and moving-iron meters both make use of the magnetic effect for their operation. The construction of the former instruments is discussed in detail in Chapter 30.

To investigate the properties of steady direct currents, one can do some experiments with simple circuits and ammeters. Suppose we have a circuit arranged as shown in Fig. 25.1, consisting of 3 torch lamps connected in series to 3 dry cells. A current, we say, flows when the switch S is closed, and we have a complete circuit. It ceases to flow, it appears (because the lamps go out), at the instant we open the switch S. So much will probably seem obvious to you. In our investigation of the behaviour of electric current, we ask two questions.

a) When the circuit is made, what is the current strength at different points in the circuit?
b) Does the current rise from zero to a steady value *instantaneously* when switch S is closed, and fall to zero instantaneously when the switch is opened? If not, how does the current change with time during the switch-on and switch-off processes?

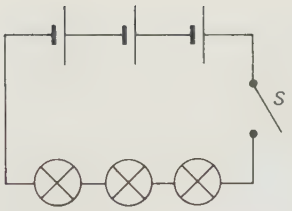


Fig. 25.1

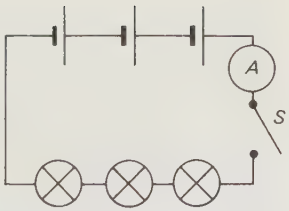


Fig. 25.2

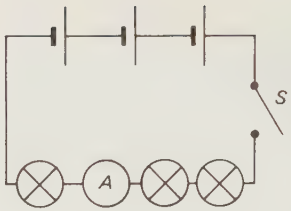


Fig. 25.3

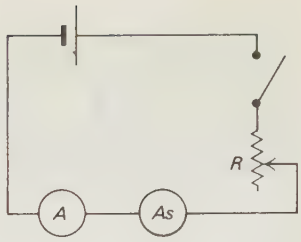


Fig. 25.4

The second question we shall leave until Chapter 31, when we have a fuller understanding of the nature of electric current. The first question can be tackled straight away. We switch on our circuit and observe that the lamps are steadily lit to equal brightness; we then break the circuit and for example insert an ammeter into the break as shown in Fig. 25.2. On switching on again, we notice that there is no observable difference in the brightness of the lamps, so we can assume therefore that the insertion of the ammeter has had negligible effect on the current supplied to the lamps. The ammeter registers a steady reading. Let us say it registers 0.25 amp. Suppose now we insert the ammeter in a different place, as in Fig. 25.3; we find that it shows precisely the same current, 0.25 amp. In fact, wherever the ammeter is inserted, it measures the same current. This suggests that the current at every point in a series circuit has the same value.

We may now see how it is possible to compare one ammeter with another, and hence to calibrate an uncalibrated ammeter against a standard. Figure 25.4 shows a circuit for doing this. Imagine that A_s is a standard ammeter and that A, having no scale yet, is to be calibrated against A_s. We have already shown by experiment that, whatever the current in a series circuit, it has the same value at all points in the circuit. So we can be sure that the same current is flowing through A as through A_s. By means of the rheostat R the current can be altered to whatever value we choose. Thus, R can be set so that the current has certain values as shown by A_s, and appropriate marks can be made on the scale of A.

The calibration of an ammeter then, is simple. But, one may ask, how was the standard ammeter, A_s, calibrated? By means of another standard ammeter, perhaps? If so, how was that ammeter calibrated? The question of how the fundamental standard is made is considered in Unit 7. For the time being we need only appreciate that, once given a standard ammeter of some kind, any other ammeter can be calibrated against it by the method described above.

An important practical point has not yet been mentioned: it concerns the *polarity* of the ammeter. A cell, or battery, has a positive terminal and a negative terminal. Likewise, an ammeter of the moving-coil variety, which is the type most commonly used for d.c. measurements, also has polarity. The polarity is marked by + and − signs near the appropriate terminals. For correct use, the ammeter in the circuit must have its positive terminal nearest the positive terminal of the cell or battery: i.e., it must be connected into the circuit with correct polarity. If the polarity is reversed the needle of the instrument moves in the opposite direction to normal, i.e., below zero, and a reading cannot be obtained, except if a centre-zero ammeter is being used. Such centre-zero instruments do not have the polarity marked.

These observations that the current at all points in a series circuit is the same, and that a moving-coil ammeter has polarity, both suggest that the invisible electric current is something which is flowing steadily through the wires and other parts of the circuit, and which has a definite direction of flow. Since nothing can actually be seen to flow, we cannot know what the direction of flow actually is. But it is useful to choose,

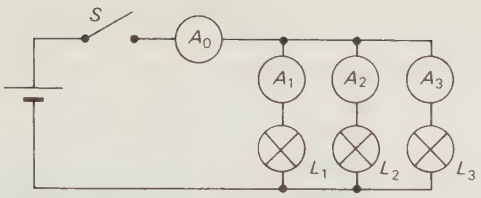


Fig. 25.5

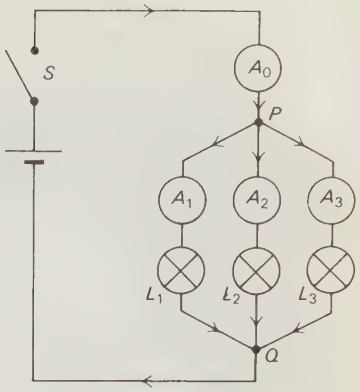


Fig. 25.6

arbitrarily, a direction for the flow of current, and by convention it is agreed the electric current flows out of the positive terminal of a cell and into the negative terminal. This is conventional current. We can put arrow-heads on our circuit diagrams to indicate this direction, and this will be done from now on.

Having considered the behaviour of current in a series circuit, one can proceed to a parallel circuit, that is, one with alternative branches through which current may flow. If we have one 1.5 volt dry cell and three 1.5 volt lamps, we can light those lamps to almost full brightness if we connect them in parallel to the cell. (If they were connected in series to the single cell, they would scarcely glow at all.) Suppose that there are a number of ammeters available, and that their calibrations are in agreement with each other, as could easily be proven by the procedure described above. A circuit is made up as shown in Fig. 25.5. When switch S is closed, the lamps all light to approximately equal brightness, and, as one would expect, the three ammeters A_1, A_2, and A_3 all register approximately the same amount. It is found that ammeter A_0, however, registers a current which equals the sum of the amounts shown by A_1, A_2, and A_3. If lamp L_3 is unscrewed from its socket, the reading of A_3 falls to zero. The readings of A_1 and A_2 change hardly at all, but it is still true that the reading on A_0 equals the sum of readings on A_1, A_2, and A_3. In practice they will always rise slightly. This effect is due to the *internal resistance* of the cell. If the circuit is redrawn as in Fig. 25.6 we see clearly that we have two distinct junction points, P and Q. What has already been discovered about the relationship of the readings on the four ammeters, adds strength to the concept of current as being the flow of something.

The current flowing through A_0 is dividing, at point P, into three "streams," a stream flowing through each of the three lamps. It is as if water were flowing through a main pipe into three smaller pipes joined to it at P. If the flow-rate through each of the three smaller pipes were 0.25 gallons every minute, then the flow-rate in the main pipe would have to be three times this amount, 0.75 gallons per minute.

In the circuit of Fig. 25.6, there are four wires joined at one point. Even if there were more than four, any number, in fact, it could be shown by experiments similar to the one described that the total current flowing into a junction-point in a circuit always equals the total current flowing out of the junction. This experimental fact is known as Kirchhoff's first law.

This flow model for what is happening in the electric circuit has been suggested by the analogy between the behaviour of the ammeters in series and parallel circuits and the behaviour of flow-meters in circuits of pipes around which fluids are being pumped. We are forced to ask what is circulating in this electrical case. All experiments described so far (and this is worth careful reflection) have exhibited the effects of some invisible "stuff" when it is flowing;

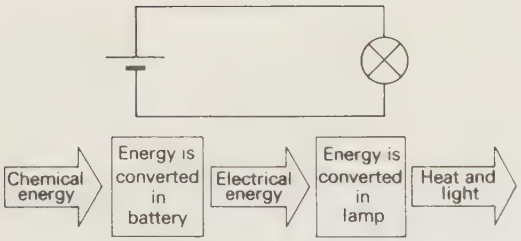


Fig. 25.7

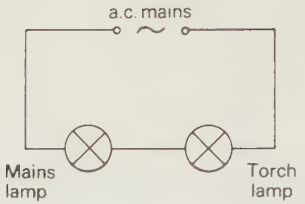


Fig. 25.8

and only when it is flowing. No effects have been observed so far when the circuit is broken; when our analogy suggests that the flow of stuff has ceased. No effects have been observed so far when we remove the battery and then complete the circuit. This suggests that the battery fills the role of a pump in the electrical circuit, re-circulating the stuff which is flowing.

The evidence we have gathered so far can be interpreted then in terms of a simple model of a stuff, which we call electricity, which flows round an electrical circuit in which there is a battery or other electrical source. This is a matter which we shall consider in greater detail in Unit 8.

The current which we measure by means of an ammeter is simply the rate of flow of electricity. Another way of putting this is to say that the ammeter measures the quantity of electricity which flows through the instrument in unit time. We use the symbol Q for quantity of electricity, and so, if a steady current flows through the meter for a time t, in the quantity of electricity which has passed through is given by $Q/t = I$. This can be written

$$Q = It. \qquad (25.1)$$

The unit of time is the second, and of current, the ampere. Thus the unit of Q could be expressed as an ampere second, and it is sometimes referred to thus, but commonly it is called the *coulomb*.

A verbal definition of the coulomb can thus be given: the coulomb is the quantity of electricity passing a given point on an electrical conductor in one second when a steady current of 1 ampere is flowing in the conductor.

So now, in answer to the question: "*what flows* when there is an electric current?" one could answer "*coulombs*." This answer makes mathematical sense, because the coulomb is defined mathematically in terms of current and time: but it has no physical reality as yet. What would one coulomb, on its own, and at rest, look or feel like? What would be its physical manifestation? This is a question we shall be better equipped to answer in Chapter 28.

25.2 POTENTIAL DIFFERENCE

Armed with the concepts of electric current and electric charge, we can proceed to the important discussion of the energy transformations in electrical circuits. As an introduction to this, consider the simple circuit shown in Fig. 25.7. Electric charge circulates at a steady rate around the circuit and the lamp is steadily lit. In the cell chemical energy is being transformed; in the lamp energy is being given out in the forms of heat and light. It is reasonable, therefore, to think of the electric charge, as it circulates around the circuit, as being the agency which carries the energy from the cell to the lamp. Since the lamp is giving out energy at a steady rate and charge is circulating at a steady rate, it makes sense to assume that each coulomb of charge is carrying a certain quantity of energy with it: it collects this energy from the cell and delivers it to the lamp. Immediately this raises the question of *how much* energy is being carried by each coulomb of electricity. Consider what would be the effect of altering the strength of the current in the circuit of Fig. 25.7. If the current is increased the lamp glows more brightly: obviously energy is being delivered to the lamp at a faster rate. Thus, clearly, the rate at which energy is delivered to

the lamp depends upon the strength of the current. But the current is not the only factor which determines the rate at which the energy is delivered. An excellent demonstration of this fact is an experiment in which two lamps of very different power are connected in series to the mains. In Fig. 25.8 L_1 is a 240 V 60 W lamp of the usual kind designed for use with the British a.c. mains. L_2 is an ordinary torch lamp designed to work at a voltage of about three volts and to take a current of about a quarter of an ampere. What do you think will happen when the supply is switched on? Both lamps will light to approximately normal brightness, although the layman would probably be willing to bet on the smaller lamp being burned out almost instantly! There is no possible doubt that energy is being converted at a much faster rate in the large lamp than in the small one: but we know, because the two lamps are connected in series, that the same current must be flowing through each. The same current is being accompanied by very different rates of energy conversion: therefore the strength of the current flowing through a device is not the only factor determining the rate of energy conversion in the device. (You are advised not to try this experiment at home.)

We now have to demonstrate what other factor, or factors, determine the rate at which energy is delivered to an electrical device. This question can be pursued by experimental investigation, using an electrical instrument called a joulemeter (Fig. 25.9). This instrument measures electrical energy and its mechanism is basically the same as that of the domestic meter which one finds in every private house and which, from time to time, a representative of the regional Electricity Board comes to read. The only major differences are that the domestic meter is designed to work at the mains voltage and to measure energy in units called kilowatt hours. One kilowatt hour equals 3 600 000 joules. The question of the units in which electrical energy can be measured will be dealt with in detail later: for the time being it is sufficient to appreciate that, when an electrical supply is connected to a load via a joulemeter, the meter will register, in joules, the amount of electrical energy transferred from the supply to the load.



Fig. 25.9 A joulemeter designed to operate at 12 V a.c. (Courtesy of Ferranti, Ltd.)

It is instructive, before actually using a joulemeter in practice, to make an approximate check of its calibration. An electric immersion heater is connected to the output (or load) terminals of the meter, and a suitable low-voltage a.c. supply is connected to the input terminals. The energy output of the immersion heater during a given time is measured by measuring the temperature rise in a measured mass of water, knowing that the specific heat capacity of water is 4180 J kg^{-1} K^{-1}. The energy output, measured in this way, can then be compared with the reading of the joulemeter.

As has been already mentioned, a laboratory joulemeter is generally a low-voltage device not suitable for use with the mains. However, using a transformer we can step down the mains voltage to a suitable value for use with the joulemeter. An instructive experiment is to take a selection of low-voltage lamps, say, a 12 V, 8 V, and a 2.5 V lamp, and to connect each one in turn into the circuit shown in Fig. 25.10. By adjusting the rheostat R it is possible

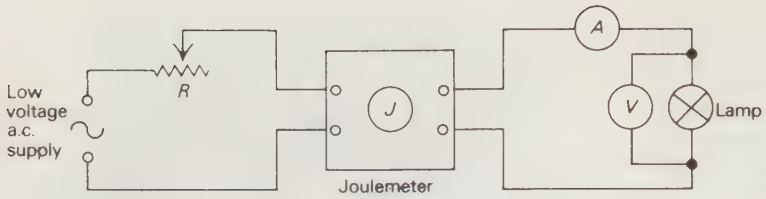


Fig. 25.10

to arrange that the same current flows in each lamp when it is connected. In an experiment of this kind the results displayed in Table 25.1 were obtained, the supply being switched on for one minute in each case.

Table 25.1

Current in each lamp: 0.20 A		Time switched on: 60 s	
Lamp type	Energy utilized (joules)	Charge passed (coulombs)	Joules per coulomb
2.5 V 0.2 A	32	12	2.7
8.0 V 0.2 A	97	12	8.1
12.0 V 0.2 A	154	12	12.8

The current through each lamp was 0.20 A and it flowed for a time of 60 s. Thus we know that the charge which has passed through each lamp is $0.20 \times 60 = 12$ C. If the supply had been switched on for *two* minutes in each case, then twice the amount of energy would have been delivered to each lamp. This suggests that the lamps differ in this important respect: the amount of energy converted per coulomb of charge is different for each lamp. These amounts have been calculated and entered in the right hand column of Table 25.1. It should not be surprising, now, that these amounts correspond roughly to the *voltage* markings on the lamps. We are now in a position to elucidate the precise meaning of the term *voltage*. The word itself means, literally, "the thing which is measured in volts," The "thing," or rather, "the physical quantity" which is measured in volts, is the *number of joules per coulomb of charge*. Thus, in the smallest of the three lamps used in the experiment described above, for every coulomb of charge which has passed through the lamp filament, approximately 2.5 joules of energy have been transformed into heat and light: and that lamp was marked 2.5 V (meaning 2.5 volts). The actual measurement made with the joulemeter gave 2.7 joules per coulomb. This is explained (assuming the joulemeter has been accurately calibrated) by the fact that the filaments of mass-produced lamps vary one from another in their characteristics, and it should not be surprising that a bulb marked 2.5 V 0.2 A may use a little more or a little less than precisely 2.5 joule per coulomb, even when the current through the filament is maintained at exactly 0.2 A. In the experiment we could have measured this quantity by a much more direct method, using a voltmeter (which has probably occurred to you already!) and it is useful to repeat the experiment of Fig. 25.9, this time connecting a voltmeter in parallel with each of the three lamps being tested. The voltages indicated by the voltmeter should then be found to agree well with the values calculated in the right-hand column of Table 25.1.

The quantity which is measured in volts is properly termed *potential difference* (p.d.). A voltmeter is always connected *in parallel with*, or, as it is often expressed, *across* an electrical device, in order to measure the p.d. between the terminals of the device. One often refers to "the p.d. across" an electrical device when signifying the p.d. thus measured.

The exact definitions of the term p.d. and the unit in which it is measured are as follows:

> The p.d. across an electrical device is the amount of energy converted per unit charge passing through the device.
> The volt is the unit of p.d. such that one volt = one joule per coulomb.

It should now be possible for you to work out how, in the circuit of Fig. 25.10, one could find the amount of energy converted in a lamp in a given time

without using the joulemeter. These facts give the clues:

a) The voltmeter measures the number of joules of energy converted in the lamp per coulomb of charge passed through the lamp. For this quantity we use the symbol V.
b) The ammeter measures the number of coulombs passing through the lamp per second. For this we use the symbol I.
c) The clock measures the time, t, during which the supply is switched on.

Using the symbol W for the amount of energy, we can write the solution thus:

(number of joules) = (number of joules per coulomb) × (number of coulombs per second) × (number of seconds),

or, in symbols:

$$W = VIt. \qquad (25.2)$$

If you do not have access to a joulemeter, the experiment can be performed directly by immersing the lamps in turn in water (taking care not to short-circuit the supply). To make sure that all the energy radiated away from the filament is transferred to the water, this should be blackened by adding black ink or colloidal graphite.

Problem 25.1 An electric room heater takes a current of 6 A from the 250 V mains supply. Calculate (a) how much energy it consumes when switched on for a period of ten hours, and (b) the cost of having the heater switched on for ten hours when the cost per *unit* is 1.5p. (1 unit = one kilowatt hour = 3 600 000 J.)

Solution To compute the answer by the method outlined above we must first express the time in seconds.

a) Ten hours $= 10 \times 60 \times 60 = 36\,000$ s
energy $= W = VIt$
$= 250 \times 6 \times 36\,000$ J
$= 54\,000\,000$ J

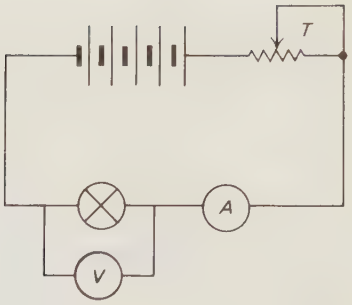


Fig. 25.11

b) Cost $= 54\,000\,000/3\,600\,000 \times 1.5$p.
$= 15 \times 1.5$p.
$= 22.5$p.

It is important for every householder, indeed for every consumer of electrical energy, to appreciate that what he pays for is not the volts or the amps separately, but the quantity of electrical energy utilized, i.e., the product VIt. This can always be calculated by the method of Problem 25.1 if the loads are resistive in nature.

25.3 RESISTANCE

One might wonder, referring back to the experiment of Fig. 25.9, how it comes about that the same current flows through each of the three tungsten filament lamps although the p.d. across each lamp is different. The difference can only be due to different characteristics of the tungsten filaments. The property of the filament which determines the relation between the p.d. across it and the current flowing through it is its *resistance*. We can begin investigating this property by means of a simple experiment with a tungsten filament lamp. In the circuit of Fig. 25.11 the current flowing through a lamp can be varied by the rheostat T. There are suitable meters in the circuit to measure the current and the p.d. A typical set of measurements (Table 25.2) shows more clearly the relationship between the p.d. and the current. In trying to analyse these results, however, one must bear in mind that in this situation we have not just *two* varying physical

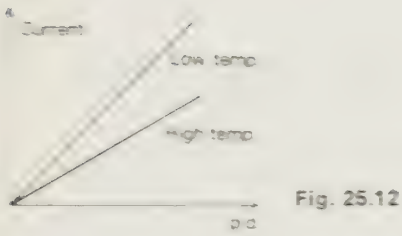


Fig. 25.12

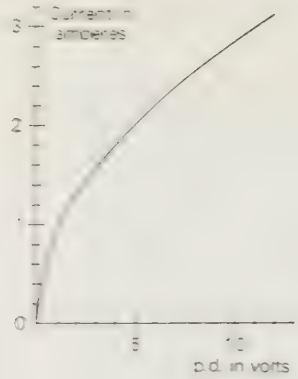


Fig. 25.13

properties (p.d. and current) but *three*. The third property is the temperature of the filament, which varies during the course of the experiment from around room temperature to a temperature of around 2000°C. Although the temperature of the filament has not been measured in the experiment, it is reasonable to suppose that the very marked variation in temperature has contributed in some way to the pattern of results obtained. To carry out a meaningful investigation of the relationship between p.d. and current in a conductor (e.g., a metal wire) we must begin by controlling any variable factors other than p.d. and current. It is not practicable to keep the temperature of a lamp filament constant while performing this kind of experiment with it, but a similar experiment can be done with a length of wire of a suitable type immersed in oil to keep its temperature constant. In such an experiment it is found that, regardless of the material of which the wire is made, the current is directly proportional to the p.d. over the whole range of measurement. If the experiment is then repeated with the wire maintained at a steady, higher temperature (by heating the wire in a sand-bath, for example) then again it is found that the current is proportional to the p.d., but for any given value of p.d., the current is less than what it was in the cool wire. The measurements of p.d. and current, when plotted on a graph, then appear as in Fig. 25.12. This kind of change in the current-p.d. characteristic as a result of change in temperature is a property of *all* wires made of pure metals within a normal working range of temperature. The reader should now be able to appreciate why the current-p.d. graph for the lamp filament, in Fig. 25.13, is not a straight line but a curve of decreasing gradient, remembering that the greater the current, the higher the filament temperature.

Table 25.2

P.d. in volts	Current in amperes
0.4	0.60
0.9	0.90
2.0	1.25
3.9	1.70
6.5	2.30
10.0	2.85
12.0	3.10

We have referred previously to the analogy between the flow of a fluid in pipes and the flow of electricity in a circuit. We observed that a pump is necessary to maintain flow in the first case and a battery or other electrical generator in the second. When that pump is operating, a pressure difference can be measured across the ends of each section of pipe in the circuit—and experiments show that the pressure difference across a particular section depends on the current flow and also on the nature of the pipe.

A pressure difference applied to the ends of a narrow-bore pipe gives rise to a smaller flow rate than the same pressure difference applied to a wide-bore pipe of the same length. The narrow pipe has a greater "resistance" to the flow. A convenient way of expressing this resistance would be to evaluate the quotient

$$\frac{\text{pressure difference}}{\text{flow rate}}.$$

A large pressure difference accompanied by a small flow rate would signify a high resistance, for example.

We may apply the same idea to an electrical circuit and define the *electrical resistance* of a conductor as the quotient

$$\frac{\text{potential difference}}{\text{current}}. \qquad (25.3)$$

In symbols

$$R = V/I.$$

The unit of electrical resistance is then the volt ampere^{-1} or *ohm*. The international symbol for the ohm is Ω.

In the experiment illustrated in Fig. 25.11 values of p.d. and current were measured for a tungsten filament lamp. These values are tabulated again in Table 25.3 and the resistance of the filament has been calculated for each pair of values. The resistance is observed to increase as the current through the filament increases. But look at the results of a similar experiment made with a metal wire kept at approximately constant temperature, in Table 25.4. It is clear in this case that resistance of the wire is constant, regardless of the strength of the current flowing in the wire. This implies that, if one knows the resistance of a wire, one can easily predict what the current in the wire will be for any given value of p.d. applied to it, or what the p.d. will be for any given current, provided that there is no change in temperature. A length of metal wire at constant temperature is one example of a linear electrical device: a device in which the current is always directly proportional to the applied p.d. The particular practical uses of linear and non-linear electrical components will be discussed in detail later in this chapter.

Table 25.3 Tungsten filament lamp.

P.d. in volts	Current in amperes	Resistance in ohms
0.4	0.60	0.67
0.9	0.90	1.00
2.0	1.25	1.60
3.9	1.70	2.29
6.5	2.30	2.82
10.0	2.85	3.51
12.0	3.10	3.87

Table 25.4 Piece of resistance wire.

P.d. in volts	Current in amperes	Resistance in ohms
1.2	0.50	2.40
2.9	1.20	2.42
4.1	1.65	2.48
5.2	2.10	2.48
6.3	2.50	2.52
7.6	3.05	2.49

The relationship of direct proportionality between the current and the p.d. in metal wires held at constant temperature was first convincingly demonstrated by Georg Simon Ohm, a German mathematician and physicist. The results of his investigations appeared in a paper published in 1826. The proportionality apparently held good for a wire of any dimensions made of any metal, over the whole range of currents which he was able to obtain, provided that the temperature of the wire was not allowed to vary.

Ohm's law can be expressed thus: a steady current flowing through a metallic conductor is directly proportional to the p.d. between the ends of the conductor, provided that the temperature and other physical conditions are kept constant. The law is found to be obeyed very closely by all metals, and by certain non-metallic conductors and electrical devices. These can therefore conveniently be referred to as *ohmic conductors* and *ohmic devices*: alternatively (as mentioned earlier) they can be called *linear* because of the linearity of their characteristic current-p.d. relationship.

Why do certain classes of conductor obey this simple law while others do not? Does the law apply

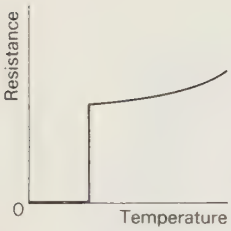


Fig. 25.14

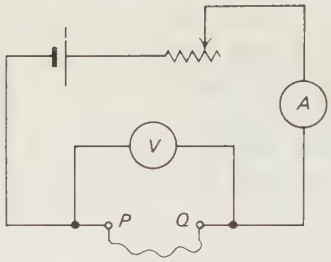


Fig. 25.15 To investigate the dependence of the resistance of a wire on its length. Various lengths of the wire are connected between the terminals *P* and *Q*.

to ohmic conductors when the current is extremely weak, or extremely strong (one might expect some kind of saturation effect with very strong currents, perhaps)? The nature of the process of electrical conduction in general will be discussed later in Unit 8. For the time being we shall be concerned with how practical use can be made of the particular conducting properties of different materials and devices.

However we must digress to note that at very low temperatures most metals depart radically from the simple behaviour described by Ohm's law. Indeed their resistance falls to zero. In this state they are superconductors. Fig. 25.14 shows graphically how the resistance of a superconductor changes at this transition.

The phenomenon, which was discovered by Kamerlingh Onnes in 1911, has the interesting implication that once a current is started in a superconducting ring, it may continue indefinitely. But, of far more significance is the application of the phenomenon to the design of super-conducting magnets. Such magnets can be made to be far more powerful and yet cheaper to run than conventional powerful electromagnets. This suggests important consequences for electric power generation and transmission and for electric motors if the upper limit of superconductivity can be raised. At present this stands at 21 K. A further rise to, say, 30 K could revolutionize electrical technology. (Matthias, B. T., 1971.)

25.4 THE RESISTANCE OF WIRES

G. S. Ohm was the first person to investigate how the resistance of a wire in a circuit depended upon the *length* of the wire. To investigate the relationship between resistance and length for a wire nowadays one could use the circuit of Fig. 25.15. It is found that the resistance is directly proportional to the length, provided that the wire has uniform cross-section. To investigate how the resistance depends upon the thickness of the wire one can use samples of wire made of one particular material, all samples having the same length but different thickness. It is found that the resistance depends upon the area of cross-section of the wire in such a way that if a wire has twice the thickness of a second wire, having the same length, then the resistance of the first wire is one quarter of the resistance of the second. The resistance is inversely proportional to the area of cross-section. These facts are summed up in a formula for the resistance, R, of a wire in terms of its length, L, and its area of cross-section, A:

$$R = \rho \frac{L}{A}. \qquad (25.4)$$

The symbol ρ represents the *resistivity* of the material of the wire, and it is a constant for a given material at a specified temperature. It should be noted that ρ is not simply a numerical constant, it is a physical quantity having physical units. If R is measured in ohms, L in metres, and A in square metres, then the unit of ρ must be: ohm metre (abbreviated: Ω m). The resistivities of some common materials are given in Table 25.5. Note that the temperature is specified here: the resistivity of any material is dependent upon temperature. A discussion of this temperature-dependence will be taken up shortly.

Table 25.5 Resistivities of some common materials, at 0° C.

Material	Resistivity ($\times 10^{-8}$) in Ω m
Silver	1.51
Copper	1.56
Aluminium	2.45
Gold	2.04
Sodium	4.2
Lead	19.0
Mercury	94.1
Nichrome	130
Constantan	48
Manganin	42

The formula $R = \rho(L/A)$ can be used to calculate the resistance of any wire or cable of uniform cross-section.

In some applications it is more convenient to use the *conductivity* σ of the material. This is the reciprocal of the resistivity and, as has been pointed out in Unit 3, varies very widely when considered over the whole range of insulators and conductors.

Problem 25.2 The heating element of an electric toaster is made of nichrome tape, thickness 0.05 mm, width 1 mm. The length of the tape is 4 m. Calculate the resistance of the element at 0° C.

Solution Cross-section area of tape

$$= 10^{-3} \times 0.05 \times 10^{-3}\ \text{m}^2,$$

$$= 5 \times 10^{-8}\ \text{m}^2.$$

Resistivity of nichrome $= 130 \times 10^{-8}\ \Omega\,\text{m}$ (see Table 25.5).

Using the formula:

$$R = 130 \times 10^{-8} \times \frac{4}{5 \times 10^{-8}}$$

$$R = \frac{520}{5}$$

$$\therefore R = 104\,\Omega.$$

If one wanted to find the resistance of a cable consisting of several strands of wire of one particular material, then one would simply substitute the total cross-section area of all the strands for A in the formula. If the strands are made of different materials, as for instance in the steel and aluminium conductors used in the overhead power cables of the National Grid in Great Britain, then one cannot use the simple formula alone: another formula, which gives the resistance of a number of resistors connected together in parallel, must be used as well.

Problem 25.3 The resistance of overhead cables of the National Grid. Estimate the resistance per kilometre of one type of aluminium-and-steel conductor used on the 132 kV grid system, given the following data and guide-lines.

The cable consists of a core of 7 steel strands, each 0.110 inch (0.28 cm) diameter: the total cross-sectional area of these steel conductors is therefore about 0.4 cm^2. Around this core are 30 aluminium strands of the same diameter: the total cross-sectional area of aluminium is thus about 1.8 cm^2.

Take the resistivity of aluminium as 3×10^{-8} Ω m, and that of steel as 12×10^{-8} Ω m.

a) Using the appropriate formula, calculate the approximate resistance of 1 km of the aluminium conductors.
b) Repeat for the steel conductors.
c) Compare the resistance of the two types of conductor: you will notice that one is considerably greater than the other. Therefore most of the current will flow in the conductor of lower resistance, and in our rough estimate we can neglect the presence of the high-resistance conductor: which conductor is this?
d) Finally, state the resistance per kilometre to one significant figure.

The results of these calculations will be used in Problem 26.3 where the power loss in the conductors is considered.

Solution From Eq. (25.4)

$$R = \rho \frac{L}{A}.$$

a) For 1 km of aluminium:

$$R = 3 \times 10^{-8} \times \frac{10^3}{1.8 \times 10^{-4}}\ \Omega$$

$$= 0.17\ \Omega.$$

b) For 1 km of steel:

$$R = 12 \times 10^{-8} \times \frac{10^3}{0.4 \times 10^{-4}}\ \Omega$$

$$= 3\ \Omega.$$

c) The resistance of the steel is thus about twenty times that of the aluminium, so most of the current flows through the aluminium.

d) The overall resistance of the cable will be only slightly less than 0.17 Ω (because of the presence of the steel), and so call it 0.2 Ω per km, correct to 1 significant figure.

Problem 25.4 In the article on microelectronics in the *Electronic Engineers' Reference Book* it is stated that, in a typical thin-film circuit, in which resistors are made by the condensation of vaporized nichrome alloy on to borosilicate glass, a 10 kΩ resistor would be 0.4 inches long and 0.01 inches wide (in metric equivalents, approximately: 1 cm long and 0.25 mm wide). The resistor is a thin, flat strip of the metal. From this information, and using the value for the resistivity of nichrome given in Table 25.5, estimate to one significant figure the thickness of the strip of metal, and express the answer in nanometres (1 nm $= 10^{-9}$ m). Then, given that the diameter of single atoms in a metal like nichrome is very roughly 0.2 nm, express the thickness of the resistor strip in terms of atomic diameters.

Solution Let the thickness of the resistor strip be x.

Using the formula $R = \rho(L/A)$ and substituting, we get

$$10^4 = 130 \times 10^{-8} \times \frac{10^{-2}}{0.25 \times 10^{-3} x}$$

$$= \frac{130 \times 4 \times 10^{-7}}{x},$$

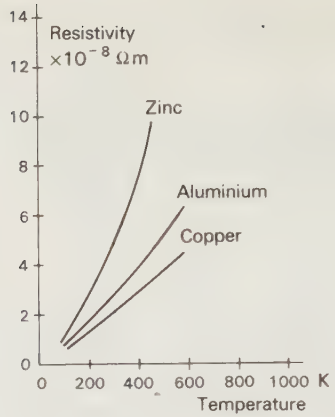


Fig. 25.16 Variation of resistivity of some pure metals with temperature.

hence

$$x = \frac{520 \times 10^{-7}}{10^4}\ \text{m}$$

$$= 5.2 \times 10^{-9}\ \text{m} \doteq 5.2\ \text{nm}.$$

If one atomic diameter is approximately 0.2 nm, then the strip is about 5.2/0.2 = about 25 atoms thick!

Layers as thin as this are achieved by condensing the vapour of the metal, in a near-vacuum, on to the surface of the substrate material (typically, borosilicate glass). It is possible to control the minute thickness very accurately since the material is being deposited, literally, atom by atom.

The resistivity of any pure metal increases with its temperature. Figure 25.16 shows the effect of temperature on the resistivities of some metals. Over a limited range of temperature the graphs are approximately linear, so that a *temperature coefficient of resistivity* can be specified for the metal concerned. This coefficient is defined as the fractional increase in resistivity per degree rise in temperature.

Certain alloys have a high resistivity and a very small temperature coefficient of resistivity in comparison with pure metals. Examples of such alloys are

Constantan and Manganin which are the metals most often used for making resistors of the wire-wound variety: and in this application it is very convenient that variation of resistance as a result of change in temperature is generally negligibly small. Table 25.6 shows the temperature coefficients of resistivity of some metals, and here it should be noted that the coefficients all have the same order of magnitude for pure metals. The coefficients for resistance alloys are in the order of one hundredth of that for a pure metal. So far as is known, it is impossible to find a "recipe" for an alloy which shows no change in resistivity whatsoever when its temperature changes.

Table 25.6

Material	Temperature coefficient of resistivity in degrees C^{-1}
Pure metals	
Aluminium	45×10^{-4}
Copper	66×10^{-4}
Silver	41×10^{-4}
Zinc	42×10^{-4}
Alloys	
Constantan	$\pm 0.2 \times 10^{-4}$
Nichrome	1×10^{-4}

The elements of electric heaters are generally made of an alloy called nichrome which (like constantan and manganin) has a high resistivity but a small temperature coefficient. Its main virtue is that it does not deteriorate mechanically or chemically when maintained at a high temperature (e.g., red heat) for prolonged periods of time. Often a heating element is made of a flat strip of the metal rather than of round wire (see Problem 25.3 above), as in an electric toaster. The flat strip form of conductor combines a high resistance with a large surface area for rapid dissipation of heat.

Chapter 26
POWER, ENERGY, AND ELECTROMOTIVE FORCE

26.1 Power
26.2 Electromotive force
26.3 Energy changes
26.4 Matching source and load
26.5 Kirchhoff's second law
26.6 Solving circuit problems

26.1 POWER

When you switch on the headlamps of a motor car, radiant energy streams out forwards and lights the way ahead. The energy flow process was discussed in Section 25.2 in which the term *potential difference* was defined. The rate at which energy is supplied by the battery is the power supplied; likewise the rate at which energy is utilized by the lamps is the power utilized. The word *power* means *rate of conversion of energy*.

The unit of power is the *watt* (W). If we say that a lamp has a power of one watt, we mean that electrical energy is being converted into heat and light at the rate of one joule per second. (See Unit 2.) We can write:

$$1\ \mathrm{W} = 1\ \mathrm{J\,s^{-1}}.$$

Let us see now how the power consumed by a passive electrical device such as a lamp is related to the current flowing through it and the potential difference across it. The current is the *quantity of electric charge flowing through the filament per unit time*, and the p.d. is the *amount of energy delivered per unit charge*: thus, if we multiply together the current and the p.d. we get the *amount of energy delivered per unit time*, which is the *power* delivered to the lamp. Expressed in terms of the units for these quantities, the product is written:

$$\mathrm{JC^{-1}} \times \mathrm{C\,s^{-1}},$$

which is simplified to $\mathrm{J\,s^{-1}}$ which, by definition, equals W. The conventional symbol for power is P, and we write the formula

$$P = VI. \qquad (26.1)$$

We can express the power in terms of the resistance, R, of the device if we use the relationship $R = V/I$, which is the definition of resistance. Substituting $V = IR$ into Eq. (26.1) we get $P = I^2R$; alternatively, substituting $I = V/R$ we get $P = V^2/R$.

Table 26.1 shows the power ratings of some electrical devices in common use. The three quantities will, of course, be observed to satisfy the relationship $P = VI$ in each case.

Multiples and sub-multiples of the basic unit, the watt, are in common use (see Appendix 2), the

Table 26.1 Power ratings of some domestic electrical appliances (in watts)

Tape recorder	50
Reading lamp	100
Television receiver	150
Iron *	625–750
Fast-boiling kettle	2750
Room heater *	2760–3250
Water heater (100 gallons capacity)	6000

* Maximum and minimum power ratings are generally marked on such appliances as these which are fitted with means of controlling the energy supplied.

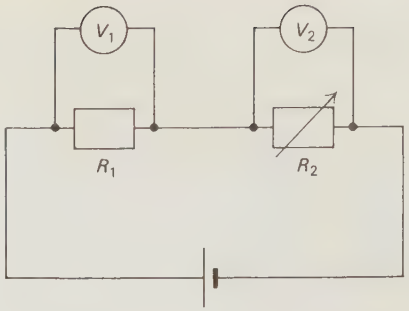


Fig. 26.1

former in the literature of electrical power generation and use, the latter in the world of electronics and telecommunication.

The power of electrical devices, both domestic and industrial, is nearly always marked on them. On a domestic lamp one may find, for example, the correct operating voltage and the power indicated thus: 240 V 60 W.

The power output of a generating station is likely to be specified in Megawatts. A typical large station may have a capacity of 2000 MW.

26.2 ELECTROMOTIVE FORCE

The role of the battery in an electric circuit has been compared with the role of the pump in a fluid circuit. It must, however, be much more than this, for not only is it necessary for the maintenance of the flow of electric charge but it is the major source of the energy which is transformed to heat, light, etc., in the circuit components. (Other sources arising from the existence of temperature gradients and contacts between different metals also contribute.)

Consider a simple circuit (Fig. 26.1) made up of a fixed resistor R_1 of about 100 ohms, a variable resistor R_2, two voltmeters of high sensitivity and high resistance, and a new $1\frac{1}{2}$ V dry battery. As R_2 is varied, readings of the voltmeters are taken. Some typical results with $R_1 = 100$ ohms are given in Table 26.2. It seems that the battery supplied about the same voltage (although the figures for V_1 and V_2 show a small rise) throughout the experiment.

Repetition of the experiment with the resistance R_1 reduced to 10 ohms yielded the results shown in Table 26.3. You should plot graphs of V_2 against R_2 for the two experiments. What do these graphs lead you to expect if now the experiment is repeated with R_1 reduced to zero?

Table 26.2 $R_1 = 100$ ohms

R_2 ohms	V_2 volts	V_1 volts	$V_1 + V_2$ volts
50	0.50	0.99	1.49
100	0.76	0.75	1.51
200	1.04	0.50	1.54
500	1.28	0.26	1.54
1000	1.41	0.14	1.55
5000	1.51	0.04	1.56

Table 26.3 $R_1 = 10$ ohms

R_2	V_2	V_1	$V_1 + V_2$
10	0.73	0.72	1.45
50	1.27	0.24	1.51
200	1.46	0.07	1.53
500	1.50	0.03	1.53
1000	1.52	0.02	1.54
5000	1.54	0.01	1.54

Table 26.4 $R_1 = 0$ ohms

R_2	V_2	V_1	$V_1 + V_2$
2	1.20	0	1.20
5	1.36	0	1.36
10	1.43	0	1.43
50	1.50	0	1.50
200	1.52	0	1.52
500	1.53	0	1.53
1000	1.54	0	1.54
5000	1.55	0	1.55

The figures for this case are given in Table 26.4.

The similarity in form between the three graphs suggests that there is, after all, some resistance other than R_2 in the circuit on this last occasion. We know that the copper connecting wires themselves can contribute very little indeed to the total resistance of the circuit so we conclude that the battery has some internal resistance itself.

Turning now to the reading of voltmeter V_2 we see that this does tend, in each case, to a maximum value as resistance R_2 is increased to a large value by comparison with other circuit resistances.

Problem 26.1

a) Why is the graph of V_2 against R_2 so much steeper when R_1 is small than when R_1 is large?

b) How can you decide, from the results of the final experiment, whether the value for the internal resistance is more or less than 5 ohms?

c) What would you expect the voltmeter V_2 to read if the resistance R_2 was made infinite?

In this last case the utilization of energy is zero (or nearly so if the resistance of the voltmeter itself is very, very high—typically 40 000 ohms).

The voltmeter readings in these experiments indicate how the utilization of energy is shared between the various components. There seems to be no way, however, in which the voltmeter can be used to measure how energy is utilized or dissipated in the battery itself. We would need to devise an experiment in which the heat produced in the cell could be measured to discover this and, in principle, this could be done. We would then be able to state, exactly, the maximum total energy per unit charge which the complete circuit (battery included) can utilize or dissipate. This maximum voltage we may call the *electromotive force* (*e.m.f.*) of the battery. We could, in principle, determine it by immersing the complete circuit in a calorimeter and measuring the current and the rate of production of heat.

Now, such a battery can provide energy to a simple circuit for a long period of time without change. The energetic conditions are steady. The amount of energy per unit charge generated by the source is equal to the total amount of energy per unit charge which is utilized in the whole circuit. So we see that the electromotive force of the battery is equal, also, to the energy per unit charge which is generated within it.

Let us now further develop these energetic aspects of the electric circuit in which a steady direct current is flowing. Consider a battery (or, indeed, any suitable electrical source) with e.m.f. E and internal resistance r connected to an external resistance R. Let the current be I.

During an interval of time t, It units of charge pass and the energy transferred in the external circuit is I^2Rt whilst the energy dissipated in the battery is I^2rt.

From the definition of electromotive force as the energy per unit charge utilized in the entire current

$$E = \frac{\text{total utilized energy}}{\text{charge passing}}$$

$$= \frac{I^2Rt + I^2rt}{It} \qquad (26.2)$$

$$= I(R + r)$$

$$\therefore I = \frac{E}{R + r}. \qquad (26.3)$$

Multiplying both sides by R

$$IR = E\frac{R}{R + r}.$$

This is the voltage V which we can measure with a voltmeter. So

$$V = E\frac{R}{R + r}$$

$$= E\frac{1}{1 + \frac{r}{R}}.$$

As R tends to an infinite resistance, so the p.d. V tends towards E. This is the justification for the direct measurement of e.m.f. by the use of a high resistance voltmeter or a potentiometer.

Equation (26.2) may be written

$$EIt = I^2Rt + I^2rt,$$

or

$$EI = I^2R + I^2r,$$

which may be written as

$$EI - I^2R - I^2r = 0.$$

In this form we are using a sign convention: we are allotting a positive sign to terms representing electrical power supplied to the circuit, and a negative sign to terms representing electrical power utilized in the circuit and in the battery. The left-hand side of the equation rewritten in this form then represents the net gain of electrical power in the circuit and battery, which is zero. The importance of using this sign convention becomes apparent in Section 26.6, where one has to be careful to distinguish regions of a circuit in which power is supplied from regions where power is utilized.

26.3 ENERGY CHANGES

The dry cell used in our experiment will eventually cease to provide a useful e.m.f. and, at this point, is discarded. It is not a rechargeable cell. The constituents of this cell (zinc case, ammonium chloride paste, a mixture of powdered carbon and manganese dioxide, carbon rod) are involved in a chemical reaction which makes it possible to transfer some energy to the circuit. We would hardly expect all the available energy to be transferred usefully: some causes the temperature of the cell to rise when in use. The e.m.f. (which is 1.5 V) describes the maximum transfer of energy to useful forms. When supplying a current to an external circuit the ratio of the energy transferred usefully to the total energy available falls and this is revealed as a fall in the external p.d.

Certain cells are reversible and chargeable. The most useful of these include the lead accumulator and the nickel iron (Nife) cell. In such cells, when in use, the e.m.f. is in the same direction as the current and energy is being converted to what we may call an electrical form. But, when on charge, the e.m.f. is in opposition to the current, and electrical energy is being transformed to another form which we may think of as chemical. In this case, too, such other energy sources as the exchange of energy with the surroundings resulting from temperature gradients are involved.

Problem 26.2 A 3 V torch battery has an internal resistance of 0.8 ohm. It is connected to a lamp which draws a current of 0.25 A.

a) Calculate the p.d. across the lamp.

The battery is then used to supply a number of lamps connected in parallel, and they take a total current of 1.25 A from the battery.

b) Calculate the p.d. across the lamps in this case.

Solution

a) From Eq. (26.3)

$$\text{p.d. across load} = IR = E - Ir,$$
$$= 3 - 0.25 \times 0.8,$$
$$= 3 - 0.2,$$
$$= \underline{2.8\ \text{V}}.$$

b) Similarly,

$$\text{p.d. across load} = 3 - 1.25 \times 0.8,$$
$$= 3 - 1.0,$$
$$= \underline{2.0\ \text{V}}.$$

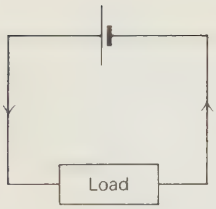


Fig. 26.2

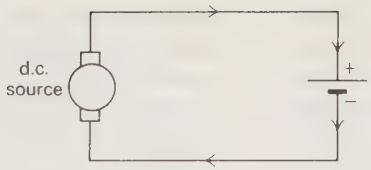


Fig. 26.3

In case (a) only 0.2 V are utilized within the battery; in case (b) 1.0 V. The greater the current taken from the battery, the greater is the utilization of energy within the battery. In practice this is not likely to be a serious problem because one would not normally take such a large current as in case (b) from such a small battery because the useful lifetime of the battery would be inconveniently short in this situation: for, in general, the smaller the chemical cell the larger is its internal resistance. It is partly a consequence of the fact that batteries have internal resistance that torch bulbs rated at 2.5 V are used with dry batteries whose e.m.f., when new, is 3.0 V. The small battery of a pocket "pen" torch may have an internal resistance of as much as 5 Ω, and if such a battery is supplying a current of 0.1 A, the "lost volts" will be 0.5 V, which in this example would be just right. However, internal resistance is not the only consideration in this practical situation: there is also the fact that, during use, the e.m.f. of a battery decreases.

We now have an important theoretical point to clarify. Looking at the circuit of Fig. 26.2 we see that we have two principal elements in the circuit, a device which converts energy from some other form into electrical form (the generator or cell), and a device which converts the energy from electrical into some other form (the load). In each case the rate of energy conversion—the power—can be calculated from the p.d. across the device and the current flowing through it. But can these measurements of p.d. and current reveal whether power is being supplied by the device or utilized by it? They can, provided that one takes into account the polarity of the p.d. across the device and the direction of the current. In the circuit of Fig. 26.2 positive electric charge flows in the direction of the arrows. This charge, when it moves through the generator, goes from a region of lower potential to a region of higher potential and thus gains energy. In the generator, then, the direction of current flow and the direction of increasing potential are the same. It is convenient to think of the potential rise across the generator, going from its negative to its positive terminal. Thus, the rule is that if the current and potential rise are in the same direction for an electrical device, then the device is supplying electrical energy to the circuit. Consider now the load. It is a passive device, and in it the electric charge loses potential energy as it flows through. The current flow is in the opposite direction to the potential rise, and here electrical power is utilized. These ideas are important when considering the circuit of Fig. 26.3. Here we have a storage battery (accumulator, or re-chargeable battery) being charged from a source of direct current. Here, the direction of current flow in relation to the polarity of the battery looks wrong: but the explanation is that, in the battery here, we are converting electrical energy back into chemical energy. Notice that the potential rise across the battery is in the opposite direction to the current: hence (by our rule) the battery is utilizing electrical energy. The internal resistance of the battery, r, is also utilizing electrical energy but, of course, not converting it into chemical form! We can write the equation for conservation of energy in the circuit thus:

$$VI - EI - I^2r = 0,$$

allotting positive signs to the terms representing electrical power *supplied* and negative signs to power *utilized*, using the sign convention explained above and the symbols of Section 26.2.

Problem 26.3 Power loss in the overhead conductors of the National Grid.

The solution to Problem 25.3 gives an estimate of the resistance per kilometre of one type of aluminium-and-steel conductor used on the 132 kV transmission system of the National Grid: to 1 significant figure the answer was $0.2\ \Omega\ km^{-1}$. A calculation accurate to 2 significant figures would have given the result: $0.16\ \Omega\ km^{-1}$ (which is about 0.25 Ω per mile).

Suppose that the loading is such that the effective value of the current (that is, the equivalent value of steady direct current) flowing in one of these conductors is 400 A.

a) Calculate the voltage drop along a 20 km run of cable.
b) Roughly what percentage of the working voltage of the system is this voltage-drop?
c) Assuming that the voltage at the supply end is actually 132 kV, and the effective value of the current is still 400 A, what is the power input to the system, to the nearest megawatt?
d) From the answer to (a), and knowing the current, calculate the power dissipated as heat in one conductor, to the nearest tenth of a megawatt.
e) What percentage of the power input is lost as heat in one conductor?

You will see from these calculations, if you have done them correctly, that the power loss is small in comparison with the input power. It is small because a very high voltage (132 kV) is being used. Suppose that the voltage were one tenth of this value: we shall now consider the consequences of this.

f) If the voltage is to be 13.2 kV, what must the current be, in order to have the same power input to the system? (Remember, power = voltage × current.)
g) What is the voltage drop along a 20 km run of cable now?
h) Compare the answers to (f) and (h): they are almost the same. This means we have lost nearly all the volts along the line! This reasoning need be pursued no further—you should now see the full implications of this argument.

Solution

a) The resistance of a 20 km run of cable is $20 \times 0.16 = 3.2\ \Omega$. The voltage drop is given by $V = IR$, whence $V = 400 \times 3.2 = 1\,280$ V.
b) The voltage drop is $1280/132\,000 \times 100$ which is roughly 1% of the working voltage.
c) The power input is $132\,000 \times 400 = 52\,800\,000$ W = 53 MW to the nearest megawatt.
d) In one conductor the power dissipated = 1280×400 W = 0.5 MW to the nearest tenth of a megawatt.
e) The percentage power loss is $0.5/53 \times 100$, that is, roughly 1% per conductor.
f) If the voltage is reduced to one tenth, then the current must be increased ten times (that is, to 4000) if the power input is to be the same, because power = the product of voltage and current.
g) The voltage drop, given by $V = IR$, will be ten times what it was before because the current I has been increased ten times (R, we assume, remains unchanged): thus the voltage drop is now 12 800 V.
h) The new supply voltage is 13.2 kV, that is, 13 200 V: so the voltage drop is only slightly less than this. Presumable therefore the power loss in the cable is also only very slightly less than the power input, and the system would be useless in this form. We see from this that a change in the supply voltage by a factor of *ten*, subject to the constraint of keeping the power input unchanged, means that the power loss is altered by a factor of about *a hundred* (from approximately 1% to 100%).

26.4 MATCHING SOURCE AND LOAD

The presence of internal resistance in an electrical generator sets a limit on the amount of electrical power which can be delivered by the generator into a load. We shall now see how this limitation arises, for it has important consequences not only in electrical power engineering but in electronics.

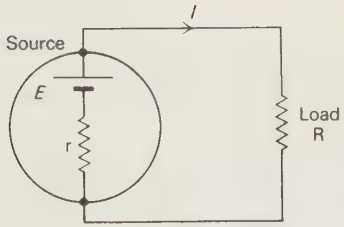


Fig. 26.4

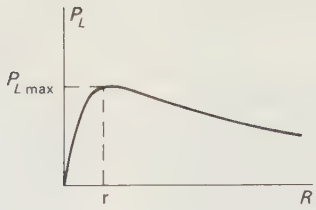


Fig. 26.5 Variation of power delivered to load with the resistance of the load. The maximum power is delivered when $R = r$.

When an electric generator is connected to a load, some of the generated power is dissipated in the internal resistance, and hence wasted, and the rest is delivered to the load. The amount of wasted power depends on the current which is being drawn from the generator, and this current depends on the resistance of the load. Alteration of the load resistance will thus cause a change in the amount of power wasted by dissipation in the internal resistance of the generator. In many practical situations it is both possible and desirable to adjust the value of load resistance so that the greatest possible power is transferred from the source to the load. We shall show here what conditions are required for maximum transfer of power from source to load.

In the circuit of Fig. 26.4 the current is given by

$$I = \frac{E}{R + r}$$

and the power (P_L) dissipated in the load, by

$$P_L = I^2R = \left(\frac{E}{R + r}\right)^2 R. \qquad (26.4)$$

Assume now that the e.m.f. and internal resistance of the source cannot be altered, but that the load resistance is adjustable. What value of R makes the power dissipated in the load a maximum? Figure 26.5 shows how the power, P_L, given by expression (26.4), varies with R. It is seen that the power is greatest when the load resistance is equal to the internal resistance of the source. Under these conditions the power wasted (by dissipation in the internal resistance of the source) is equal to the power delivered to the load, and each is equal to $E^2/4r$. The value of R which makes P_L a maximum can be found by differentiating the expression for P_L with respect to R and equating to zero, as follows

$$P_L = \left(\frac{E}{R + r}\right)^2 R$$

$$\frac{dP_L}{dR} = E^2 \frac{d}{dR}\left[\frac{R}{(R + r)^2}\right]$$

$$= E^2 \left[\frac{(R + r)^2 - 2R(R + r)}{(R + r)^4}\right].$$

Equating this to zero gives

$$r^2 - R^2 = 0.$$

Hence

$$R = \pm r. \qquad (26.5)$$

Since only the result $R = + r$ has physical meaning in this context, we conclude that the condition for P_L to be a maximum is that $R = r$.

This case is analogous to one we have already examined in Unit 2—that of the transfer of energy when two masses collide. In that case, the maximum transfer occurred when the masses were equal to one another.

When a source of electrical energy is connected to a load in such a way that maximum power is transferred from source to load, the source and load are said to be *matched*. The problem of matching a source to a load is particularly important in certain branches of electronics, for instance in the design of audio amplifiers. The theory of matching, as applied to audio amplifiers, will not be elaborated upon here because an understanding of alternating current theory is essential: but the general idea can be simply



Fig. 26.6

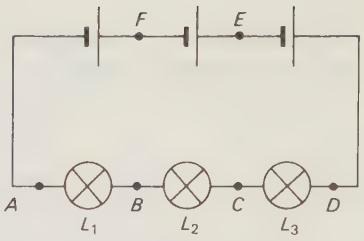


Fig. 26.7

described as follows. Figure 26.6 represents an audio amplifier being used to amplify the signal from a microphone and to feed it into a loudspeaker. For this system to work efficiently the microphone must be matched to the input of the amplifier, and the loudspeaker matched to the output of the amplifier. Any microphone has a characteristic *impedance*, and it is sufficient to say here that impedance is, in a.c. devices, equivalent to resistance in d.c. devices. The impedance of a good quality moving-coil microphone is small, in the order of a few ohms. The audio amplifier has a specified input impedance and output impedance. For maximum power transfer from the microphone to the amplifier input, the microphone impedance should be equal to the input impedance of the amplifier: likewise for maximum transfer of power from the amplifier output to the loudspeaker, the amplifier output impedance must be equal to the loudspeaker's impedance. A certain amount of *mismatch* is acceptable in practice: one would not hesitate to connect an 8 Ω loudspeaker to an amplifier whose output impedance was 7.5 Ω, for instance. An idea of just how great a mismatch is tolerable can be obtained by arguing as follows. We have shown that, for a situation like that of Fig. 26.4, maximum power is transferred from source to load when $R = r$, and this maximum power is $E^2/4r$. If $R = 2r$, then substitution in expression (26.4) shows that $P_L = 2E^2/9r$, which is only slightly smaller than $E^2/4R$. If, on the other hand, $R = r/2$, then $P_L = 2E^2/9r$ once again. Thus, from the point of view of power transfer, connecting a 15 Ω loudspeaker to an amplifier whose output impedance is 7.5 Ω or even 30 Ω would not produce a bad mismatch. Also, as you can verify by doing a little algebra with expression (26.4), the values of R which make P_L equal to half its maximum value of $E^2/4r$ are $R = (3 \pm 2\sqrt{2})\, r$, that is, $R = 5.82r$, or $R = 0.172r$. Thus, an approximately six-to-one mismatch would reduce the power transfer to about half its optimum value. However, power transfer is not the only consideration when matching items of audio equipment, and in general, mismatches as bad as these just discussed are not advisable, for other reasons.

26.5 KIRCHHOFF'S SECOND LAW

It is necessary now to find out what happens when there are two or more electrical devices in a circuit. Look at the circuit of Fig. 26.7. This circuit has been used earlier in elucidating the behaviour of electric current: now we shall use it to throw more light on the concept of p.d. To measure the p.d. across the lamp L_1 we would connect a voltmeter to points A and B: we can say alternatively that we are measuring the p.d. between points A and B. If we then measure the p.d. between points A and C we find that it equals the sum of the p.d. between A and B and the p.d. between B and C. This result can be expressed in symbols thus:

$$V_{AC} = V_{AB} + V_{BC}.$$

Similarly it is found by experiment that

$$V_{BD} = V_{BC} + V_{CD},$$

and, as one would expect (and as must follow logically from the previous two results),

$$V_{AD} = V_{AB} + V_{BC} + V_{CD}.$$

Table 26.5

Negative terminal of meter connected to point	Positive terminal of meter connected to point	Voltmeter reading
A	B	+1.2
B	C	+1.5
C	D	+1.4
D	E	−1.3
E	F	−1.3
F	A	−1.5

Table 26.6

Point	Potential (in volts)
A	0
B	+1.2
C	+2.7
D	+4.1
E	+2.8
F	+1.5
A	0

Another way of expressing these results would be to say that the voltages add up as one goes around a circuit. The same rule is found to be obeyed by the cells when the voltmeter is connected across each cell separately, then across two of the cells, then across all three. One could then express the result thus: the cell voltages when added up equal the lamp voltages added up. But in fact the situation is a little more subtle than this. To clarify the subtlety we need to make voltage measurements on the circuit of Fig. 26.7, using a centre-zero voltmeter. With such a voltmeter, if a cell whose e.m.f. is 1.5 V is connected to the meter with the correct polarity the needle swings to the right of zero, registering + 1.5 V. If the cell is connected with reverse polarity (i.e., with the connections to the meter interchanged) then the needle swings to the left of zero, registering − 1.5 V. Suppose now that we go around the circuit of Fig. 26.7 with our centre-zero voltmeter, starting with the meter's negative terminal connected to point *A* and its positive terminal to *B*. The results of a typical set of measurements are shown in Table 26.5.

As expected, the negative voltages in the table of results (which represent the p.d.'s across the three cells), when added together, equal the sum of the positive voltages (the p.d.'s across the three bulbs). We can say, in a sense, that "the voltage goes up" as we go from *A* to *B* around the circuit, and then "goes down" from *D*, via *E* and *F*, back to *A*. But our crude language, indicated by inverted commas, is not really precise enough, because the ideas of "voltage going up" and "voltage going down" are not exact. The word "voltage" so far has been an alternative word for "potential difference". The word "difference" is important to appreciate here: there is a certain potential at point *A*, and another potential at point *B*, in the circuit of Fig. 26.7, and the difference in potential between *A* and *B* is 1.2 V. What is more, the potential at *B* is 1.2 V higher than the potential at *A*. Alternatively one can say that the potential at *A* is 1.2 V lower than the potential at *B* (not surprisingly, perhaps). Thus, in going from *A* to *D*, we find that the potential increases by a certain amount; and in going from *D* via the cells back to *A* we find that the potential decreases by the same amount. We can, if we choose, call the potential at point *A* zero. If we do this, then we can tabulate the potentials at different points in the circuit as in Table 26.6. From one point of view the results may seem obvious: one does not expect to arrive back at point *A*, having gone around the circuit, to find a potential there different from what it was when one started the journey at point *A*!

The results of this experiment are an illustration of *Kirchhoff's second law*, which can be expressed as follows: "The sum of potential rises (taken as positive) and potential drops (taken as negative) along any closed path is zero." The rises occur in active components; the drops in passive ones, if one goes round the circuit with the current.

We have considered only a simple series circuit, constituting a single closed path: but Kirchhoff's second law can be applied to any circuit, however complex. Figure 26.8 shows a bridge circuit (this has various applications, as we shall see). There are

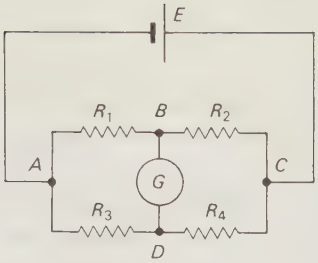


Fig. 26.8 A bridge circuit. There are seven closed paths in this circuit. Can you trace them all?

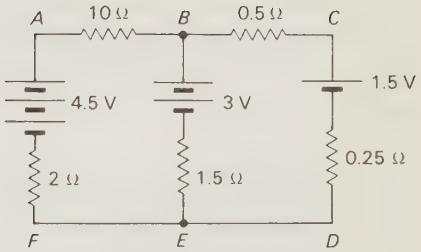


Fig. 26.9

several closed paths which can be followed round in this circuit:

a) $A - R_1 - B - G - D - R_3 - A$,
b) $B - R_2 - C - R_4 - D - G - B$,
c) $A - R_1 - B - R_2 - C - R_4 - D - R_3 - A$,
d) $A - E - C - R_2 - B - R_1 - A$,
e) $A - E - C - R_4 - D - R_3 - A$.

Each of these paths has been followed around clockwise, but it could just as well have been followed around anticlockwise. See whether you can trace two more, perhaps less obvious, closed paths in this circuit. If we were to go around any one of these closed paths with a centre-zero voltmeter in the manner already described, we would find that Kirchhoff's second law is obeyed.

Kirchhoff's second law often enables us to write equations relating the potential differences across the several devices in a complex electrical circuit, and together with Kirchhoff's first law (the current law) it provides a powerful means of analyzing complex circuits.

Before we proceed to seeing how simple circuit problems can be solved, we must consider one further fact about circuits in general. In the simple series circuit of Fig. 26.7 the current is flowing clockwise around the circuit. In the lamp L_1 the current flow is therefore from right to left, from B to A, and we note also that the potential at B is higher than at A. Similarly for the other two lamps: in "going across" a lamp in the same direction as the current flow, we always experience a drop in potential. And this is a general fact; that the current flow through any passive electrical device (that is to say, any device in which electrical energy is converted into some other form) is from the high potential side of the device to the low potential side. But in going across a cell (and neglecting for the moment any internal resistance it may have) from its negative terminal to its positive terminal, there is a potential rise: another way of expressing this fact is to say that the positive terminal of the cell is always at a higher potential than the negative terminal. Such a potential rise, associated as it is with an active driving source, is the e.m.f. of the source.

26.6 SOLVING CIRCUIT PROBLEMS

We shall now apply the laws and the ideas discussed above to the solving of a not-too-complex circuit problem.

The problem we shall set ourselves is to determine the current in each part of the circuit of Fig. 26.9. The small resistors just below the cells represent the internal resistance of the cells. In the diagram we first indicate the assumed directions of the currents: we start by drawing an arrowhead above the 4.5 V cell and labelling it I_1; which is one of the currents to be calculated. Then, above the 3 V cell, we indicate the current I_2. Now we apply Kirchhoff's first law to the junction point B: the current flowing away from B, towards C, must be equal to the sum of the currents flowing into B. Thus the current flowing from B to C

must be $(I_1 + I_2)$. Now we apply Kirchhoff's second law, firstly to the closed path *A-B-E-F*, which will give us one equation containing the unknown quantities I_1 and I_2; and secondly to the path *B-C-D-E*, which will give us a second equation containing I_1 and I_2. Thus we shall obtain a pair of simultaneous equations which can then be solved to yield the values of I_1 and I_2.

Consider first the path *A-B-E-F*. We shall write potential rises as positive quantities and potential drops as negative. Going from *A* to *B* we are travelling in the same direction as the current, so we have a potential drop which can be written $-10I_1$. From *B* to *E* we have firstly a potential drop as we go across the cell from its positive to its negative terminal: thus we write this potential drop as -3 V. In going down across the internal resistance of the middle cell we are going in a direction *opposite* to that of the current, and so we have a potential rise: this will be written $+1.5I_2$. From *E* to *F* we have a direct connection with (we assume) no resistance, so there is no change in potential. From *F* towards *A*, across the internal resistance of the left-hand cell, we are going in the same direction as the current, so here is a potential drop, written $-2I_1$. Finally, in returning to *A* across the cell, we have a potential rise of $+4.5$ V. Now, Kirchhoff's second law says that the sum of potential rises and potential drops along any closed path is zero: so we can write, for the path *A-B-E-F*:

$$-10I_1 - 3 + 1.5I_2 - 2I_1 + 4.5 = 0.$$

And for the path *B-C-D-E* (you should work through this equation term by term):

$$-0.5(I_1 + I_2) - 1.5 - 0.25(I_1 + I_2) - 1.5I_2 + 3 = 0.$$

These two equations can be simplified and rewritten as follows:

$$-12I_1 + 1.5I_2 + 1.5 = 0,$$

$$-0.75I_1 - 2.25I_2 + 1.5 = 0.$$

If we multiply the upper equation by 1.5, then the equations can be added together and the terms in I_2 will disappear:

$$-18I_1 + 2.25I_2 + 2.25 = 0,$$

$$-0.75I_1 - 2.25I_2 + 1.5 = 0.$$

Adding the equations together, we obtain

$$-18.75I_1 + 3.75 = 0$$

therefore

$$I_1 = \frac{3.75}{18.75} A$$

$$= 0.2 A.$$

Substituting this value for I_1 in one of the equations above containing I_1 and I_2 enables us to find I_2, and the result is

$$I_2 = 0.6 \text{ A}.$$

The current $(I_1 + I_2)$ flowing from *C* to *D*, via the 1.5 V cell, is thus 0.8 A. You will note that we have here a rather curious situation: the 1.5 V cell, if it were operating normally in a circuit, would cause current to flow in a direction opposite to what we have here. In this particular circuit it is, as it were, being overpowered by the effect of the other two cells. This is a situation which can happen in practice. What really happens is that whereas the 4.5 V and 3 V cells are being discharged (or more precisely, in them chemical energy is being converted into electrical energy) in the 1.5 V cell the reverse process is happening, with electrical energy being converted back into chemical energy. This is called *charging*. If the 1.5 V cell is a rechargeable cell, e.g., a lead-acid accumulator or a Nife cell, then indeed it *will* be charged by this process: but an ordinary dry (Leclanché-type) cell is not ordinarily rechargeable.

Having computed the currents in the circuit we can now determine the potential difference between any two points we choose. For example, the p.d. between *B* and *C* is simply the product of the current in the resistor and the resistance: 0.8×0.5 V $= 0.4$ V. The potential at *B* is higher than at *C* because the current is flowing from *B* to *C*. Suppose we wish to find the p.d. between *B* and *E*, and to determine

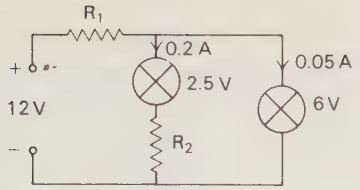


Fig. 26.10

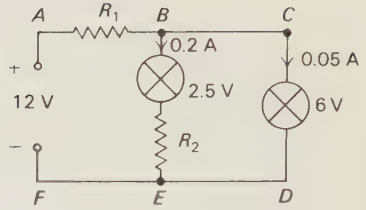


Fig. 26.11

which point is at the higher potential. In going from B downwards across the internal resistance of the cell, against the direction of the current, we find a rise in potential of an amount 0.6×1.5 V $= 0.9$ V. So, *down* by 3 V and then *up* by 0.9 V means a net drop of 2.1 V. Thus the point B is at a potential of 2.1 V higher than point E.

Now you should check your understanding of the ideas, by trying Problem 26.4, and carrying out similar calculations for other pairs of points in the circuit of Fig. 26.9. It is a good idea, when solving the problem, to copy the circuit diagram, choose one point in the circuit and make it the zero of potential (say point F), and then to label the other points in their circuit with their potentials when they are calculated.

Problem 26.4

a) Find the p.d. between points A and F. (See Fig. 26.9.) Which of the two points is at the higher potential?

b) Find the p.d. between C and D. Which of these two points is at the higher potential?

c) Calling the potential of F, E, and D zero, label the points A, B, and C with their appropriate potentials. Hence calculate the current in AB and in BC: these results should check with the values of current already calculated.

Problem 26.5

a) Suppose you have a 12 V d.c. supply, and from this you wish to light two lamps, one rated at 6 V 0.05 A and the other at 2.5 V 0.2 A. Their voltage and current ratings are different. Because their current ratings are different they cannot be connected simply, in series. The circuit shown in Fig. 26.10 could be used. Using Kirchhoff's two laws, find the values of the two resistors needed, R_1 and R_2.

b) If the 6 V lamp blows, then it becomes an open circuit and takes no current. Try to predict, by calculation, what happens to the 2.5 V lamp. You may assume that the resistance of the 2.5 V lamp does not alter: then, afterwards, you may need to take into account that in fact the resistance will increase if the lamp glows more brightly.

c) Consider what would happen if the 2.5 V lamp were to blow: would the 6 V lamp's behaviour alter? Again, first assume that the resistance of the 6 V lamp does not alter.

Solution

a) Let us label some of the points in the circuit, as shown in Fig. 26.11. Now we know that the p.d. between C and D must be 6 V: hence also the p.d. between B and E. Thus the p.d. across R_1 must be $12 - 6 = 6$ V. The current flowing from A to B divides at junction B, hence this current is $0.2 + 0.05 = 0.25$ A. Thus we can find R_1, since

$$R_1 = \frac{6}{0.25} = 24\ \Omega.$$

Now to find R_2. The p.d. between B and E is 6 V. But we must have 2.5 V across the lamp, so the p.d. across R_2 must be $6 - 2.5 = 3.5$ V. The current flowing through it is of course 0.2 A. Hence

$$R_2 = \frac{3.5}{0.2} = 17.5\ \Omega.$$

b) If the 6 V lamp blows, then the circuit reduces to the one shown in Fig. 26.12. Assuming that the resistance of the 2.5 V lamp (which is 2.5/0.2 =

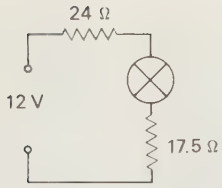


Fig. 26.12

12.5 Ω) does not alter, then the total circuit resistance is now 24 + 12.5 + 17.5 = 54 Ω, hence the current is 12/54 = 0.22 A. This implies a slight increase in the brightness of the lamp: 10% over-running, but since its resistance will increase as a result of the higher temperature, the current will not be quite so large as the value calculated. The precise value is indeterminate.

c) If the 2.5 V lamp blows, then we have simply the 6 V lamp, whose resistance in normal operation is 6/0.05 = 120 Ω, in series with R_1 across the supply. The current we thus calculate as 12/(24 + 120) = 0.083 A. Therefore, we might well expect this lamp to blow also.

Chapter 27
SOME USEFUL CIRCUITS

27.1 The potential divider
27.2 The slide wire potentiometer
27.3 Bridge circuits
27.4 Combining resistances

27.1 THE POTENTIAL DIVIDER

The purpose of a voltage dropper is to split the supply voltage into two parts: one part is the voltage across some electrical device—the device's correct operating voltage—and the other part is the voltage drop across a resistor. It is often desirable to divide a supply voltage into two or more (not necessarily equal) parts by means of two or more resistors connected in series across the supply. The simplest example of this potential divider is shown in Fig. 27.1. Here we have two resistors connected in series to a battery. We know that the sum of the p.d.'s across R_1 and R_2 must equal the p.d. across the battery terminals, V.

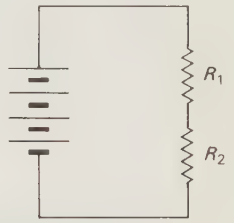


Fig. 27.1

Suppose the current in the circuit is I. Then

$$V = IR_1 + IR_2,$$

hence

$$I = \frac{V}{R_1 + R_2}$$

and the p.d. across R_1 is

$$IR_1 = \frac{R_1}{R_1 + R_2} V \qquad (27.1)$$

similarly, the p.d. across R_2 is

$$IR_2 = \frac{R_2}{R_1 + R_2} V.$$

Thus the p.d. of the battery has been divided in the ratio $R_1 : R_2$ across the two resistors. By choosing suitable values of R_1 and R_2 we can obtain any voltage we wish, between zero and V, across one of the resistors.

If a continuously variable voltage source is required, the pair of fixed resistors can be replaced by

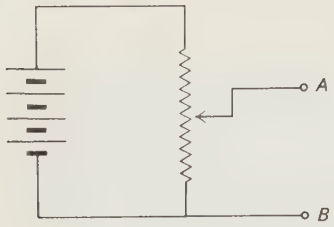


Fig. 27.2 The potentiometer. The arrowhead represents the "wiper" or sliding contact. The p.d. between A and B is proportional to the distance between the wiper and the lower end of the resistance track in a linear potentiometer.

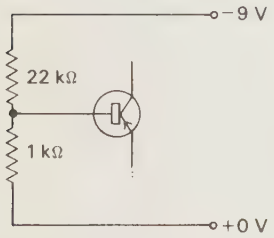


Fig. 27.3 A two-resistor potential divider used to supply the base of a transistor with a small negative voltage.

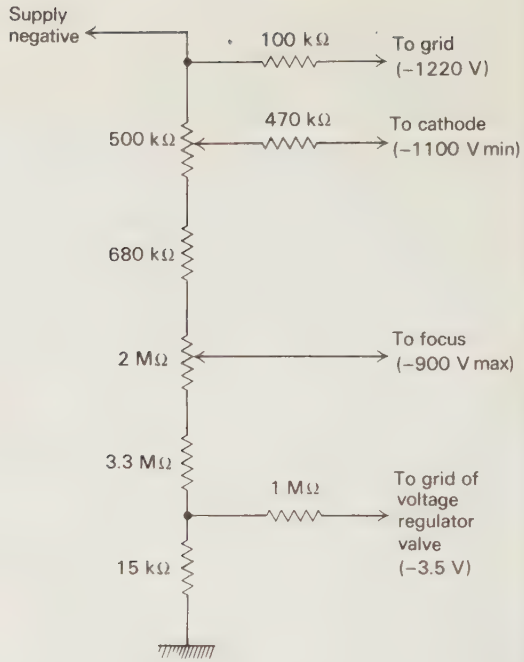


Fig. 27.4 A potential divider used to supply the electrodes of a cathode ray tube in a laboratory oscilloscope.

a *potentiometer*: this is basically a resistor to which a sliding contact (a *wiper*) has been added. If the potentiometer is made from a length of resistance wire (often coiled around a former), then we have a linear potentiometer, so called because the resistance between the wiper and one end of the resistance wire is directly proportional to the distance between the wiper and that end. Hence, in Fig. 27.2, the p.d. between the points A and B is directly proportional to the distance of the wiper from the lower end of the resistance track. You may wonder what determines the value of resistance which a potentiometer should have. This depends on the particular application for which the potentiometer is being used. If the resistance is small, a large current is drawn from the battery, and this may be undesirable (one should bear in mind that the current flowing in the resistor heats it, and this energy, which must be dissipated, is wasted). If, on the other hand, the resistance of the potentiometer is large, then although only a small current is taken from the battery, it is impossible to take much current from the terminals A and B. However, in many applications of the potential divider, the current taken from the terminals A and B is much smaller than the current flowing through the resistance track: under these conditions no difficulty arises in using it to provide any desired voltage within the range from zero to the full battery voltage.

One important use of a two-resistor potential divider is to supply the correct bias voltage to the base of a transistor. Figure 27.3 shows how this can be done. Typical values for the resistors are shown, suitable for a low-power transistor being used for small-signal amplification. The base of the transistor is here maintained at a negative potential with respect to the positive rail whose potential is zero. The p.d. between the base and the positive rail is

$$\frac{1}{22 + 1} \times 9 = 0.39 \text{ V}.$$

This voltage has been calculated assuming that no current flows between the base of the transistor and the junction of the two resistors. In practice a small current will flow, from the base into the junction of the two resistors, and the effect of this is to reduce the p.d. across the lower resistor of the pair. But again, in practice, this current is generally much smaller than the current flowing in the two resistors already, and so the calculation above is only slightly in error.

A potential divider sometimes consists of a chain of several resistors. Figure 27.4 shows an arrangement used to maintain some of the electrodes of a cathode ray tube at the required operating potentials with respect to the cathode. In this instance the currents taken by the electrodes are very small, and so large values of resistance can be used, drawing very little current from the high voltage supply.

The potential divider principle can be used to convert changes in resistance into changes in voltage. For example, a certain type of photosensitive cell, known as a light-dependent resistor (which we shall abbreviate to LDR), has a resistance which depends upon how strongly it is illuminated. Variations in the intensity of the light falling upon the LDR cause variations in its resistance. If the LDR is connected in series with a fixed resistor across a battery (or other voltage source) then variations in illumination give rise to changes in p.d. across the LDR. These changes in p.d. can be used to trigger an electronic switching circuit and thus to switch on (or off) some piece of apparatus. Problem 27.1 concerns an LDR and gives you an appreciation of the magnitudes of the variations involved.

Problem 27.1 A typical cadmium sulphide, light-dependent resistor, has a resistance which is about 100 Ω when it is strongly illuminated and about 1 M Ω when in almost total darkness. By using the LDR as one member of a two-resistor potential divider, as shown in Fig. 27.5, it is possible to convert the changes in resistance into changes in voltage for the purpose of operating a voltage-controlled switching circuit. By choosing a suitable value of R one can arrange that the voltage across the LDR is nearly zero when it is illuminated, and very nearly the full supply voltage, 6 V, when unilluminated. Suggest a suitable value for R and explain why it is a suitable value.

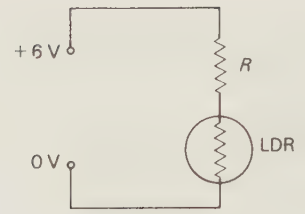


Fig. 27.5

Solution For the voltage across the LDR to be nearly zero when the LDR is illuminated we require that its resistance—call it x_{bright}—should be much smaller than R. We also require that x_{dark} should be much greater than R. In terms of ratios, we require:

$$\frac{x_{\text{bright}}}{R} \ll 1 \qquad \text{and} \qquad \frac{x_{\text{dark}}}{R} \gg 1,$$

or,

$$\frac{R}{x_{\text{dark}}} \ll 1.$$

Suppose we simply equate the two ratios which are to be much less than unity, we get

$$\frac{x_{\text{bright}}}{R} = \frac{R}{x_{\text{dark}}}$$

hence

$$R^2 = x_{\text{bright}} \times x_{\text{dark}} = 100 \times 10^6$$

hence

$$R = 10^4\ \Omega, \qquad \text{or} \quad 10\ \text{k}\Omega.$$

The two ratios thus become 1/100, so that when the LDR is illuminated the voltage across it is only about 1% of the supply voltage, and 99% of the supply voltage when it is dark.

27.2 THE SLIDE WIRE POTENTIOMETER

For certain applications, a potential divider is made very easily, using a length of uniform resistance wire AB, connected directly across a battery, as shown in

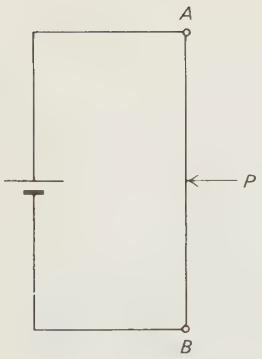


Fig. 27.6 A slide-wire potentiometer.

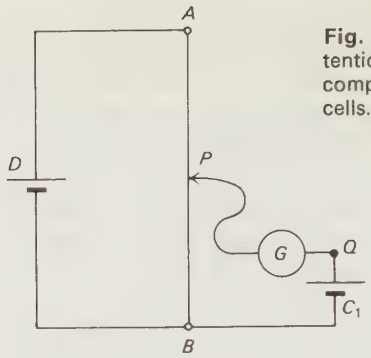


Fig. 27.7 A slide-wire potentiometer in use for the comparison of e.m.f.s of cells.

Fig. 27.6. In a simple laboratory version of this, the wire AB is 1 m in length and is mounted directly above a metre scale divided into centimetres and millimetres. A moveable contact, P, can be set at any position on the wire. In this way, a potential can be obtained between P and B, having any desired value in the range from zero to the full p.d. provided by the battery. The usefulness of this type of potentiometer lies in the fact that the p.d. between P and B is directly proportional to the length of the section of wire between P and B so long as no current is taken from P. This is a consequence of the wire's uniformity of thickness which means that the resistance of any section of the wire is directly proportional to that section's length.

The slide-wire potentiometer is used chiefly for comparing voltages, that is to say, finding the ratio of two voltages. It cannot be used for making absolute measurements of voltage. Let us see how the instrument can be used for comparing the e.m.f's of two chemical cells. If one of the cells is a standard cell, for example, a Weston Cadmium Cell which has an e.m.f. of 1.015 9 V at 20° C, then the e.m.f. of the other cell can be calculated accurately when the ratio of e.m.f.'s has been found by means of the potentiometer. Figure 27.7 shows the arrangement for these measurements. To understand its operation, we shall consider typical numerical values in an experimental situation. D is the battery which is used simply to maintain a steady current flowing in the line AB, and thus a constant p.d. between A and B. For D one would use a lead-acid accumulator, giving a p.d. of about 2 V, or a pair of Nife cells giving a p.d. of about 2.5 V. The exact value of this p.d. is unimportant: what is essential is that it should remain constant throughout the experiment. We find that when the moveable contact, P, is touched to the wire near to B, the needle of the centre-zero galvanometer G is deflected strongly one way: let us say this deflection is to the left of zero. When P is close to A, a strong deflection to the right is obtained. As one would expect, there is one point on the wire at which, when P is touched on to that point, the galvanometer shows no deflection. We can interpret these results as follows:

Galvanometer deflection	Interpretation
To the left	Potential at P less than potential at Q, so current flows $Q \rightarrow P$.
To the right	Potential at P greater than potential at Q, so current flows $P \rightarrow Q$.
Zero	Potential at P the same as at Q, so no current flows between P and Q.

In the third of the situations above, P is at what is often called the balance point. In this condition, since there is no p.d. between P and Q, the p.d. across the section of wire PB must be equal to the p.d. across the cell C_1. Also, in this condition of balance, no current is being drawn from the cell C_1, and the p.d. across it is its open-circuit p.d., that is, its e.m.f.

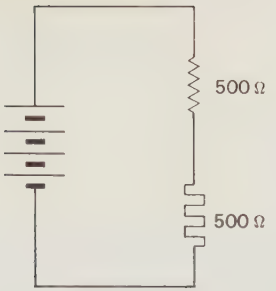


Fig. 27.8 An electrical strain-gauge as one member of a potential divider.

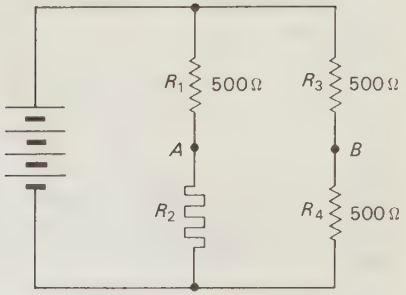


Fig. 27.9 A strain-gauge as one of the members of a four-resistor bridge circuit.

Let us suppose that the length PB which was obtained with cell C_1 was 60.3 cm, and that C_1 was a small dry cell of the type used in a pocket-size electric torch. We are intending to find its e.m.f. accurately. We now replace the dry cell C_1 with a standard cell C_2, and find the balance-point by the same method as before. The result of doing this, together with the previous result, is tabulated. Because the p.d. across any section of the potentiometer wire is directly proportional to the length of the section and no current is drawn from P, the ratio of the e.m.f.'s of the two cells must be equal to ratio of corresponding lengths, that is,

$$\frac{E_1}{E_2} = \frac{L_1}{L_2}. \qquad (27.1)$$

Using the numerical results of Table 27.1, we find

$$\frac{E_1}{1.015\,9} = \frac{60.3}{43.1},$$

whence $E_1 = 1.48$ V. Thus we have found that the e.m.f. of our dry cell is 1.48 V.

With a potentiometer wire one metre long, measurements of length can easily be made to the nearest millimetere, so the length-measurements are reliable to 3 significant figures. Since the e.m.f. of a standard cell is reliable to at least 4 significant figures, our calculated e.m.f. is reliable to 3 significant figures. This degree of accuracy is greater than can be obtained with even a very expensive moving-coil voltmeter.

Table 27.1

Length PB for balance (cm)	e.m.f. of cell (V)
$L_1 = 60.3$	$E_1 = ?$
$L_2 = 43.1$	$E_2 = 1.015\,9$

27.3 BRIDGE CIRCUITS

Suppose now that one wishes to use an electrical strain-gauge in a potential divider circuit in a similar manner. A typical strain-gauge has a resistance of about 500 ohms (see Fig. 27.8), and when subjected to strain the resistance can change by as much as $\frac{1}{2}\%$, i.e., by about 2.5 Ω in this case. If it is in series with a 500 ohm resistor, the p.d. across it changes by a fraction $((502.5/1002.5) - (500/1000))$, or approximately 2.5/1000, of the battery voltage, i.e., about $\frac{1}{4}\%$ only!

One may well wonder if such a small percentage change can be measured accurately at all. We shall tackle this problem now, and see how it can be solved. Consider the circuit of Fig. 27.9. Here there are two potential dividers connected in parallel. One of the total of four resistors is a strain-gauge. Suppose that all four resistive elements have a resistance of precisely 500 Ω. The p.d. across each of the four resistors is half the battery voltage, i.e., it is 3 V. The p.d. between points A and B is zero, because points A and B are at the same potential. If now the strain-gauge, R_2, is subjected to strain, as it was earlier, so that its resistance increases by, say, 2.5 Ω, then the p.d. across R_2 becomes

$$\frac{502.5}{(500 + 502.5)} \times 6 = 3.015 \text{ V},$$

and now there is a p.d. of 0.015 V between A and B. If a sensitive enough voltmeter is connected between A and B, then subjecting the strain-gauge to a little strain can make the pointer of the instrument move from zero to full-scale deflection. Compare this situation with that of Fig. 27.8. A voltmeter connected directly across the strain-gauge there would need to have a full-scale deflection of, say, 5 V. The change of voltage from 3.000 to 3.015 V would obviously produce a negligible movement of the pointer. Thus we see that our twin potential divider arrangement provides a highly sensitive means of making very small fractional changes of resistance observable and measurable. But this arrangement has its practical snags. One must not forget that small fractional changes in resistance, of the magnitude we are concerned to observe, are readily produced by changes of temperature. It is a sobering fact that the touch of a fingertip, on any one of the four resistive elements, will warm it up and alter its resistance sufficiently to produce a change in p.d. between A and B as great as, if not greater than, the change due to strain applied to R_2. It is physically impossible to prevent the resistances from altering when the temperature changes, and certainly difficult in practice to provide thermal insulation of the resistors adequate to prevent any changes in temperature. But what can be done, and what eliminates the unwanted effect of temperature fluctuation, is to ensure that all four elements have the same temperature coefficient of resistance, and then to fix them on a common mounting so that they are always at the same temperature as each other. In this way, if the resistance of one changes by a certain fraction, then the resistance of all the others changes by the same fraction, and these changes produce no alteration in the p.d. between A and B. If a strain-gauge is being used for R_2, as in Fig. 27.9, then the simplest way of fulfilling these requirements is to use four identical strain-gauges for the four elements, with just one of them (R_2) being stuck to the sample which is being subjected to strain.

It is worth pointing out here that, in the arrangement described above, it is not essential that the four resistors (Fig. 27.9) should be equal for the p.d. between A and B to be zero. Let us find what relationship must be satisfied by R_1, R_2, R_3, and R_4, in order that the p.d. between A and B should be zero. We require:

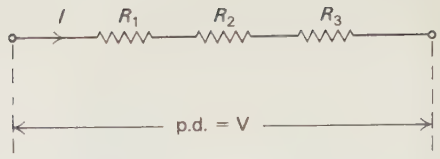


Fig. 27.10

$$\text{p.d. across } R_1 = \text{p.d. across } R_3$$

which implies that

$$\frac{\text{p.d. across } R_1}{\text{p.d. across } R_2} = \frac{\text{p.d. across } R_3}{\text{p.d. across } R_4}.$$

If I_A is the current flowing in R_1 and R_2, and I_B the current in R_3 and R_4, we have

$$\frac{I_A R_1}{I_A R_2} = \frac{I_B R_3}{I_B R_4} \quad \text{i.e.,} \quad \frac{R_1}{R_2} = \frac{R_3}{R_4}.$$

This means that the four resistances can all have different values, e.g., $R_1 = 3$, $R_2 = 6$, $R_3 = 8$, $R_4 = 16$, so long as the relationship

$$\frac{R_1}{R_2} = \frac{R_3}{R_4} \tag{27.2}$$

is satisfied.

Alternatively, this may be written

$$\frac{R_1}{R_3} = \frac{R_2}{R_4}.$$

The arrangement of two twin-resistance potential dividers in parallel is often referred to as a *bridge* circuit, and when the null condition is satisfied, so that there is no p.d. between A and B, we have what is called a *balanced bridge* circuit.

27.4 COMBINING RESISTANCES

When several resistors are connected together in series (see Fig. 27.10) the resistance of the combination

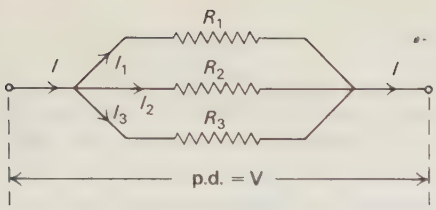


Fig. 27.11

is equal simply to the sum of the individual resistances. This can easily be proved as follows. Consider three resistances, R_1, R_2, R_3, connected together in series, and connected to some source of e.m.f. so that a current, I, flows through them. We already know that the current in each resistor must be the same. If the p.d. between the ends of the combination is V, then (from energy considerations)

$$V = V_1 + V_2 + V_3, \tag{27.3}$$

where V_1, V_2, V_3, are the p.d.'s across the individual resistances R_1, R_2, and R_3. But, by the definition of resistance

$$R_1 = \frac{V_1}{I}, \; R_2 = \frac{V_2}{I}, \; R_3 = \frac{V_3}{I}.$$

Dividing Eq. (27.3) by I throughout we get

$$\frac{V}{I} = \frac{V_1}{I} + \frac{V_2}{I} + \frac{V_3}{I},$$

but V/I is the resistance of the whole combination, which we can call simply R. Thus

$$R = R_1 + R_2 + R_3. \tag{27.4}$$

If three resistances are connected in parallel, as in Fig. 27.11, then the resistance of the combination is given by a formula different from the one above. We derive it as follows. Calling the currents in each of the three resistors I_1, I_2, and I_3, we know that the current entering the combination and leaving it, at the junction points, is given by

$$I = I_1 + I_2 + I_3. \tag{27.5}$$

But, as follows from the definition of resistance,

$$I_1 = \frac{V}{R_1}, \; I_2 = \frac{V}{R_2}, \; I_3 = \frac{V}{R_3}.$$

And if we call the resistance of the combination simply R, then $I = V/R$. Hence, substituting in Eq. (27.5) we get

$$\frac{V}{R} = \frac{V}{R_1} + \frac{V}{R_2} + \frac{V}{R_3},$$

and dividing this throughout by V we obtain

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}, \tag{27.6}$$

which we can call a reciprocal addition law, for resistances in parallel.

A knowledge of these two laws, for resistances in series and for resistances in parallel, is useful to anyone building electronic circuits. Suppose you require a 1 kΩ resistor for some particular purpose, but you have not got one. However you find that you have a 330 Ω and a 680 Ω resistor, each one marked with a silver ring indicating that it has a tolerance of 10%. Then, connecting the two resistors in series, you know that you have now a resistance of (330 + 680) Ω ± 10%, that is, 1010 ± 101 Ω. If a 10% tolerance resistor is good enough for your purpose, then this combination is good enough, because the total resistance could lie anywhere in the range 909 Ω to 1111 Ω. Similarly, if you wanted a resistance of 500 kΩ, but had not got this particular value, or anything near it, in your stock, you could make do with a pair of 1 MΩ resistors connected in parallel.

Chapter 28
ELECTRIC CHARGE

28.1 Electric charge
28.2 Capacitors
28.3 Discharge of a capacitor
28.4 Dielectrics and capacitors

28.1 ELECTRIC CHARGE

What is an electric current? In this section we take up this question again, having left it at the end of Chapter 25. So far none of our experiments have made the flow of electricity visible although we have implicitly assumed that there exists such a flow. Let us now ask this awkward question: assuming that an electric current is the flow of something we have agreed to call electric charge, how does this electric charge behave when not flowing? We know that we must have a complete electrical circuit in order for there to be a flow, and there must be a source of e.m.f. to maintain the flow. Does nothing at all, then, happen if we have a source of e.m.f. connected to a circuit which is not complete? Suppose that we have a very small break in a circuit consisting of long wires: we might imagine that when we connect a battery into this incomplete circuit (but do not close the small gap), that there is at any rate some movement of electric charge along the wires although there cannot be any continuous flow. If we wish to attempt detecting this movement we can make an arrangement like that of Fig. 28.1. When we close the switch S we observe whether the moving coil galvanometer, G, shows any deflection. In practice we will be almost certainly disappointed, even if the galvanometer is the most sensitive of its type in the laboratory. If we cannot obtain a more sensitive galvanometer, can we devise any way of enhancing the effect which we are trying to observe? Suppose we think of the electricity as a kind of invisible fluid which flows readily within the confines of a metal, and which is highly incompressible. Imagine the water-analogue of the circuit of Fig. 28.1. The water-analogue would have a pump instead of the battery, a flow-meter instead of the galvanometer, and pipes instead of wires. At the small gap in the circuit it is as if we had inserted an impermeable flexible diaphragm in the water pipe. When the pump is switched on the water cannot circulate and the flowmeter cannot therefore register any flow. But water is not entirely incompressible. Suppose that, if we connected the end A to a watertight vessel of large volume completely filled with water. We also connect a similar vessel to B. When

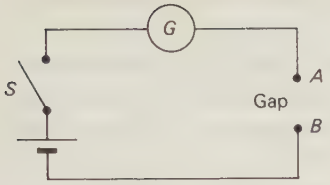


Fig. 28.1

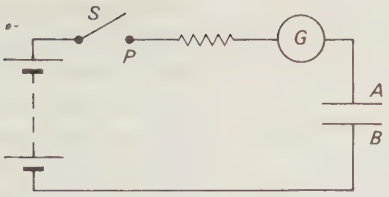


Fig. 28.2

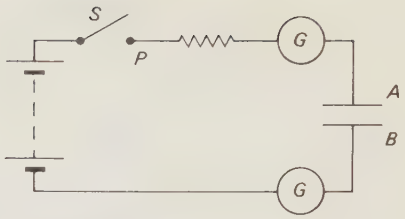


Fig. 28.3

the pump is started it provides a pressure which tends to compress the water into the large vessel joined to *A*. Since this vessel has a large volume the effect of the very slight fractional decrease in volume resulting from the increase in pressure may now be observable in the flow-meter as a momentary movement of a small volume of water into the vessel. There would be a similar momentary flow from the vessel joined to *B* into the pump. Analogous action to enhance the hoped-for momentary movement of the electrical fluid might be to attach a large amount of metal to points *A* and *B* in the electrical circuit of Fig. 28.1. In practice it is found that with two moderately sized metal plates (as little as 20 cm wide is adequate) set up with a small gap between them, insulated from each other, and connected to *A* and *B*, that there *is* an observable effect. A momentary deflection of the galvanometer is seen at the instant of switching on. The circuit of Fig. 28.2 shows the arrangement. If the switch *S* is opened and the point *P* connected to the negative terminal of the battery, a momentary deflection of the same magnitude, but in the opposite direction, is observed. This suggests that electric fluid flows on to the plate *A* when *S* is closed and then flows back when *P* is connected to plate *B* (or, possibly, that the electricity flows away from *A* when *S* is closed, then back on to *A* when *P* is joined to *B*).

The experiments in Fig. 28.2 show that having the plates *A* and *B* closer to each other results in larger deflections on the galvanometer when the switch *S* is closed. The investigation can be extended by inserting a second galvanometer into the circuit, as in Fig. 28.3. When *S* is closed (the plates having first been discharged by joining *P* to the negative terminal of the battery), the two galvanometers show deflections, usually of about equal magnitude, and in such a direction that if electric charge has flowed on to *A* it must have flowed away from *B*; alternatively if electric charge has flowed away from *A* then it must simultaneously have flowed on to *B*. We say that plate *A* has become *positively charged*, and *B* *negatively charged*, as a result of closing *S*. When *S* is opened and *P* is joined by a wire to the negative terminal of the battery, by-passing the battery, then the two galvanometers show equal deflections in the opposite direction to what they showed when *S* was closed. We interpret this result as the discharge of the plates *A* and *B*.

We have, in a sense, captured the electricity which hitherto had seemed only to exist in a state of motion. We can now investigate the properties of this static electricity charge and make quantitative measurements of it. We would like to find out first of all, perhaps, how many coulombs went on to plate *A* when we charged the plates by closing the switch *S* in the circuit of Fig. 28.3. But this measurement needs a new technique, because the deflections of the galvanometers in the experiments described above are momentary, and it is impossible to make any meaningful current measurement. We know, from previous discussion in Chapter 25, that to find the amount of electric charge which has flowed through a galvanometer coil we need to be able to measure a steady current through the coil and the time during which the current is flowing, and that the charge is the product (current × time). This approach is of no use if the current is not steady and the time is extremely short. Fortunately we may show that, when a charge flows in a very short time through a galva-

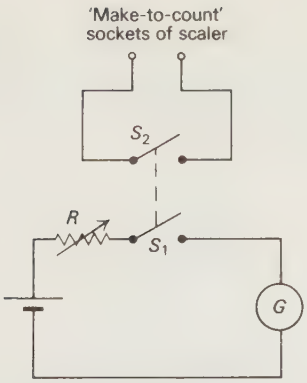


Fig. 28.4 Circuit used to calibrate a ballistic galvanometer for measuring electric charge.

nometer which has very little coil-damping, the *fling* (the amplitude of the momentary deflection) is directly proportional to the charge which has passed through the coil. Thus, if a galvanometer of this type can be calibrated by passing a known amount of charge through it and measuring the fling produced, it can be used to measure charge in experiments of the type we have been discussing.

A very quick and straightforward way of calibrating such a ballistic galvanometer involves the use of an electronic scaler to measure a very short time interval. Figure 28.4 shows the arrangement. A two-pole switch (S_1, S_2) is used. One gang of the switch is in a non-inductive circuit containing a low-voltage battery, a variable resistor large enough to reduce the current suitably, and the ballistic galvanometer. The other gang of the switch is connected to "make-to-count" sockets of the scaler so that, when the switch is closed, the scaler counts the number of millisecond pulses produced by its internal oscillator, and thus gives the time during which the switch is closed to the nearest millisecond.

The principle of the calibration is as follows.

a) When the switch S_1-S_2 is closed momentarily, a current whose strength depends on the e.m.f. of the battery and the total circuit resistance passes through the galvanometer for a time which can be accurately measured by the scaler. The fling of the galvanometer is measured.

b) The switch S_1 is now closed for a long enough time for the galvanometer to register a steady deflection: this gives the measurement of the current which must have flowed during (a) also. Multiplying this current by the time which the scaler registers in (a) one obtains the charge which passes through the galvanometer in (a). Further measurements of charge and fling can be made by altering either the variable resistor, in order to vary the current, or the time during which the switch is closed. A double pole snap-action toggle-switch is ideal and, with a little practice, a quick flick "on-off" by hand lasting only tens of milliseconds can easily be achieved.

A galvanometer used to measure charge (rather than a steady current) in this way is called a ballistic galvanometer. If it is to give accurate results it is essential that it has very little coil-damping, and that the time during which the charge flows through the coil is small in comparison with the period of free oscillation of the galvanometer coil. In carrying out the calibration as described above you need to be aware of these conditions, and it is worth extending the calibration experiment as follows, in order to make yourself aware of the limitations of a ballistic galvanometer.

Suppose that R has been set to give a steady current of 0.5 μA. If the switch is closed for a time of 50 ms, then the charge passed, in coulombs, is

$$(0.5 \times 10^{-6}) \times (50 \times 10^{-3}),$$

which is 2.5×10^{-8} C. The same quantity of charge could be produced by a current of half this value, 0.25 μA, flowing for twice the time, 100 ms. In either case the fling is found to be the same. This is difficult to verify in practice unless one has a means of switching a switch on for the desired, very short, precise time interval. However, with an ordinary toggle-switch one can cause a variety of different currents to flow for a variety of different times. Then, if a graph of the flings is plotted against the products (current $\times$ time), the points will be found to lie on a straight line. This shows that the time *by itself* does not affect the result. In a typical ballistic galvanometer

the period of free oscillation of the coil is just under two seconds, and provided the time during which the charge flows through the coil is not more than about a quarter of a second, the fling is directly proportional to the charge, within the limits of observational accuracy.

28.2 CAPACITORS

We proceed now to investigate the action of the parallel-plate arrangement as a holder of electric charge. This investigation will lead to the concept of *capacitance*. We have just discussed how a ballistic galvanometer can be used to measure quantities of electric charge: it is a well-established method of making such measurements. However, a different method for measuring charge will shortly be introduced because it is easier to perform than using a ballistic galvanometer.

Some quick, preliminary investigations of the charge-storing ability of a pair of parallel plates are conveniently carried out using the apparatus suggested in Fig. 28.5. S is a single-pole, two-way switch which can be operated quickly: an old-fashioned morse telegraph key is ideal. G is a ballistic galvanometer protected by a resistance R whose value should be about one megohm. The two plates can be of any shape, but should be accurately plane on their inner surfaces and of area 400 to 900 cm^2. The lower plate B is earthed. When switch S is in the left-hand position, plate A is connected to the positive terminal of the battery and becomes charged. When the switch is thrown to the right the plates discharge through G and R, and the charge can be measured by noting the fling of the galvanometer. The two plates are separated and insulated from each other by flat spacers of equal thickness and area about 0.5 cm^2, made of plastic insulating material: celluloid, polythene, PVC, are all suitable if clean and dry. The plates can be separated by three or four spacers evenly spaced from each other. The separation of the plates can be varied by using more spacers, stacked on top of each other. First of all, the plates are charged and then discharged using different voltages from the battery, and it is observed that the larger the voltage used, the greater the charge. Keeping the voltage constant one can then test the effect of altering the distance between the two plates: increasing the distance reduces the charge which is stored. The area of overlap of the two plates can be varied by sliding the top one horizontally away from the lower one: reducing the area of overlap is found to reduce the charge. The nature of the medium between the plates also affects the amount of charge which can be stored at a given voltage: this can be verified by inserting a sheet of the insulating material which was used to make the small spacers in such a way that the distance between the two plates is not altered. It is then found that substituting a solid plastic insulator for what was mostly air causes a marked increase in the charge. If the bottom plate is removed altogether (leaving the negative terminal of the battery still earthed, say, to a water-pipe, gas-pipe, metal mains cable sheath, or mains power-point earth socket and the upper plate insulated) it is found that the top plate alone can still store charge although considerably less than it did when the lower plate was present close to it. With the top plate alone it can also be shown that the presence of any other conducting object, close to the top plate but insulated from it, increases the charge-storing ability of the plate. Even your hand held at a distance of one or two centimetres (a steady hand is needed to avoid getting a shock!) from the plate demonstrates this fact nicely: the hand is effectively a good earthed conductor.

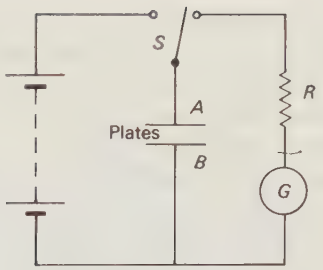


Fig. 28.5

These quick and rough experiments suggest more controlled and accurate ones. We need to investigate

carefully how the charge-storing ability of the parallel-plate arrangement depends on these factors.

a) The voltage used to charge the plates.
b) The distance between the plates.
c) The area of overlap of the plates.
d) The insulating material between the plates.

To carry out these investigations we shall use a technique which is easier than using a ballistic galvanometer. The technique employs a *reed switch*, a kind of miniature magnetic relay which can perform a large number of switching operations per second. First of all, keeping the arrangement of Fig. 28.5, we observe what happens if we operate switch S rapidly. If we perform about one charge-discharge operation every second the galvanometer registers distinct flings of approximately equal magnitude. Now speed up the operation of the switch to about five charge-discharge actions per second, and the galvanometer registers a reading which is almost steady: only a small pulsation at the rate of switching is seen. When demonstrating these effects it may be necessary to select a less sensitive range on the galvanometer, or exchange it for a less sensitive instrument. If the switching repetition rate is increased to as fast as the operator can perform (usually not more than about six or seven per second) then the galvanometer shows an even steadier reading. Now an automatic reed-switch can be substituted for the manually-operated one. If the reed-switch coil is driven from a low-voltage 50 Hz a.c. supply it will perform a hundred on-off operations per second. With the reed-switch in the circuit the galvanometer reading will be completely steady, and the current will be about ten times as great as it was when the manual switch was used. A few quick tests of the effect of varying the factors listed above will show that the steady galvanometer reading is affected in just the same way as the fling of the ballistic galvanometer. Thus it is reasonable to suppose that the steady current registered by the galvanometer is a measure of the charge released when the plates are discharged.

One final test before proceeding to accurate measurements. Bring your hand close to the upper plate, while the reed-switch is running, and observe the galvanometer reading, and then move your hand away again. This shows the effect of the hand on the charge-storing ability of the plates much more directly and convincingly than by using the ballistic galvanometer.

Suppose the charge released by the plates when they are discharged through the galvanometer is Q. The reed-switch causes n discharges per unit time. Thus the average current through the galvanometer is nQ and we will assume that this is the steady reading indicated by the galvanometer. If n is kept constant, then the current is directly proportional to the charge Q. Using a low-voltage a.c. supply derived from the 50 Hz a.c. mains ensures the constancy of n. Knowing the current sensitivity of the galvanometer, and knowing n to be 100 s^{-1}, one can calculate Q from the current readings.

The first accurate measurement involves setting the plates at a fixed separation and using different values of battery voltage V. It is found that the charge on the plates Q is directly proportional to V. Then, using a constant charging voltage, and setting the plates at different distances d apart, one finds that, provided d is not too large, that $Q \propto 1/d$. Then, keeping constant d and voltage, but altering the area of overlap, A, one finds that $Q \propto A$. Finally, keeping the geometry of the plates constant and using a constant voltage, but putting a solid insulator instead of air between the plates, one discovers that Q is increased. The amount by which it is increased depends on the *relative permittivity* of the material, a term which will be defined in Section 28.4.

The pair of parallel plates which we have used for these experiments is one example of a *capacitor*. A capacitor is any arrangement of a pair of electric conductors, insulated from each other, and designed as a device for storing electric charge. Its ability to store charge is specified by its capacitance. Capacitance is defined as the amount of charge stored per unit potential difference between the plates. (One generally refers to the conductors as "plates" even if they are in fact rolls of metal foil.) If the charge stored is Q when the p.d. is V then the capacitance is given by

$$C = \frac{Q}{V}, \qquad (28.1)$$

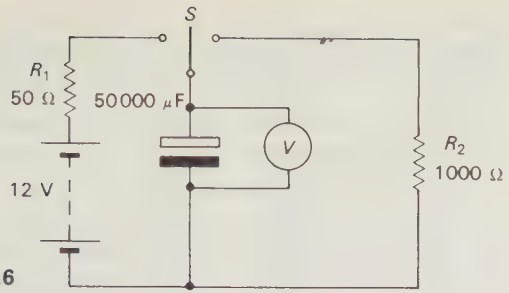


Fig. 28.6

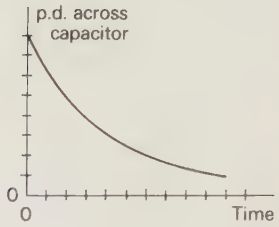


Fig. 28.7 The discharge of a capacitor through a resistor.

and, as the experiments have shown, this is a constant for a pair of conductors having a given geometry. The unit of capacitance is the farad (F) and one farad is one coulomb per volt. The farad is not a conveniently sized practical unit. The parallel-plate arrangement used in the experiments described has a capacitance of only a few millionths of a millionth of a farad. Submultiples in common use include the microfarad (μF) and the picofarad (pF). (See Appendix 2.)

A parallel-plate capacitor is of little practical use: its importance lies in what it can reveal in experimental use about the nature of the electric field. Electric field theory is dealt with in Unit 7. The uses of capacitors of various types in electric circuits are what concern us now.

28.3 DISCHARGE OF A CAPACITOR

The discharge of a capacitor is never an instantaneous process, although we assumed, when doing the experiments with the apparatus of Fig. 28.5, that the time taken for the capacitor to discharge itself was negligibly small. And we assumed this in spite of the fact that the reed-switch was discharging the capacitor one hundred times per second. We shall now make a careful investigation of the discharge process. This is most conveniently done using a very large capacitance. Figure 28.6 shows the circuit and suitable component values for this experiment. The capacitor is charged by putting the switch to the left-hand position. The resistor R_1, whose value is not at all critical, simply limits the initial charging current. The voltmeter should be one which draws negligible current and it should therefore have a resistance in the order of a megohm. A valve-voltmeter, or high-impedance transistorized voltmeter (see Section 29.2) is ideal. Capacitors whose values are as large as the one shown have low maximum working voltages (typically 12 V or 15 V) and this should never be exceeded. The capacitor is charged up to a suitable initial voltage: for a 12 V working capacitor an initial voltage of about 10 V would be suitable. Then the switch is thrown to the right-hand position and, with a stopwatch to hand, the experimenter takes readings of the p.d. across the capacitor at convenient intervals of time, for instance, every ten seconds. If the laboratory is fortunate enough to possess a pen-recording voltmeter or a suitable X-Y plotter, a lot of labour is saved. The result appears as in Fig. 28.7. The rate of discharge is rapid at first, continuously becoming less rapid. It appears to be a curve of the exponential decay type (see Chapter 19), but rather than take this for granted we shall make a theoretical analysis of the process.

At any instant during discharge the p.d., V, across the capacitor and the charge, Q, are related by

$$Q = CV,$$

where C is the capacitance. This follows from the definition of capacitance which can be written in symbols: $C = Q/V$. The current, I, flowing in the resistance R is given by

$$V = IR,$$

but the current is equal to the rate of loss of charge

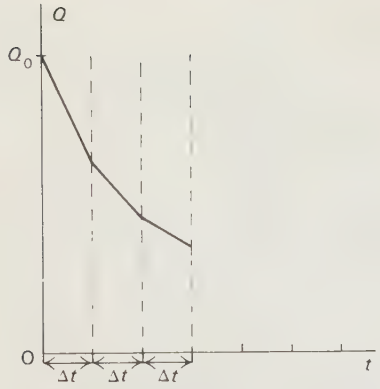


Fig. 28.8 Step-by-step computation of the p.d. across a capacitor during its discharge.

Table 28.1

Time, in seconds	Charge Q coulombs	P.d. V volts	Current I amperes
0	0.400	8.00	8.00×10^{-3}
10	0.320	6.40	6.40×10^{-3}
20	0.256	5.12	5.12×10^{-3}
30	0.205	4.10	4.10×10^{-3}
40	0.164	3.28	3.28×10^{-3}
50	0.131	2.62	2.62×10^{-3}
60	0.105	2.10	2.10×10^{-3}
70	0.084	1.68	1.68×10^{-3}
80	0.067	1.34	1.34×10^{-3}
90	0.054	1.07	1.07×10^{-3}
100	0.043	0.86	0.86×10^{-3}

from the capacitor, and thus we can write, in calculus notation:

$$I = -\frac{dQ}{dt}$$

the negative sign indicating that as time, t, increases, the charge, Q, *decreases*. Combining the three equations we obtain

$$\frac{Q}{C} = -R\frac{dQ}{dt}$$

or

$$\frac{dQ}{dt} + \frac{Q}{RC} = 0. \qquad (28.2)$$

This is a differential equation of the first order for which we shall carry out a step-by-step approximate numerical solution. There are two good reasons for doing this: firstly, it is a technique which can be used by the student who has no knowledge of calculus; secondly, it is a method easily programmed on a digital computer. Having obtained numerical values of V one can plot a graph and study the properties of the curve.

The steps of the computation are as follows.

a) Choose initial value of voltage, V_0 at time $t = 0$.
b) Calculate the charge, Q, using $Q = CV$.
c) Calculate the current, I, flowing in the resistor using $V = IR$.
d) Because the current is the rate of loss of charge from the capacitor, then $-dQ/dt$, assumed to remain constant over a short time interval, can be plotted as a downward sloping straight line on the graph (Fig. 28.8).
e) Using this straight line we can plot the point on the graph representing the charge on the capacitor at the end of the first time interval.
f) Use the new value of charge to calculate the p.d., as was done in step (b), and then follow the same steps.

Suppose we start the clock when the p.d. across the capacitor in Fig. 28.6 is 8 V: then the charge on the capacitor will be 0.4 C. The student should work through several steps of the computation, plotting the graph as he proceeds. Table 28.1 can be used to check the steps of calculation.

When the graph has been plotted (it is, of course, made up of straight-line segments which approximate to a smooth curve) we can apply a test to see whether or not it fits the exponential form. In Chapter 19 this test was applied to the decay curves for damped oscillatory motions: there it was shown that, if a quantity x decays according to an exponential law, then it is reduced by the same fraction in successive

equal intervals of time. For example, if the quantity is halved from x to $x/2$ in time $t_{1/2}$, then it will fall from $x/2$ to $x/4$ in the next time interval $t_{1/2}$, and then from $x/4$ to $x/8$ in the next interval $t_{1/2}$, and so forth. In this case, the time $t_{1/2}$ taken for the quantity to halve its value is known as the *half-life* of the decaying quantity. If the graph for the capacitor discharge has been plotted as far as $t = 100$ s it will be discovered that the p.d. across the capacitor is halved from 8.0 to 4.0, then from 4.0 to 2.0, and then from 2.0 to 1.0, in very nearly equal intervals of time which are, in this example, 31 s to the nearest second.

The exact solution to Eq. (28.2) is

$$Q = Q_0 \, e^{-t/RC} \qquad (28.3)$$

and the half-life $t_{1/2}$ is found by setting $Q = Q_0/2$ and solving for $t_{1/2}$, thus:

$$\frac{Q_0}{2} = Q_0 \, e^{-t_{\frac{1}{2}}/RC}$$

$$\therefore \; e^{t_{\frac{1}{2}}/RC} = 2$$

$$\therefore \; t_{1/2} = RC \ln 2 = 0.693 \, RC. \qquad (28.4)$$

The product RC is known as the *time-constant* of the circuit, and in this example it is $1000 \times (50\,000 \times 10^{-6})$ s $= 50$ s. Hence $t_{1/2} = 0.693 \times 50$ s $= 34.7$ s. The result obtained from our approximate numerical solving of the equation was less than this, the discrepancy being about 10%. We could make a better approximation to the truth by using a shorter time interval in the computation, using, for instance, $\Delta t = 5$ s instead of $\Delta t = 10$ s. If you have already decided to programme the computation on a digital computer, you may find it an interesting exercise to use these two values of Δt and see how much better a result is achieved by using $\Delta t = 5$ s. However, the coarse technique we have used suffices to show the method, and to demonstrate the exponential nature of the decay.

The time-constant of a circuit containing resistance and capacitance is an important parameter in electronics. At first it may seem surprising that multiplying together these two physical quantities, the one measured in ohms and the other in farads, produces units of time. It is easily shown that this must be so by writing each unit in terms of units from which it is derived, thus:

$$\Omega = \mathrm{V\,A^{-1}} = \mathrm{V\,(C\,s^{-1})^{-1}} = \mathrm{V\,C^{-1}\,s},$$

and

$$\mathrm{F = C\,V^{-1}},$$

therefore

$$\Omega \times \mathrm{F} = (\mathrm{V\,C^{-1}\,s}) \times \mathrm{C\,V^{-1}} = \mathrm{s}.$$

28.4 DIELECTRICS AND CAPACITORS

The capacitance of a pair of parallel metal plates, area 400 to 900 cm^2 and spaced a few millimetres apart, is in the order of a few hundred picofarads. This is far too small a capacitance for most applications of capacitors in electronics. A quick look at the lists of capacitors in the catalogue of a manufacturer of electronic components gives a good idea of the ranges of values most frequently used. If one attempted to make a one microfarad capacitor in the form of parallel plates with an air-gap in between, then the area of each plate would have to be hundreds of square metres. But because the capacitance can be increased by reducing the separation of the plates and substituting a solid insulating material for the air between the plates, a capacitor can be made compact by having two layers of thin metal foil separated from each other by a very thin layer of insulating material. Waxed paper has been used for many years, and is still used, in the type of capacitor known simply as a paper capacitor. You should, if given the opportunity, cut open and dissect an old or unwanted paper capacitor, and carefully separate from each other the layers of foil and paper.

Instead of waxed paper, mica and various ceramic materials (for example, titanium oxide) are used commonly as the dielectric material in capacitors. Mica is a naturally occurring mineral which has a crystalline structure such that it can be split into very thin sheets of almost perfectly uniform thickness. It can withstand high voltages, and high temperatures, and it is suitable for use in capacitors in circuits where there are ultra-high frequency (that is, in the order of hundreds of megahertz) alternating currents. The ceramic materials specially developed for use in

capacitors can also withstand high temperatures and voltages: they also have very high relative permittivities. The meaning of this term is discussed below. Paper, mica, and ceramic capacitors are not generally made with capacitances greater than about one microfarad. For appreciably larger capacitance one generally uses an electrolytic capacitor, in which the thickness of the dielectric layer between the layers of aluminium foil is very much less than can be achieved using paper, mica, or ceramic. In an electrolytic capacitor the layer of dielectric, which can be less than a thousandth of a millimetre thick, is produced by electrolysis. Paper impregnated with a solution of aluminium borate is used to separate the layers of aluminium foil. This solution is a conductor; but when a p.d. is applied to the two plates of the capacitor, the positive plate becomes *anodized*, that is, a very thin layer of aluminium oxide is deposited by electrolytic action on the surface of the metal. It is this layer of oxide which is the insulator between the two plates, and it is of course sandwiched between the positive aluminium foil and the layer of impregnated paper which makes direct electrical contact with the negative foil. This deposit of aluminium oxide has remarkably good insulating properties and, although it is so thin that one might suppose that it would break down when only a small p.d. is applied to the plates, in fact an electrolytic capacitor of this type can be made to work at voltages of several hundreds of volts. Electrolytic capacitors cover a range of capacitance from about 1 μF up to several thousands of μF. It is essential to connect such capacitors into circuits the right way round. (Why is this so?)

The different insulating materials used in capacitors have different *relative permittivities*. To understand what the term *relativity permittivity* (or *dielectric constant*) means, consider what was done in the experiment described in Section 28.3. The factors determining the capacitance of a parallel-plate capacitor were investigated, and it was shown that inserting a slab of solid insulating material between the plates, without altering their spacing, increased the capacitance. The dielectric, air, was replaced by a solid dielectric. If the plates were in a vacuum we would find that the capacitance was not noticeably different from that when air was present. In fact the capacitance with air is very slightly greater than with a vacuum.

The relative permittivity ε_r of a material is defined thus:

$$\varepsilon_r = \frac{\text{capacitance of an ideal parallel-plate capacitor with material between plates}}{\text{capacitance of same capacitor with plates in a vacuum.}}$$

The relative permittivities of some commonly used insulating materials are given in Table 28.2. These are simply numerical constants.

Table 28.2

Material	Relative permittivity
Polythene	2.3
Perspex	2.6
Paper (waxed)	2.7
Paraffin	2.7
Paper (oil impregnated)	3.6
PVC	4
Mica	7
Barium titanate	1200

The name relative permittivity arises in the following way. The capacitance of an ideal parallel-plate capacitor can be expressed in terms of the geometry of the plates thus:

$$C = \varepsilon \frac{A}{d}, \qquad (28.5)$$

where A is the area of each plate, d is the distance between the plates, and ε is a constant for the insulating material between the plates, called the *permittivity* of the material. If the plates are in a vacuum the constant is written ε_0 and is called the *permittivity of free space*. The significance of this constant, for free space, and how it is measured, are discussed in Unit 7. If the medium between the plates is anything other than a vacuum the value of the constant is greater than ε_0. If we use the symbol ε_x for the permittivity of a medium X, then the relative permittivity ε_r of

medium X is defined as the ratio $\varepsilon_x/\varepsilon_0$. Now suppose we have a parallel plate capacitor in a vacuum, and nothing between the plates, then we can write, for its capacitance

$$C_0 = \varepsilon_0 \frac{A}{d}. \tag{28.6}$$

Then if we insert a material whose permittivity is ε_x between the plates, the capacitance is given by

$$C_x = \varepsilon_x \frac{A}{d}.$$

Then the relative permittivity

$$\varepsilon_r = \frac{C_x}{C_0}. \tag{28.7}$$

So far nothing has been said about the fundamental physical reason why the insertion of a slab of insulating material between the plates of a capacitor has the effect of increasing its capacitance. The fact is that the electric field between the plates is modified by the presence of the dielectric. The atoms and molecules of all materials contain positively- and negatively-charged particles. In an insulating material these particles are not free to move away from their mean positions, but they can move a little, so that the distribution of positive and negative charge in the material can be altered slightly. In the presence of an externally produced electric field the distribution of charge in the material is changed. This process is known as *polarization.*

The uses of capacitors can be roughly classified as follows:

a) To store electric charge which, after charging, leaks through some resistance in a circuit at a rate determined by the capacitance and the resistance combined. This process can be employed in electric circuits for determining a time-interval. This category includes the smoothing action of a capacitor, used, with a rectifier, in the conversion of a.c. to d.c.

b) To store a small quantity of electrical energy. A large electrolytic capacitor is used in an electronic flash-gun, for use in photography. The capacitor is charged and then discharges very rapidly through a gas discharge tube (usually containing xenon gas). A small quantity of electrical energy is thus converted into light and some heat.

c) To filter out a.c. from d.c. In many applications a capacitor is used in such a way that it has, in effect, a very low resistance to an a.c. but a very high resistance to a d.c.

Chapter 29
USING CIRCUIT ELEMENTS

29.1 Some circuit elements
29.2 Circuit elements in combination

29.1 SOME CIRCUIT ELEMENTS

Resistors, cells, and capacitors are examples of *circuit elements:* they are self-contained devices which are commonly used in electrical circuits. So far we have been concerned principally with the function of these particular circuit elements: *what* they do rather than *how* they do it. We have not considered, for instance, the question of what actually happens, on the atomic scale of size, within a resistor when it conducts an electric current. Mechanisms of conduction are studied in Unit 8. In this section we shall study some additional circuit elements, concerning ourselves at this stage entirely with their function.

The conducting behaviour of a circuit element which has only two terminals can often be concisely described by its *voltage-current characteristic* (often referred to simply as *characteristic*). Figure 29.1 shows the voltage-current characteristics of three devices. The characteristics of nearly all circuit elements are dependent upon temperature, as Fig. 29.1(c) makes clear. This is often of great practical importance with semiconductor devices, as will become apparent later.

The voltage-current characteristic tells a circuit designer what current will pass through the element when a given p.d. is applied to it, or, put another way, what p.d. appears across it when a given current passes through it. The skill of the circuit designer lies in combining different circuit elements so that they produce the required overall function, and often much ingenuity is exercised in counteracting shortcomings in the behaviour of one circuit element by combining its function with another element which annuls, partly if not totally, those shortcomings. For instance, the often very undesirable sensitivity of transistors to changes of temperature can be counteracted by pairing suitable transistors together so that the temperature effects cancel each other.

We shall now discuss a few simple circuit elements and try to show how their functions are useful in practice, and to explain the principles of the designing of circuits.

The first is the *rectifier diode.* The name *rectifier* is given to any device which converts an alternating

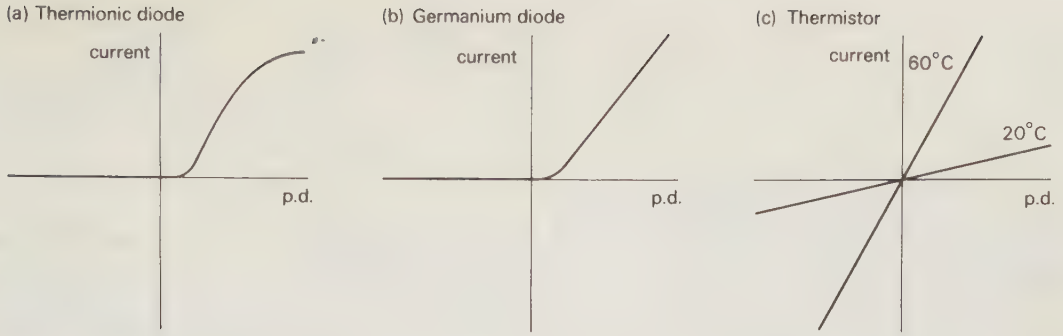


Fig. 29.1 Characteristics of some circuit elements.

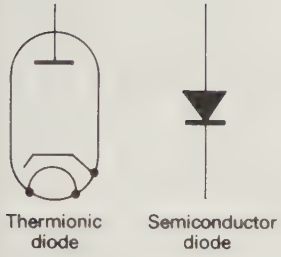


Fig. 29.2 Circuit symbols for two types of rectifier diode.

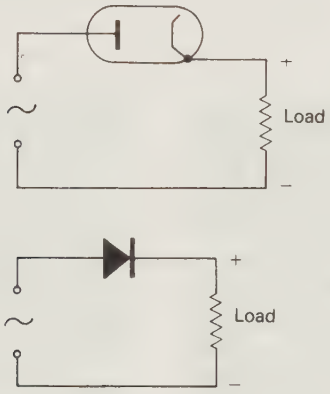


Fig. 29.3 Basic circuits for rectification.

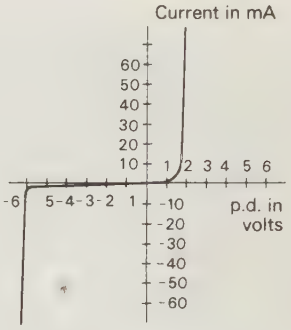


Fig. 29.4 Characteristic of a typical zener diode.

current into a direct current. "Diode" is a modern word constructed from Greek roots to signify "two electrodes." The two main categories of rectifier diode are *thermionic* diodes and *semiconductor* diodes. In the thermionic diode one electrode, the cathode, has to be heated in order for the diode to function. Its action is described in Unit 8 where the thermionic emission of electrons is discussed. For low-power rectification of a.c. thermionic diodes are nowadays much less used than semiconductor diodes. Figure 29.1 shows the voltage-current characteristics of two typical diodes of each type. Figure 29.2 shows the conventional circuit symbols for the two kinds of diode. In Fig. 29.3 each type is shown incorporated in the simplest "half-wave" rectifying circuit.

An "ideal" diode would conduct perfectly a current in one direction, that is, it would have zero resistance to current in one direction; and it would not conduct at all a current in the reverse direction. The characteristic of such a diode would therefore be a straight line lying exactly along the voltage axis, the whole length of the negative voltage region, becoming at the origin of the graph a straight line lying precisely along the positive current axis. Although the graphs of Fig. 29.1 do not show it, the thermionic diode's characteristic has a slight "tail" stretching into the negative voltage region, because it still conducts a slight forward current even when small reverse voltages are across it. Any semiconductor diode conducts *slightly* in the *reverse* direction, when reverse

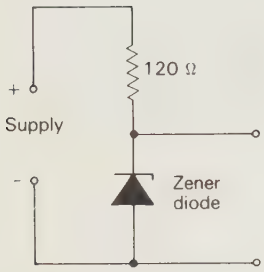


Fig. 29.5 The simplest form of voltage stabilizing circuit using a zener diode.

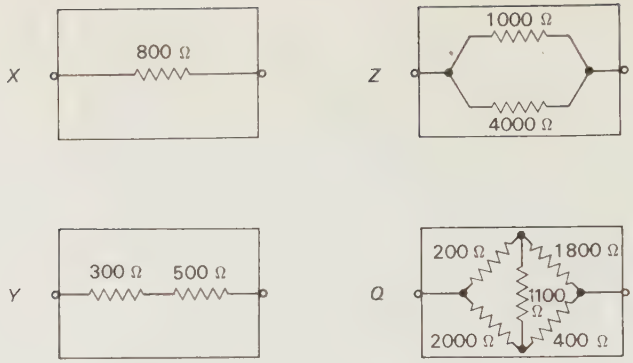


Fig. 29.6 The resistance between the terminals is the same for each box.

voltages are applied, the reverse current being greater for larger reverse voltages. Typically, a small semiconductor diode which is used to conduct forward currents in the order of one ampere will have a reverse current in the order of several microamperes.

If a sufficiently large reverse voltage (the breakdown voltage) is applied to a semiconductor diode it will suddenly conduct strongly and most probably become quickly overheated and permanently damaged.

A special type of semiconductor diode, called the Zener diode, is designed to have a specially low value of breakdown reverse voltage. Zener diodes are made from silicon, and Fig. 29.4 shows the characteristic of such a diode. If a small reverse voltage is applied to a Zener diode it scarcely conducts at all (just like a normal rectifying diode) but when the breakdown voltage is reached, very little further increase of voltage is needed to produce a large current. This property enables the Zener diode to be used in voltage stabilizing circuits. Figure 29.5 shows the principle upon which voltage stabilizing circuits can be designed. Suppose the Zener diode shown has a Zener voltage (that is, its breakdown voltage) of 6.0 V: it is being used here to provide a reliably constant voltage of 6.0 V. The supply is, say, nominally 12 V; but it may in fact vary from 12 V to 9 V. When the supply voltage is 12 V a current of 50 mA flows through the diode and the resistor, and thus the p.d. across the resistor is 6 V and so also is the p.d. across the Zener diode. When the supply voltage drops to 9 V, the current also drops, but, as Fig. 29.5 shows, there is very little corresponding change in the p.d. across the Zener diode: it will stay (to the nearest 0.1 V) at 6.0 V, and the p.d. across the resistor will then be 3 V, and the current flowing in the diode and resistor will be 25 mA.

What is more, if current is taken from the terminals (in Fig. 29.5) connected across the Zener diode, say, 10 mA, then the current through the diode will fall by 10 mA. Thus the current through the resistor will remain unchanged, and hence the p.d. across the resistor, and the p.d. across the diode, stay very nearly constant.

The *thermistor* is a semiconductor device whose resistance depends upon temperature in a sensitive way. The characteristic of a typical thermistor is shown in Fig. 29.1(c). Unlike a metal conductor, its resistance decreases with rise in temperature, falling by about 75% as the temperature rises from 25° C to 60° C. Contrast the temperature sensitivity of a pure metal, such as copper; the resistance of a pure copper wire would increase by approximately 20% for the same rise in temperature.

A thermistor, unlike a diode, has no rectifying property: it conducts equally well a current in either direction through it. Certain types of thermistor are

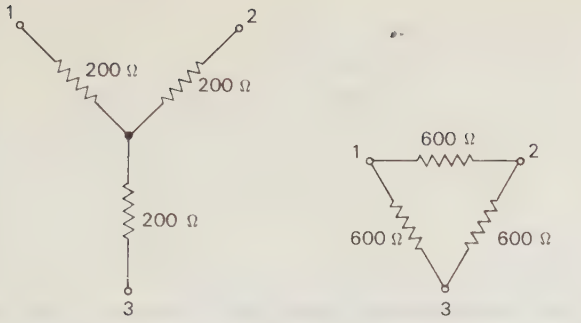


Fig. 29.7

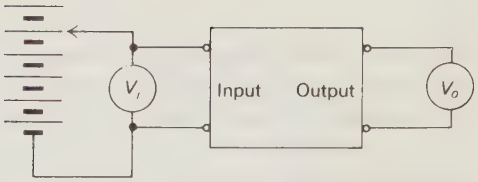


Fig. 29.8 Arrangement for measuring the input-output characteristic of the "black box."

used in electrical temperature-measuring or controlling devices; other types are used to limit surge-currents in circuits.

Another useful semiconductor device is the *light-dependent resistor* (LDR), sometimes called a *photo-sensitive resistor*, generally made of cadmium sulphide. It behaves as an ordinary resistor, but the resistance depends upon the amount of illumination incident upon it: the greater the illumination, the smaller the resistance. A number of ingenious photo-electronic devices can be designed around an LDR, ranging from burglar alarms to exposure meters. Problem 27.1 concerns such a resistor.

29.2 CIRCUIT ELEMENTS IN COMBINATION

The job of a circuit designer involves putting together a number of circuit elements in order to perform a function which cannot be performed by a single circuit element. Look at the circuit shown in Fig. 24.3. This device is an integrated circuit which is supplied ready-made, encapsulated, and fitted with connecting terminals. It required an expert to design that circuit, which combines the functions of a large number of different components. Obviously we cannot here explain in full how such a circuit is designed. But we can lay the foundations of an understanding of the principles of circuit design, and get you to think like a circuit designer.

To begin thinking in the right way, consider a simple practical problem. There are three small 'black boxes', labelled X, Y, and Z. Each box is fitted with a pair of terminals. The problem is to deduce what is inside each box, having made electrical measurements externally. We are told that each box contains nothing but resistors, and that the contents of each box is distinctly different. Suppose now that we take box X and measure the resistance between its terminals. We find that it is 800 Ω. We then take box Y: again we find that its resistance is 800 Ω, and finally, we discover that box Z has precisely the same resistance. We know, of course, that a resistance of 800 Ω can be produced by combining two or more different resistances, either in series or parallel: the contents of the box does not have to be a single resistor whose value is 800 Ω. We now examine what each box contains, and find that they are as shown in Fig. 29.6.

Of course, values different from those in boxes Y and Z could have been used to produce an overall resistance of 800 Ω. Box Q may be shown to have an overall resistance of 800 Ω. (You may prove this if you wish.) The examples remind us that there is in fact an infinity of possible ways of combining resistors to get a particular desired value.

If the measured resistance of all the boxes of Fig. 29.6 is the same, is there any way of telling, by means of external tests only, the difference between

their contents? The answer is no. All the circuits are, electrically, precisely equivalent.

We turn now to a slightly more complicated situation. Consider the two three-terminal circuits shown in Fig. 29.7. A little calculation will show that, for each circuit, the resistance between any pair of terminals is 400 Ω, and in this respect the two circuits are identical. Perhaps it seems that, in this case, there *might* possibly be some external measurements which would reveal the difference between the two circuits. This is something you should think about. Think, for instance, of what would happen if you applied a known voltage across terminals 1 and 2 and at the same time measured the voltage across terminals 2 and 3: would this reveal the difference between the two circuits?

We shall come closer still to the real problems besetting the circuit designer if we study some simple four-terminal boxes. Suppose that, on each of our four-terminal boxes, one pair of terminals is labelled "input" and the other pair labelled "output." One way of testing the function of such a box is suggested in Fig. 29.8: different voltages are supplied from a battery to the input pair of terminals with a voltmeter being used to measure the input voltage, V_i; at the same time the output voltage, V_0, is measured with another voltmeter.

Suppose that the contents of one particular box is (Fig. 29.9) a simple voltage-divider circuit. We know that whatever input voltage is applied to the left-hand pair of terminals, the output voltage will be given by

$$V_0 = \frac{1000}{2000 + 1000} V_i = \frac{V_i}{3}.$$

We could say that the function of this 'black box' is to reduce any given voltage to one-third of its value. A name given to a voltage-reducing device (of which this is the simplest possible example) is *attenuator*. An attenuator is often used in electronics, connected to the input of an amplifier for instance, to avoid overloading the amplifier with too large a signal.

By means of external voltage measurements alone we could discover the voltage-reducing property of this particular box. However, the same voltage-reduction ratio could be achieved with any pair of resistors inside the box provided that their resistances were in the ratio 2 : 1. Clearly, to determine the actual value of the resistances we should need to make resistance measurements across each pair of terminals separately. Problem 6.11 at the end of this unit involves deducing the contents of a similar 'black box' from the results of external measurements.

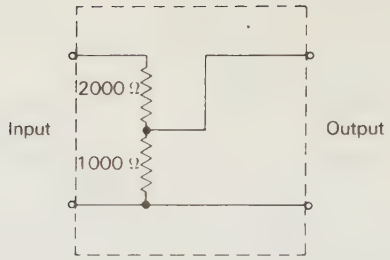


Fig. 29.9 A simple "black box." The output voltage will always be one third of the output voltage.

We now have to tackle a practical difficulty which can arise when making voltage measurements on circuits of the type we are discussing. This difficulty is best illustrated by the results of a simple experiment which present an apparent paradox: by solving the paradox we shall be able to understand why voltage measurements on certain circuits can be misleading.

The situation is this. We have three voltmeters: the best, and most expensive (a), is a large multi-range meter; the second (b) is a cheap laboratory voltmeter; and the third (c) is a small voltmeter salvaged from some old radio equipment. Each of the three reads to 5 volts full scale (the multi-range meter has its selector switch set on the 5 V range). First of all we check the three meters by connecting them in turn to a 4.5 V dry battery (comprising three 1.5 V dry cells). We find that they all read 4.5 V, and so we are satisfied with their accuracy. Then we decide to make some voltage measurements with the 'black box' of Fig. 29.9. We connect a 12 V battery to the input terminals of the box, and measure the output voltage with each of the voltmeters in turn. The results we obtain are as follows:

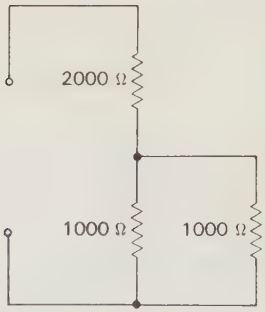


Fig. 29.10 This circuit is equivalent to the circuit of Fig. 29.9 when the voltmeter is connected across the output terminals.

a) multimeter 4.0 V;
b) lab. voltmeter 2.4 V;
c) ex-equipment voltmeter 3.5 V.

Here we have discrepancies which are certainly serious! You may have guessed a reason for the disagreement: but consider the results of further measurements. The multimeter seems in order but the two other meters seem to be reading low. Suppose now we connect meters (a) and (b) simultaneously (i.e. in parallel) to the output terminals of the box: we find that they both read 2.4 V. Likewise, when we connect (a) and (c) simultaneously they both read 3.5 V. Then, when (b) and (c) are connected simultaneously we find that they agree with each other, but this time the reading is 2.2 V, slightly less than the reading obtained with (a) and (b) together. The truth of the matter is that, although none of the meters is inaccurate, the very act of connecting a voltmeter to the terminals of the box alters the voltage between those terminals. This happens because the voltmeter itself has a finite resistance and when we connect it across the terminals we are thereby adding a resistance in parallel with the output terminals of the box. Let us see exactly how this happens in one particular instance. Figure 29.10 represents the situation when voltmeter (b), which incidentally has a resistance of 1000 Ω, is connected to the output terminals. We shall predict the p.d. which should appear across the terminals. The combined resistance of the lower resistor of the voltage divider and the voltmeter is given by

$$\frac{1}{R} = \frac{1}{1000} + \frac{1}{1000}$$

whence

$$R = 500\ \Omega.$$

Hence the p.d. across this parallel pair is given by

$$V = \frac{500}{2000 + 500} \times 12 = \frac{12}{5} = 2.4\ \text{V}.$$

We can safely say, therefore, that this particular voltmeter is simply not suitable for measuring voltages in this particular situation. The voltmeter (c), which (we can now reveal) has a resistance of 5000 Ω, does not cause so serious a disturbance of the p.d.'s in the circuit, but it is nevertheless unsuitable. The multirange meter (a), however, has a resistance of 10^5 ohms: connecting it in parallel with the 1000 Ω resistor of the 'black box' produces a combined resistance of very slightly less than 1000 Ω, and the change in voltage caused by this is negligibly small. The meter is nothing like sensitive enough to detect such a change. An ideal voltmeter would be one which possessed infinite resistance. A good approximation to this ideal is the *valve voltmeter* which is a voltmeter incorporating a d.c. amplifier so that its resistance becomes effectively thousands of megohms, or even greater. Valve voltmeters are becoming obsolete, and the modern equivalent is a voltmeter which incorporates a transistor amplifier. But, in the absence of such a very high resistance voltmeter,

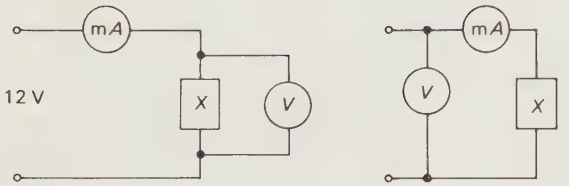


Fig. 29.11

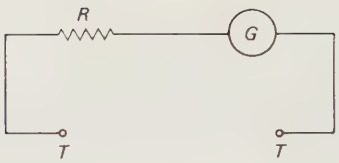


Fig. 29.12

what determines whether or not a particular meter is suitable for a particular situation? The foregoing discussion should make the answer clear. If the connection of a voltmeter to a circuit is not to alter appreciably the p.d.s in the circuit, the voltmeter's resistance should be in the order of about a thousand times the resistance of the part (or parts) of the circuit across which the p.d. is to be measured. Thus, a voltmeter like (b), having a resistance of only 1000 Ω, is really only suitable for measuring the p.d. across something which has a resistance of not more than a few ohms at most. It is satisfactory for measuring the p.d. across the terminals of a battery because the internal resistance of a battery whose e.m.f. is 5 V or less is not likely to be greater than an ohm or so. But suppose you want to check the voltages supplied to the electrodes of a cathode ray tube in an oscilloscope, where the voltages are derived from a voltage divider in which the resistors are in the order of one megohm. A valve (or transistor) voltmeter would be the only suitable instrument in this case.

Having learned to exercise caution in the choice of a voltmeter when making measurements on a live circuit, we may wonder if similar care is needed with ammeters. The problem of choosing a suitable ammeter is generally much simpler to solve. When we use an ammeter to measure the current in a part of a circuit we have first to break the circuit and insert the ammeter into the break. We want the current which flows when the ammeter is present to be no different from the current when the ammeter was absent. If the addition of the ammeter appreciably increases the resistance of the circuit in which it has been inserted, then it will, in general, reduce the current appreciably. So we require simply that the resistance of the ammeter should be much less than the resistance of the part of the circuit in which it is to be inserted. But, typically, the resistance of a cheap ammeter reading up to 5 A will be as little as 0.01 Ω, and this is considerably lower than the resistance of *most* circuits in which one is likely to want to measure current with a cheap ammeter! Problem 29.1 concerns a situation where the measurement of current in a circuit would be difficult because of the finite resistance of the ammeter.

Problem 29.1 Suppose you are going to make measurements of current and voltage on some electrical devices. You have a milliammeter whose resistance is 5 Ω (reading up to 15 mA) and a voltmeter whose resistance is 1 kΩ (reading up to 15 V). Let us say that the device *X* has a resistance of about 5 kΩ, and that you have connected up the circuit as shown. Calculate the readings of the meters in the situation shown in Fig. 29.11(a) when the supply voltage is 12 V.

Consider the alternative arrangement shown in Fig. 29.11(b). In what way (if any) is it better than the previous circuit? Explain carefully.

For what kind of electrical device would the first arrangement of ammeter and voltmeter be suitable?

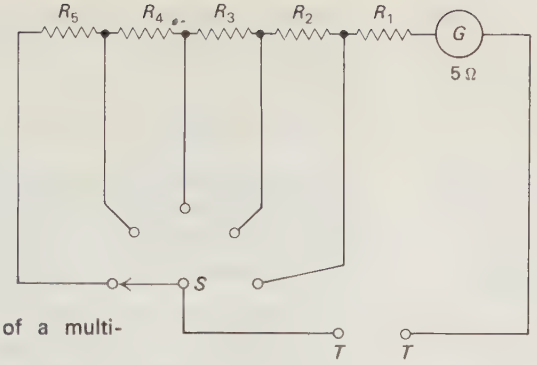


Fig. 29.13 The circuit of a multi-range voltmeter.

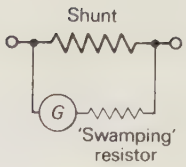


Fig. 29.14 In an ammeter, the effects of variation of coil resistance as a result of temperature change can be reduced by means of a "swamping" resistor.

Problem 29.2 A typical moving-coil galvanometer has a resistance of 5 Ω and needs a current of 100 μA to produce full scale deflection. It can be used, in certain limited applications anyway, as a voltmeter for measuring small voltages, because the current in the meter, and hence the scale-reading, will be proportional to the voltage applied to its terminals, because the meter itself is an ohmic conductor.

a) Calculate the voltage which, applied to the meter terminals, produces full scale deflection, expressing the answer in millivolts. To convert the meter into a voltmeter capable of reading higher voltages we connect a resistance in series with it, as in the diagram (Fig. 29.12).

b) Calculate the value of R needed to make the meter give full scale deflection when the voltage applied to the terminals is 5 V.

c) In a multi-range voltmeter a chain of resistors is used, as shown in Fig. 29.13. By means of the rotary switch S different resistances can be connected in series with the meter. Suppose we want voltage ranges of 5, 25, 100, 250, and 500 V. Calculate the values of the resistors needed, as shown in the diagram.

Solution

a) Using $V = IR$, we get $V = (100 \times 10^{-6}) \times 5 = 0.5 \times 10^{-3} = 0.5$ mV.

b) $5 = (100 \times 10^{-6}) \times (R + 5)$, hence $R + 5 = 5 \times 10^4 = 50\,000\ \Omega$, and so $R = 49\,995\ \Omega$.

c) R_1 is clearly the resistor which gives the 5 V scale, and so it is 49 995 Ω, as already calculated.

For the 25 V scale the second position of the switch is used (going anticlockwise), bringing R_1 and R_2 into the circuit. Hence

$$R_1 + R_2 + 5 = \frac{25}{100 \times 10^{-6}} = 250\,000\ \Omega$$

and since $R_1 + 5 = 50\,000\ \Omega$, this means that R_2 must be 200 000 Ω (200 kΩ). To find R_3, we know that $R_1 + R_2 + 5 = 250\,000\ \Omega$, so

$$R_3 + 250\,000 = \frac{100}{100 \times 10^{-6}} = 1\,000\,000\ \Omega.$$

Hence $R_3 = 750\,000\ \Omega$ (750 kΩ). In a similar way we find that $R_4 = 1.5$ MΩ and $R_5 = 2.5$ MΩ.

Problem 29.3 Errors can arise, when using a moving-coil meter, as a result of changes in resistance of the coil (generally made of copper wire) due to changes in temperature. Firstly we shall see how great these errors can be, and then we shall see how they can be overcome in one particular instance.

a) Consider a moving coil meter whose coil has a resistance of 10.00 Ω at a temperature of 20° C. The temperature coefficient of resistivity of standard annealed copper is $43 \times 10^{-4}\,°C^{-1}$: this is also the *fractional* change in resistance, of any particular copper wire, per degree change in temperature. Use this data to find the increase in resistance when the coil warms

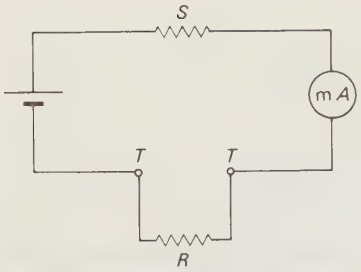


Fig. **29.15** A simple ohmmeter.

from 20° C up to 40° C, and hence find the resistance (to four significant figures) at 40° C.

b) Suppose our meter gives full scale deflection for a current of 100 μA. What must the voltage, in microvolts, across the coil be, at 20° C, to produce full scale deflection?

c) Suppose the same voltage as calculated in (b) is applied to the coil when the temperature is 40° C: what will the current be now? Express the difference between this reading and the reading at 20° C as a percentage of the reading at 20° C. This error is excessive if the meter is a good-quality one in which the true current is guaranteed to be within, say, $\pm 2\%$ of the scale reading.

d) The effect of this temperature-dependent change in resistance can be reduced by connecting in series with the coil a resistor which changes very little with temperature, one made of constantan, for example.

Suppose we have a constantan resistor of 90.00 Ω connected in series with the coil, and that we can neglect any changes in resistance of this resistor due to changes in temperature. The total resistance at 20° C is thus 100.00 Ω. Calculate the voltage which must be applied to the combination coil + resistor to produce full-scale deflection at 20° C now.

e) Suppose this same voltage is applied to the meter when the temperature is 40° C. What current flows now? Express the difference between this current and the current at 20° C as a percentage of the current at 20° C, and comment on the result.

The constantan series resistor is generally known as a swamping resistor, and such a resistor is generally used in a meter designed to have a low-resistance shunt connected to it, to convert it into an ammeter, as in Fig. 29.14.

Solution

a) The increase in resistance is calculated thus:

$$\begin{pmatrix}\text{fractional increase in resistance per degree}\end{pmatrix} \times \begin{pmatrix}\text{resistance at 20° C}\end{pmatrix} \times \begin{pmatrix}\text{rise in temperature}\end{pmatrix}$$

putting in numerical values, we get $(43 \times 10^{-4}) \times 10.00 \times 20 = 0.86\ \Omega$. Hence the resistance at 40° C is 10.86 Ω, to 4 significant figures.

b) Using $V = IR$. Voltage needed to produce fsd $= (100 \times 10^{-6}) \times 10.00 = 10^{-3}$ V, or 1 mV.

c) At 40° C the current is found by using $V = IR$, hence $I = V/R$, thus:

$$I = \frac{10^{-3}}{10.86} = 92.1\ \mu\text{A}.$$

The discrepancy is thus $100 - 92.1 = 7.9\ \mu$A, and expressing this as a percentage we find it to be

$$\frac{7.9}{100} \times 100 = 8\%,$$

to nearest percent.

d) With a total meter resistance of 100.00 Ω we need a voltage of $(100 \times 10^{-6}) \times 100 = 10^{-2}$ V, or 10 mV, to produce full scale deflection.

e) If this voltage is applied to the meter at 40° C, when the resistance will be 100.86 Ω, the current will be given by

$$I = \frac{10^{-2}}{100.86} = 99.1\ \mu\text{A}.$$

Thus the error is now 0.9 μA, and consequently a percentage error of 0.9%, or 1% to the nearest percent, and this is generally within the acceptable limits.

Problem 29.4 Figure 29.15 shows the simplest form of meter for measuring resistance directly. The resistance to be measured in connected to the terminals *TT*, the scale of the milliammeter being calibrated to read the resistance directly. Suppose we have a

meter which reads to 15 mA full-scale, an angular deflection of the needle of 90°; the coil resistance of the meter being 5 Ω, and the battery (B) voltage 1.5 V. We shall now see, working step-by-step, ho v to construct the ohm-meter.

a) Clearly, the larger the resistance R connected to the terminals TT, the less the current. Hence if the current scale is calibrated, as is usual practice, from left to right, the resistance scale will have its zero on the right. We want the zero of the resistance scale to correspond to full-scale deflection. Thus when the angular deflection of the needle is 90°, $R = 0$. The current flowing in the circuit here is, of course, 15 mA. Calculate the value of S needed.

b) What is the value of current in the circuit when $\theta = 75°$? Knowing this, calculate the value of R which will produce this deflection of the needle. Enter the answers in a table like the one below.

c) Repeat the calculation for the other values of θ given in the table. Make a rough sketch of the new meter scale, marked off in intervals of 50 Ω, and comment upon the form of this scale.

Pointer deflection, θ	0°	15°	30°	45°	60°	75°	90°
Current (mA)							
R (Ω)							

When you next see a multi-meter (with an ohms scale), look at it critically.

Solution

a) Using $V = IR$. The total circuit resistance (assuming the battery has negligible internal resistance) is $(S + 5\ \Omega)$: thus $1.5 = (15 \times 10^{-3}) \times (S + 5)$, hence $S + 5 = 100$, therefore $S = 95\ \Omega$.

b) When $\theta = 75°$ the current is $\frac{75}{90} \times 15 = 12.5$ mA. To find the value of R:

$$\text{total resistance} = S + 5 + R = \frac{1.5}{12.5 \times 10^{-3}}$$

but $S = 95\ \Omega$, therefore

$$100 + R = 120$$

hence

$$R = 20\ \Omega.$$

c) We can save time in the long run by making a formula for R in terms of θ. The current needed to produce a deflection θ is $\theta/90 \times 15$ mA, that is, $\theta/6$ mA, or $\theta/6 \times 10^{-3}$ A. Proceeding as in (b), we can write

$$100 + R = \frac{1.5}{\frac{\theta}{6} \times 10^{-3}}$$

which reduces to

$$R = \frac{9000}{\theta} - 100.$$

It is a quick business now to substitute values of θ into this and find the corresponding values of R.

Chapter 30
ELECTROMAGNETISM

30.1 Electromagnetic forces
30.2 Electric motors

30.1 ELECTROMAGNETIC FORCES

Electromagnetic forces are the forces between moving electric charges. As we shall see in Unit 7, the theoretical understanding of these forces was developed much later than the theory of *electrostatic* forces which, as the name implies, are the forces between static electric charges. These latter forces are often known as *coulomb* forces, after Charles Augustin Coulomb who investigated their nature in the eighteenth century. In this section we are not concerned with the development of the theory of electromagnetic forces: we shall simply state what is known about them and discuss in some detail a variety of practical situations in which these forces are manifested or put to a useful purpose.

In 1820 Oersted discovered that an electric current in a wire influenced a magnetic compass-needle placed near to it. In his experiments the force of interaction between a current and a magnet was revealed. In the same year, Ampère demonstrated to the Académie des Sciences the force of interaction between two coils of wire carrying electric currents: this was just three weeks after Arago had demonstrated Oersted's experiments to the same academy. Ampère's name has been used for many years for the fundamental unit of electric current, but it is only relatively recently that this unit of current has been defined in terms of the force between two wires carrying a current, the theory of which was developed for the first time by Ampère between 1820 and 1830. We shall not retrace the details of Ampère's work here, but many of the statements we shall make were made, although not perhaps in quite the same way, by Ampère himself.

We investigate the characteristics of the forces between current-carrying conductors by means of a simple current balance (Fig. 30.1). The two conductors can be connected in series so that the current flows in the same direction through each. When this is done the conductors attract each other. When the connections are altered so that the current flows in the opposite direction in one conductor, there is a force of repulsion. That the size of the force depends on the strength of the current is easily demonstrated by adjusting a rheostat in the circuit. To measure

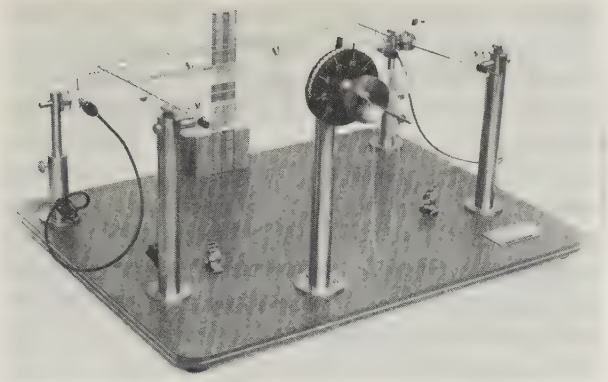


Fig. 30.1 A current balance. (Courtesy of G. Cussons, Ltd.)

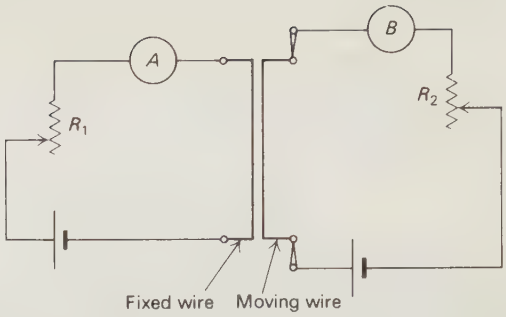


Fig. 30.2 Using the current balance to investigate the electromagnetic forces between two current carrying conductors.

the force between the conductors one can proceed as follows: with the current switched off adjust the chaindial counterweight on the balance arm until it is horizontal. Switch on the current and then re-adjust the chaindial until the arm is restored to its original position. The change in the reading of the chaindial is a measure of the required force.

To make more detailed tests we need an independent means of monitoring the strength of the current. For this we can use an uncalibrated ammeter (or an ammeter with its scale-markings obscured). The reason for using an uncalibrated ammeter is that we wish to lead up to the idea of using a current balance for making an absolute measurement of a current—a measurement which does not employ a calibrated instrument. Figure 30.2 shows the circuit arrangements. The uncalibrated ammeters, A and B, will be used simply to indicate whether or not the currents remain constant.

The moving conductor is included in a circuit which includes the uncalibrated ammeter B and the rheostat R_2. This enables us to maintain a constant unknown current in this moving conductor.

The fixed conductor is included in a second circuit which includes the uncalibrated ammeter A, the rheostat R_1 and a long flexible insulated lead. This is long enough to be laid close to and parallel to the fixed conductor without disturbing the circuit in any way.

The rheostat R_1 is adjusted to give a value of the current which produces a deflection on A of a little less than half full scale. The force between the two current-carrying conductors is measured and the position of the needle of A is marked.

The flexible conductor is then laid alongside the fixed conductor in such a way that the currents in the two conductors are in the same direction and flowing side by side. Having checked that the current as indicated by A is unchanged, the force is again measured. Not surprisingly we find that it is just double what it was originally. We can safely assume that, if we repeated the experiment with three fixed conductors laid side by side instead of two, we would get three times the original force. Returning now to using a single fixed conductor, we find that we can get two units of force simply by adjusting the rheostat to increase the current in it. We do this, and make a second mark to record the new pointer reading of A. Thus it is perfectly reasonable to say that a current in a single conductor which produces two units of force is exactly equivalent to the sum of the currents in two conductors side by side where each current produces one unit of force. The results of this experiment may seem obvious: the idea of the magnetic effects of currents being arithmetically additive in this way is, after all, already familiar in the form of Kirchhoff's first law. The simple experiments demonstrating

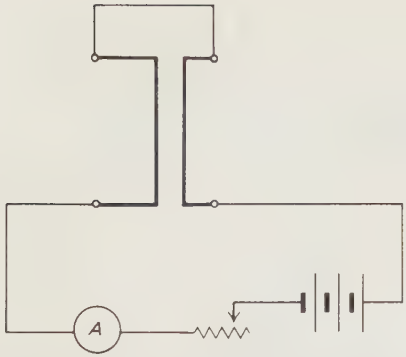


Fig. 30.3 In this circuit the same current flows in the moving and in the fixed wire.

Kirchhoff's law (Section 25.1) make use of the magnetic effect of an electric current, but in those experiments, calibrated ammeters were used—ammeters which, ultimately, had been calibrated by means of a current balance. The experiment just performed confirms the validity of our deduction of Kirchhoff's first law from those earlier experiments with ammeters.

This is not quite the whole story about the additive effects of currents. We should also demonstrate the effect of connecting two fixed conductors, lying side by side, in series and in such a way that the same current flows in opposite directions in each wire. When we do this we find that the force between the double, fixed, conductor and the moveable one is precisely zero. Thus two equal and opposite-going currents, flowing in paths which are a negligible distance apart, produce magnetic effects which exactly cancel each other.

Our uncalibrated ammeter A now has two marks on it, corresponding to one unit and two units of current. We can use this to make the remaining investigations. So far we have only varied the current in the fixed conductor: the current in the moving conductor monitored by the uncalibrated ammeter B, has been kept unaltered. Using the single fixed conductor, we now connect the fixed and the moving conductor in series with ammeter A, the rheostat and the battery, as in Fig. 30.3. We measure the force produced when one unit of current flows (indicated by getting the pointer of A to the first mark), and then measure the force having adjusted the rheostat so that two units of current are flowing. We find now that the force is four times what it was. This suggests that the force has doubled twice, firstly as a result of doubling the current in the fixed conductor, secondly as a result of doubling the current in the moving conductor. Arguing by the logical process of induction we can suppose that if we had x units of current in one conductor and y units in the other, then we would get xy units of force.

It remains now to investigate the effects of varying firstly the spacing, and, secondly, the length of overlap, of the two conductors. The first is difficult in practice and needs a skilled hand. However, with care one can show that, keeping the current constant, the force between the conductors is inversely proportional to the distance between their centres. The length of overlap of the conductors is varied by substituting fixed conductors of different lengths. Again one makes sure that the current is kept constant throughout: then it is easily shown that the force is directly proportional to the length of overlap.

Summing up the results of the experiments we have outlined, we can say that the force F between two parallel straight conductors:

$$F \propto \frac{I_1 I_2 L}{a}, \tag{30.1}$$

where I_1 is the current in one, and I_2 the current in the other conductor, L is the length of overlap of the two conductors, and a is the distance between them. As yet we have no defined unit of current. The formula above helps us to define the unit: we define the unit of current in terms of the force. If L and a are each one metre, if the currents in the two conductors are equal, and if they are in a vacuum (free space), then the ampère is the value of current which produces a force of 2×10^{-7} newtons. Why, one may ask, choose a force of 2×10^{-7} N and not simply one newton, since one has chosen unit values for L and a? The reason is an historical one. The ampère was defined internationally until 1948 in terms of the mass

of silver deposited by an electric current in electrolysis of silver nitrate solution. For practical purposes the unit was one of convenient magnitude. When the MKS system of units (based on the metre, kilogramme, and second) was being adopted it was thought wise not to alter appreciably the magnitude of the fundamental unit of current. Hence the numerical factor 2×10^{-7} was introduced. Thus the formula for the force between two conductors which are parallel, straight, and in a vacuum, can be written

$$F = 2 \times 10^{-7} \frac{I_1 I_2 L}{a} \qquad (30.2)$$

provided that the appropriate units are being used for F, I, L, and a.

In connection with this formula there are two important points.

a) End-effects

Around an infinitely long straight wire carrying a current, and isolated in space, the magnetic field is uniform. If the wire is finite in length, bending around at its ends as it must if it is to be part of a circuit so that a current can flow in it, then the magnetic field will be non-uniform. We assume when we use the formula above that such non-uniformities in the magnetic field due to the finite length of the wire are negligible. This is justifiable if the lengths of the straight portions of the wires are large compared with the spacing of the wires apart from each other.

b) Finite thickness of the conductor

The formula is strictly applicable only if the area of cross-section of the conductor is negligible, and thus all the moving electric charges in the conductors can be considered as being on paths which are precisely the same distance apart. In practice, errors introduced by the fact that the wires have finite thickness are negligible provided that they are separated from each other by several diameters.

The magnitude of the forces involved in the experiments with a simple parallel-wire current balance (described above) is small—much less than one newton —although the current used is in the order of an ampère or more. Typically in such a current balance we might have: $I_1 = I_2 = 2$ A, $L = 25$ cm, $a = 1.0$ cm, and hence

$$F = 2 \times 10^{-7} \times \frac{2 \times 2 \times 0.25}{10^{-2}}$$

$$= 2 \times 10^{-5} \text{ N},$$

or about two milligrams weight.

Not surprisingly, therefore, the force between the two current-carrying wires in the flex leading to, say, an electric heater is not likely to be large enough to make the wires fly apart. It is perhaps fortunate that, in familiar situations anyway, the force between adjacent wires carrying currents are very small. There are, however, some interesting rare situations where such forces are dangerously large. The Russian physicist Peter Kapitza, working in Cambridge in the nineteen-twenties, developed techniques for producing extremely large magnetic fields by sending huge, but very short-lived, pulses of current through large coils. One of the problems encountered was the fact that the huge forces between the wires of the coil themselves caused several coils to burst. But the currents there were in the order of millions of ampères. Even the very large currents flowing in the overhead supply cables of the National Grid in Britain produce forces which are far too small to be troublesome.

When considering the force between conductors carrying currents we can say that one of the conductors produces a magnetic field, and the other experiences a force as a result of being in this magnetic field. We will turn our attention for a while to situations in which the magnetic field is produced by a magnet, or combination of magnets, and discuss the force on conductors in the magnetic field.

To become familiar with the force in this situation we can usefully do some simple experiments, and it is convenient to use again the ammeter, A, of the previous experiments, having made the two calibration-marks on its otherwise blank scale. The force in this new experiment is going to be considerably larger than in the parallel-wire current balance experiments, and we can use a crude but quick way of measuring

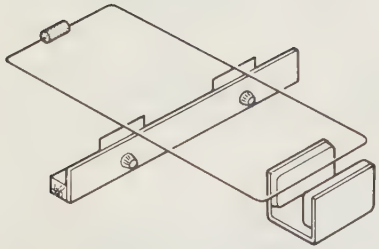


Fig. 30.4 Simple current balance for investigating the force on a conductor in a magnetic field.

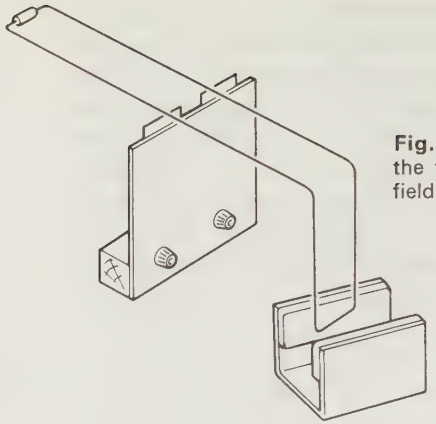


Fig. 30.5 Dependence of the force on angle with the field.

the force: simply a length of wire (or even paper tape) cut off, bent, and used as a rider to counterweight the balance-arm of the simple current balance (shown in Fig. 30.4). The magnetic field is provided by ferrite slab magnets in soft iron "yokes." Although we cannot alter the strength of the magnetic field conveniently in these experiments, we can vary the extent of the region of field affecting the wire, simply by placing a second, and perhaps a third, yoke of magnets next to and in contact with the first one. What we can find from this experiment is as follows.

a) In a magnetic field of constant strength, the force on the wire is directly proportional to the strength of the current.
b) Keeping the current constant, the force on the wire is directly proportional to the length of wire affected by the field.

In symbols:

$$F \propto IL.$$

Obviously if we had used stronger magnets we would have obtained greater forces. Thus, for a given value of I and a given value of L, the force on a wire lying at right-angles to the direction of the magnetic field can be used as a measure of the strength of the field. Note that we have said *wire lying at right-angles*. As will be shown, the angle between the wire and the direction of the magnetic field affects the size of the force. A convenient way of measuring the strength of a magnetic field, which we suppose for the time being to be a *uniform* field, is by defining it as the force per unit length per unit current upon a straight wire lying at right-angles to the field. If the actual current in the wire is I and the length affected by the field is L, then the strength of the field is given by

$$B = \frac{F}{IL}. \qquad (30.3)$$

The unit of this quantity in the Système International is the newton per ampère metre, or *tesla*. The name commemorates the remarkable work of Nikola Tesla (1856–1943), born in Yugoslavia, who founded the Tesla Laboratory in New York. He developed and patented various types of a.c. motor, new forms of generators and transformers. He also invented a resonant high-voltage induction coil which has come to be known as the Tesla coil.

What happens when a wire carrying a current lies in a magnetic field but not at right-angles to it? This can be investigated using a modified form of current-balance. Figure 30.5 shows this. The "business end" of the balance armature is short: this is so that it is affected only by the limited, nearly uniform, region of field between the magnet poles. The vertical parts of the armature in the field also experience forces due to the field, but these forces are very nearly horizontal, and therefore tend only to rotate the armature in a horizontal plane, and it is constrained

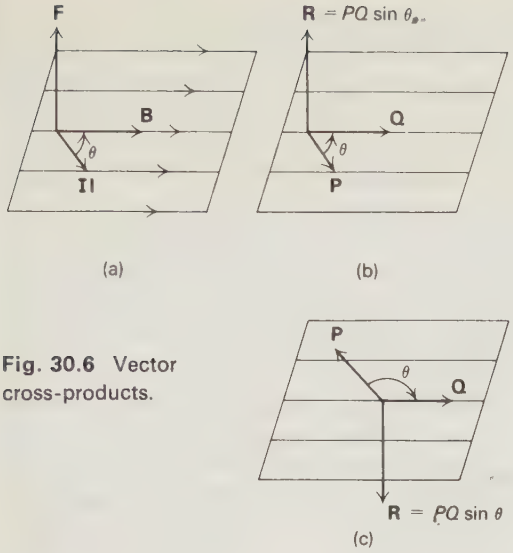


Fig. 30.6 Vector cross-products.

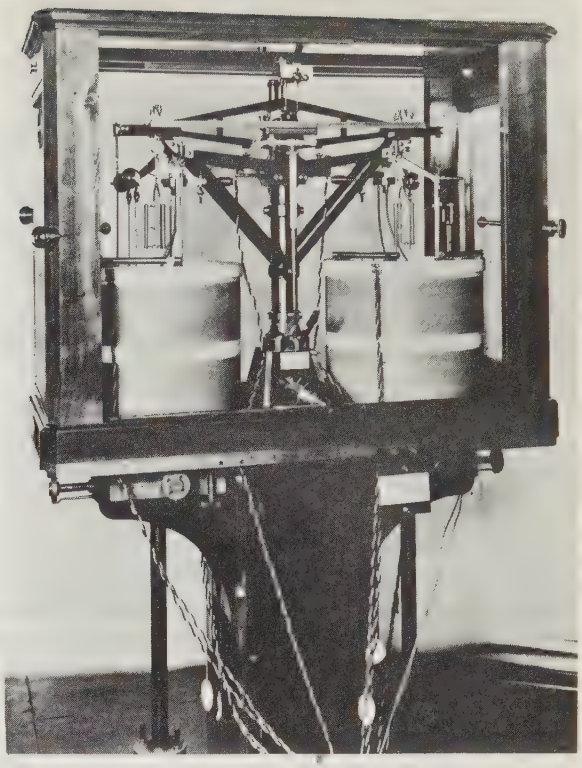


Fig. 30.7 The National Physical Laboratory current balance which is used from time to time to realize the ampère. (Crown copyright. Reproduced by permission of the National Physical Laboratory.)

not to do this. Thus the only measured force is that on the "business end" of the armature, and this piece of wire can easily be set at different angles to the field. It is found that the force is greatest (keeping constant current) when the wire makes a right-angle with the field, and is zero when the wire is parallel to the field. Careful measurement will reveal that the force is proportional to the sine of the angle, θ, between wire and field. Thus the complete expression for the force on a straight conductor of length L in a field of strength B is written

$$F = BIL \sin \theta. \qquad (30.4)$$

The direction of this force is always at right angles to the plane containing the direction of I and the direction of B.

(*Armature*, as used above, is a term generally signifying the moving parts of a piece of electromagnetic machinery. In a motor or dynamo it is the rotating part: in a magnetic relay it is the pivoted piece of magnetic material which moves a small distance to open or close the relay's switch contacts.)

The magnetic field intensity, B, or magnetic flux density as it is more correctly called (see Unit 7), can only be specified completely by stating both a magnitude and a direction. It must, therefore, be classified with the vector quantities.

In Eq. (30.4), which summarizes the relationship between the field $\boldsymbol{B}$, the product IL and the force on the wire, the directions of the three quantities may be represented as in Fig. 30.6(a).

This situation is typical of a *vector cross-product*. The direction of such a product is determined by a right-hand screw rule. If we use the symbols $\boldsymbol{P}$, $\boldsymbol{Q}$, and $\boldsymbol{R}$ (Fig. 30.6b) to represent three vectors in the three dimensions, we may say that if vector $\boldsymbol{P}$ is

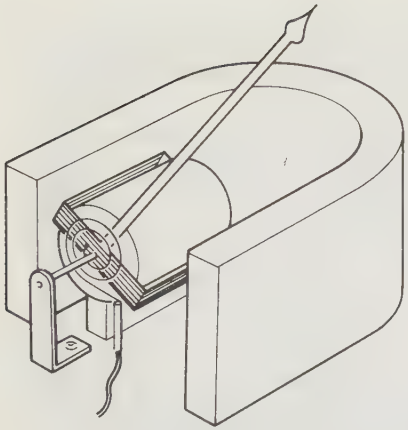


Fig. 30.8 Working model of a moving coil galvanometer.

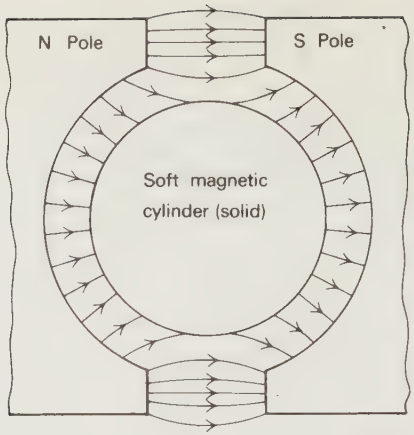


Fig. 30.9 The direction of the magnetic field is radial except near the gaps between the poles.

rotated into the vector $\boldsymbol{Q}$ through the angle between the two vectors θ (which is less than 180°), a right-handed screw rotated in the same sense would advance in the direction of $\boldsymbol{R}$. The magnitude of the cross product is $PQ \sin \theta$.

If the angle between $\boldsymbol{P}$ and $\boldsymbol{Q}$ is greater than 180° (i.e. is obtuse) one still performs the rotation of $\boldsymbol{P}$ into $\boldsymbol{Q}$ through the acute angle $(180° - \theta)$, and so the direction of the cross product is downwards with respect to the plane containing $\boldsymbol{P}$ and $\boldsymbol{Q}$ in Fig. 30.6(c).

Reflect for a moment upon the problem of measuring an electric current. An ammeter is simple to use, but must first be calibrated. A current balance, of the parallel-wire type described above, can be used to make an absolute measurement of a current (provided it is in the order of an ampère; for if it were much smaller the force would be impossible to measure accurately), but the process is tedious and not capable of high accuracy anyway. An accurate current balance is shown in Fig. 30.7. The current to be measured passes through coils and the force between the coils is weighed, and then the current can be calculated using a formula appropriate to the particular design of balance.

A current-measuring meter of the moving-coil type makes use of the force on wires carrying current in the field of a strong permanent magnet. The action of this type of meter is best understood by first referring to a simplified model which demonstrates its action (Fig. 30.8). The coil is caused to rotate by the action of two forces, one on each side of the coil, and acting in opposite directions (a pair of equal forces, referred to as a *couple*). When a current is switched on, the coil rotates and comes to rest when the opposing turning effect of spiral springs (or the torsion of the suspension wire, in the case of very sensitive galvanometers) is equal to the turning effect of the electromagnetic forces on the coil. If such an instrument is to have a linear scale then the field in which the sides of the coil move must be at right-angles to the sides of the coil, regardless of the position of the coil. (When we say "linear" we do not mean that the scale is straight: we mean that the size of the interval between two marks on the scale representing a small increase in current should be the same over the whole extent of the scale. This requires that the angular deflection of the coil should be directly proportional to the current in the coil.) The magnetic field must therefore be radial over the whole region in which the sides of the coil are allowed to move. Figure 30.9 shows the arrangement of magnet poles, and a solid cylinder

made of a soft magnetic alloy, which produces such a field. The coil is not generally allowed to swing through a total angle of more than 60° at most. In this way the fact that the field becomes appreciably non-radial near the gaps between the poles is no worry.

In a typical school laboratory milliammeter the magnetic flux density in the region of the coil is likely to be in the order of 10^{-1} tesla. The sides of the coil, assumed roughly square, might be about 0.5 cm, and the number of turns could be, say, 50. From this data we can estimate the forces on the sides of the coil when the current which produces full-scale deflection is flowing. Let us suppose that this meter reads up to 10 mA.

Considering just one side of the coil, the total length of wire affected by the radial field is $50 \times 0.5 = 25$ cm. With a current of 10 mA flowing the force is given by

$$F = BIL$$

hence

$$F = 10^{-1} \times 10^{-2} \times 0.25$$
$$= 2.5 \times 10^{-4} \text{ N}.$$

This is about 25 milligrams weight: a small force, which can be balanced by the action of a pair of fine hair-springs. Naturally, such a meter is fairly delicate. A pair of hairsprings is used because these springs serve conveniently as a means of making electrical connections to the coil.

The moving coil of the instrument is an oscillatory system. If there is no damping apart from the friction in the bearings, and air-drag, then, when a current is switched on, the coil will perform many cycles of oscillation before coming to rest and permitting a reading to be taken. Extra damping is needed so that the needle of the instrument comes to rest in a short time. Electromagnetic damping is used, and it is provided by the metal former upon which the coil is wound. When the coil is actually moving a current is induced in the former itself, and this current is in such a direction as to set up forces which oppose the movement of the coil (Lenz's law). These resisting forces are velocity-dependent, like viscous forces. Ideally the design of the coil should provide critical damping, that is, just enough damping to bring the coil to rest without it overshooting the mark. You may remember that the different degrees of damping in an oscillatory system were demonstrated by means of a galvanometer in Unit 5, except that in those demonstrations the damping was controllable externally.

A very sensitive galvanometer, for example the type which does not have a pointer but in which a beam of light is reflected on to a translucent scale, often has a switch (or other means) for short-circuiting the terminals when the meter is not in use. This actually protects the coil and its delicate suspension from damage. If the coil receives a sudden impulse because the instrument has been roughly handled, for instance, a current is induced in the coil as it moves, passing through the short-circuit connection between the terminals, and damps the motion of the coil considerably more heavily than the damping due to the metal frame of the coil. Some sensitive galvanometers, especially older models, have a mechanical clamp to protect the coil when the meter is being shifted from place to place. Whatever the protective technique employed to prevent the coil moving, it should always be used before moving the instrument.

A more detailed discussion of the design principles of a moving-coil galvanometer can be found, in the form of Problem (30.1), at the end of this chapter.

30.2 ELECTRIC MOTORS

If you have never built a simple model electric motor for yourself you should, if possible, seize the opportunity now, for there is no better way of beginning to understand the action of motors than feeling for yourself the forces which are responsible for a motor's action. Figure 30.10 is a drawing of a model motor which can be quickly assembled from a kit of parts. Figure 30.11 is a simplified, schematic, diagram of the motor, to assist an explanation of its action. A rectangular coil of wire, *QRSTUV*, is free to rotate about an axis 00′ between the poles of magnets *N*

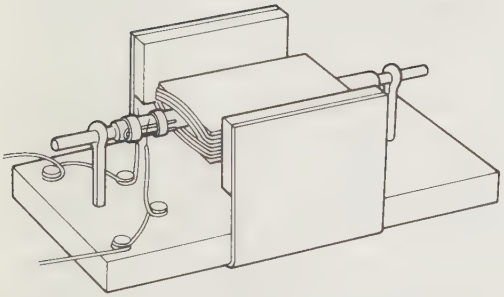


Fig. 30.10 A simple d.c. electric motor. (From Osborne, J. M., *Electromagnetism*, Longmans, 1970. Reproduced by permission of the publisher.)

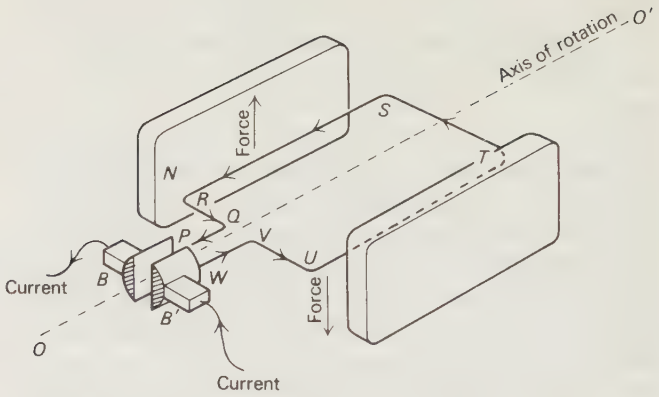


Fig. 30.11 Schematic diagram of a simple d.c. motor.

and *S*. The ends of the coil are connected at *P* and *W* to a pair of almost semicircular pieces of copper foil. Two brushes, *B* and *B'*, make electrical contact with these pieces of foil without appreciably impeding the rotation. In the model (Fig. 30.10) the magnetic field is by no means uniform, but for the purpose of our description here we shall assume that the field in Fig. 30.11 is uniform. When the coil is in the position shown in Fig. 30.11 the force on the side *RS* is directed vertically upwards, and the force on the side *TU*, vertically downwards. Since the parts of the coil *ST*, *QR*, *UV* are all parallel to the field, there is no force on them. The pair of equal forces on the sides *RS* and *TU* constitute a couple which rotates the coil clockwise. Consider now the situation when the coil has rotated through 45° from the position shown. The forces on each side of the coil will be the same as before, but their turning effect is less because the lines of action of these two forces are closer to the axis than before. When the coil has rotated through another 45° the two forces simply pull against each other, because side *RS* is now vertically about side *TU*, and there is no turning effect. Assuming that the coil has acquired some angular momentum, we can see that it will overshoot this "dead" position. As it does so, the gaps between the semicircular segments pass under the brushes, and an instant later the electrical connections to the coil are reversed. The segments and brushes constitute the *commutator*. A commutator is any kind of switch which reverses a pair of electrical connections. Thus the direction of current through the coil is reversed, and hence the directions of the forces on the sides *RS* and *TU* are reversed. The force on side *RS* now pulls it downwards, and the force on *TU* pulls upwards. This change-over process is shown in Fig. 30.12. The coil then makes a half revolution, the current is reversed once again, and hence the forces on the sides *RS* and *TU* are again reversed. In this way, the forces acting upon the sides of the coil are always in such directions as to make the coil rotate continuously in the same sense (clockwise, in our example).

Although the forces on the coil make it turn continuously the same way, the torque produced by them is not constant. We need now to look closely at how the torque on a coil in (for simplicity's sake) a uniform magnetic field depends on the coil's orientation in relation to the field. Consider the single-turn,

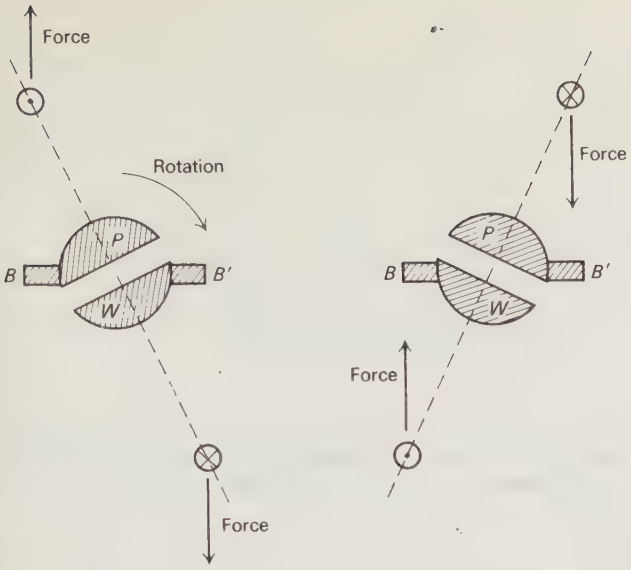


Fig. 30.12 A commutator (a) just before and (b) just after the reversal of the current.

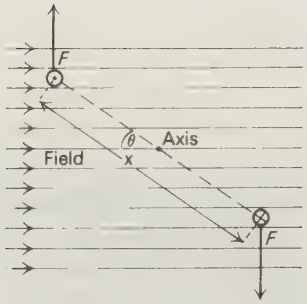


Fig. 30.13 The torque on a current carrying coil in a uniform magnetic field.

rectangular coil of Fig. 30.11. Suppose that the side ST has length x, and the side RS has length y. Imagine now that you are looking at it in a direction along the axis: along side RS a current is coming towards you, and along TU it is going away from you. Now transfer your attention to Fig. 30.13, which represents the two wires RS and TU sliced through and looked at end-on. This is a simpler representation than the perspective diagram of the other figure. The plane of the coil makes an angle θ with the magnetic field. The force on each side is given by $F = BIy$, where B is the flux density of the field and I is the current in the coil. The torque (T) on the coil is the combined moment of the two forces, and hence is given by

$$T = 2BIy \times \frac{x}{2} \cos \theta$$

$$= BIxy \cos \theta.$$

Since the area enclosed by the coil is given by $A = xy$, we could write the expression

$$T = BIA \cos \theta. \qquad (30.5)$$

Do forces on the other two sides on the coil have any effect? As stated earlier, when the sides RU (neglect the small gap $Q - V$) and ST are parallel to the field, as in Fig. 30.11, there is no force on them. When the plane of the coil is vertical (and you should argue this out for yourself), there are equal and opposite forces in RS and TU, but their lines of action lie along the axis, and thus exert no turning effect: they simply tug against each other. For any other position of the coil, these two forces just oppose each other and make no contribution to the turning effect.

The dependence of the torque upon the position of the coil in the field can be convincingly felt, using the model motor of Fig. 30.10, if one prevents the coil from rotating by means of one's fingers, and feels how strong the torque is when the coil is held at different positions. Expression (30.5) derived above shows that, in a uniform field, the torque is a sinusoidal function of the angular position of the coil, and thus reaches a maximum value twice per revolution, and zero twice per revolution. This unevenness of torque is undesirable. It can be overcome by two means: firstly, by having, effectively, a number of separate coils, mounted on the rotating part of the motor (called the *armature* or *rotor*), the planes of these

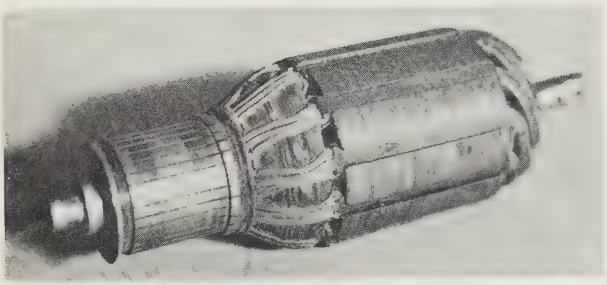


Fig. 30.14 The armature and commutator of a small d.c. motor.

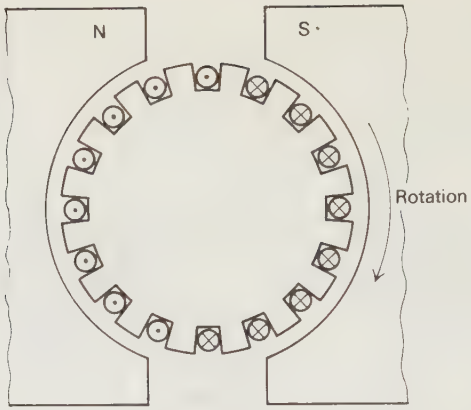


Fig. 30.15 Schematic diagram of the cross-section through the windings of a d.c. motor, showing the directions of the currents.

coils being set at different angles, and secondly, by arranging that the magnetic field is more nearly radial (as in the moving-coil galvanometer). Each coil is connected to a pair of segments in the commutator (these can be seen in Fig. 30.14) in such a way that, whatever the position of the armature as it rotates, the currents in the coils are as shown in Fig. 30.15. Thus, in such a motor, the torque can be made to be nearly constant as the armature rotates.

Before discussing the practical forms of motors further, let us consider how a motor converts electrical energy into mechanical forms. A simple experiment will open the discussion. Using, for convenience, a low-voltage d.c. motor, measure the current taken by the motor under different conditions. You will find two important facts.

a) When the motor is first switched on, the current is large: then, as the motor speeds up, the current decreases until the motor is running at a steady speed.
b) If the motor is forcibly slowed down (by gripping the shaft with a cloth, as a kind of crude friction-brake), the current is observed to rise, and it falls again when the motor is allowed to run freely.

A further experiment gives clues which help us to explain the facts just demonstrated. In the arrangement shown in Fig. 30.16 a motor has a flywheel attached to it so that, when the current is switched off, the armature keeps on spinning for a time. The switch S enables the motor to be disconnected from the battery and immediately connected to the lamp. The ammeter A must be a centre-zero type, because it is used here to indicate currents in both directions. Firstly, the switch is set to the left-hand position: the battery is thus connected to the motor which speeds up, and the current is shown on A. When the motor has reached its steady maximum speed the switch is flicked over to the right. Now acting as a dynamo while the flywheel is still spinning, the motor generates a current which lights the lamp. The important thing is that now the current registered by the ammeter is in the opposite direction to what it was before. Now, presumably, when the motor was connected to the battery it was nevertheless acting as a dynamo, producing a current in the opposite direction to the current forced through it by the battery. In more precise terms, the armature spinning in the magnetic field was generating an e.m.f. in opposition to the e.m.f. of the battery. This opposing e.m.f. is generally called a back e.m.f. The faster the armature spins, the greater is this back e.m.f., and hence the smaller

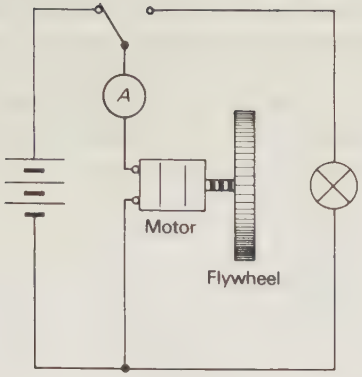


Fig. 30.16

the current flowing around the battery–motor circuit. This may become clearer when we construct a quantitative relationship, as follows. The net e.m.f. in the battery–motor circuit is the difference between the battery e.m.f. and the motor's back e.m.f. By the definition of resistance we can therefore write

$$\text{total circuit resistance} = \frac{(\text{battery e.m.f.}) - (\text{back e.m.f.})}{\text{current}}$$

or, in symbols,

$$R = \frac{E - \mathscr{E}}{I}$$

thus the current in the circuit is given by

$$I = \frac{E - \mathscr{E}}{R}. \qquad (30.6)$$

The back e.m.f., $\mathscr{E}$, can be assumed to be proportional to the speed of the armature—just as the e.m.f. generated by a dynamo is proportional to the speed. Thus, the faster the armature spins, the greater the value of $\mathscr{E}$, the smaller the value of $(E - \mathscr{E})$ and hence the smaller the current.

Turning now to the question of energy converted by the battery–motor circuit, we can write, for the rate at which energy is converted (from chemical to electrical form) in the battery, EI. Now, expression (30.6) can be rewritten

$$E = \mathscr{E} + IR$$

and multiplying this by I we get

$$EI = \mathscr{E}I + I^2R. \qquad (30.7)$$

The second term on the right-hand side is easily recognised as the rate at which electrical energy is converted into heat in the resistive parts of the circuit. The term $\mathscr{E}I$ represents the rate at which electrical energy is converted into mechanical form in the motor. Thus Eq. (30.7) is an expression of the principle of conservation of energy applied to the battery–motor circuit.

The existence of this back e.m.f., when a motor is running, can be made completely convincing by slightly modifying the experiment of Fig. 30.16. Instead of an ammeter connected in series with the motor, connect a centre-zero voltmeter across the motor. If such an instrument is not available one can modify a centre-zero galvanometer by adding a suitable series resistance, and perhaps making temporary new markings on the scale of the instrument. It will be found that, if the switch S is flicked across rapidly from battery to lamp position, a reverse voltage is registered by the voltmeter. In an efficient motor, running at working speed, the back e.m.f. is usually only slightly less in magnitude than the battery (or other supply) voltage. Typically, a motor designed to work from a 12 V supply may generate a back e.m.f. of 11 V or even greater. Since the resistance of the motor windings (i.e., the coils) is very small, even a net e.m.f. of (12 − 11) = 1 V can cause a large current to flow.

Little has been said so far about the magnetic field in which a motor's armature spins. In most motors this field is provided by an electromagnet, rather than a permanent magnet. In a *series-wound* motor the field winding is connected in series with the armature winding. In a *shunt-wound* motor the two windings are connected in parallel. For reasons which there is not room enough to discuss here, a series-

wound motor has the advantage that it can provide a larger starting torque than a shunt-wound motor, whereas a shunt-wound motor is preferable in applications where precise control of speed is desirable, because this type of motor can provide the extra torque needed to cope with an increased load without appreciable change of speed.

Problem 30.1 A galvanometer is said to be *sensitive* if it shows a large deflection for a small current. The sensitivity can be defined as the angular deflection of the coil per unit current. A design formula can be derived for the moving-coil galvanometer as follows. Since the field provided by the permanent magnet is radial in the region where the coil is permitted to move, the plane of the coil is always parallel to the field, and so $\theta = 0$ in the expression (30.5). Hence the torque on the coil is given by

$$T = BInA$$

if the coil has n turns.

The coil rotates against the opposing torque provided by hair-springs (or in a very sensitive instrument, by the torsional support). The torque provided by these springs can be written $k\phi$ where k is the torsional stiffness factor for the springs, and ϕ is the angular deflection of the coil. Hence, when current I is flowing in the coil and it is in equilibrium,

$$k\phi = BInA$$

and so the sensitivity (deflection per unit current) is given by

$$\frac{\phi}{I} = \frac{BnA}{k}.$$

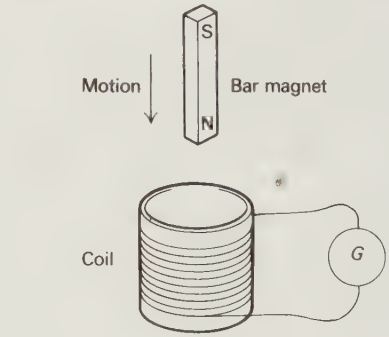
The sensitivity is large if B, n, or A are made large, and if k is made small. But there is a limit to how large any one of the parameters B, n, and A can be made, and to how small k can be made. And if B is to be large, this sets a limit on how large n and A can be. Discuss the problem of trying to maximize the expression BnA/k.

Chapter 31
ELECTROMAGNETIC INDUCTION

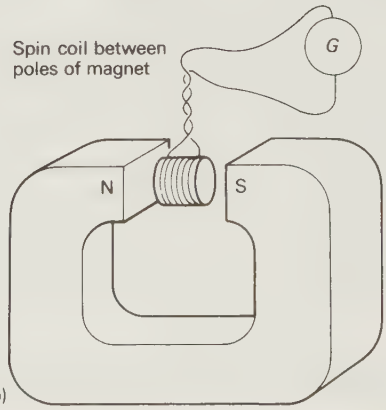
31.1 Electromagnetic induction
31.2 The disc dynamo
31.3 Induction without movement

31.1 ELECTROMAGNETIC INDUCTION

We know that a current-carrying conductor in a magnetic field generally experiences a force, and Chapter 30 was concerned with the practical applications of this force. When a current flows through a wire a magnetic field is set up around the wire, and the interaction of this field with some other magnetic field usually gives rise to a force on the wire. The converse effect, in which a magnetic field causes a current to flow in a wire, is known as *electromagnetic induction*, and was first fully investigated by Michael Faraday. If you have not done so before you should try the simple demonstrations of electromagnetic induction illustrated in Fig. 31.1. In each experiment it is found that the galvanometer registers a small current when the coil and the magnet are in relative motion, and that this current ceases as soon as the



(a)



(b)

Fig. 31.1

relative motion stops. The direction of the current is seen to be related to the direction of the relative motion of coil and magnet, and the strength of the current is related to the speed of the relative motion: the greater the speed, the greater the current. Also, if you increase the number of turns of wire in the coil in each case, you will find that this increases the current generated.

The simple demonstrations just described do not enable us to investigate precisely what factors determine the strength of the induced current, and a carefully designed experiment is necessary. Figure 31.2 shows an arrangement for moving a wire (or rather, a bundle of wires) at a steady, measurable speed between the poles of an electromagnet. It is found that a current is generated while the wires are moving, and only while they are moving. The parameters which can be varied in this experiment are:

a) the strength of the field,
b) the speed of the wire,
c) the number of turns in the coil,
d) the resistance of the circuit.

We can alter each parameter singly, and we find the following.

a) If the strength of the field is increased, by stepping up the current in the electromagnet, the induced current is greater, with the coil moving at the same speed. If one uses a suitably-shaped current balance to measure the flux density of the magnetic field (see Section 30.1), then one can show that the strength of the induced current is directly proportional to the flux density of the magnetic field.

b) Increasing the speed of the wire increases the induced current, and it is easily shown that the current is directly proportional to the speed.

c) If different numbers of turns in the coil are connected into the circuit with the galvanometer, then it is found that, for motion with the same constant speed, the current is directly proportional to the number of turns. Hence, presumably, the total length of wire interacting with the field is the important factor.

d) Varying the total circuit resistance, while keeping parameters (a), (b), and (c) constant, shows that the greater the resistance the smaller the current. One deduces from this that the parameters (a), (b), and (c) determine not the current but the e.m.f. being generated in the circuit. Thus, by means of this experiment, we find that the induced e.m.f. $\mathscr{E}$ must be given by this expression:

$$\mathscr{E} \propto BLv,$$

where B is the flux density of the magnetic field, L is the total length of wire interacting with the field, and v is the speed. The theory of what happens when an electrical conductor moves across a magnetic field in this way is developed in Unit 7 which shows that, if SI units are used for all the quantities, the expression can be written

$$\mathscr{E} = BLv, \qquad (31.1)$$

the constant of proportionality being unity.

In the situations discussed so far, the motion of the conductor was entirely in a plane perpendicular to the magnetic flux. Put in a slightly different way, we can say that, in those situations, the conductor was cutting the lines of magnetic flux at right angles. Imagine we have a straight conductor, and we are looking along it, and it is moving at right-angles to the lines of magnetic flux, as in Fig. 31.3(a). Here we represent the uniform magnetic field by a set of equally spaced parallel lines. If it goes from P to Q

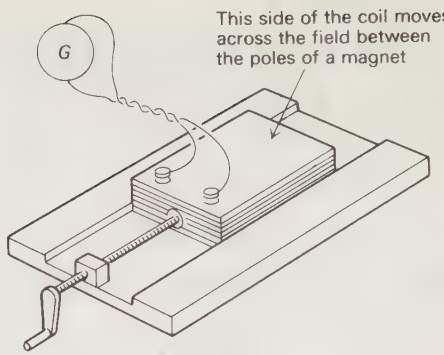


Fig. 31.2 Device for moving a wire through a magnetic field at a steady speed.

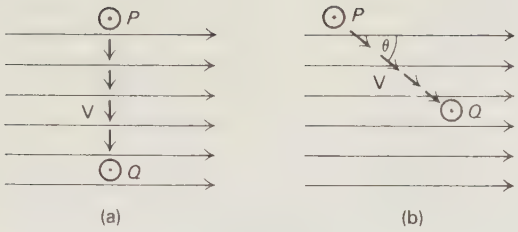


Fig. 31.3

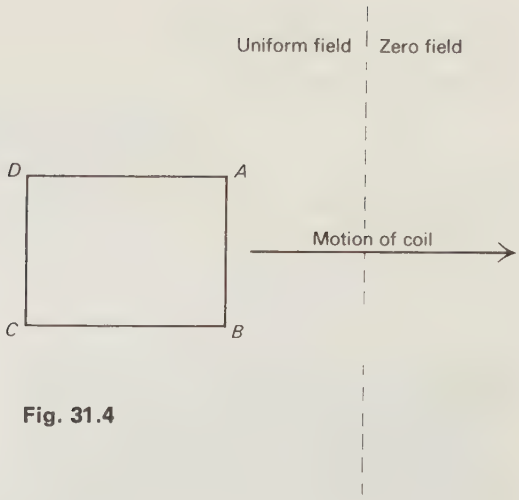


Fig. 31.4

it cuts 5 lines. In Fig. 31.3(b) it goes the same distance PQ, but only cuts 3 lines. If in both cases the conductor is moving with the same speed, it is found by experiment that the e.m.f. induced in Fig. 31.3(b) is less than in (a). If the motion of the conductor is parallel to the flux, then the induced e.m.f. is zero: no lines of flux are cut. Michael Faraday developed the concept of lines of flux (although he called them *magnetic curves*, and later, *lines of magnetic force* and in particular, the fact that an e.m.f. is induced in a moving conductor only if it cuts lines of flux. Figure 31.3(a) and (b) shows that the number of lines cut by the conductor per unit time must be proportional to the component of velocity of the conductor perpendicular to the magnetic flux, that is, to $v \sin \theta$, where v is the velocity of the conductor and θ is the angle between its direction of motion and the direction of flux.

To test whether you have understood these concepts, study Problem 31.1.

Problem 31.1 A *search coil* is a small coil which is used to measure the flux density in a magnetic field. It is placed in the field with its plane perpendicular to the field and then quickly removed to a region of zero field strength. As this is done a small charge circulates around the coil. Why is this? The following argument will show you how this comes about.

Figure 31.4 shows a plane, single-turn, rectangular coil in a uniform magnetic field whose direction is perpendicular to the plane of the diagram and directed away from us. The coil starts fully immersed in the field. It is then moved from left to right out of the field in the direction shown by the arrow. The field-region ends at the broken line. Consider the process stage by stage and fill in the information in a table as below. We will assume that the coil is moving at a steady speed from left to right.

	Direction of force (if any) on electrons in side:				Net current in coil
Coil totally in field	*AB*	*BC*	*CD*	*DA*	
Side *AB* to right of broken line					
Coil totally out of field					

Now we are equipped with the concepts necessary to explain the action of a simple dynamo. A schematic

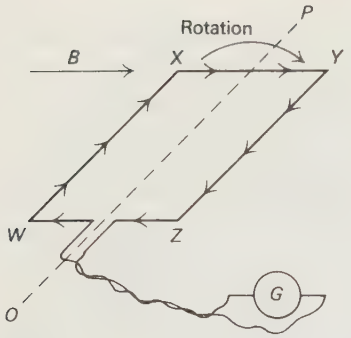
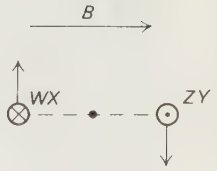


Fig. 31.5



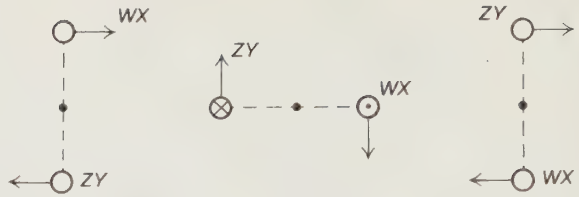


Fig. 31.6

diagram of the very simplest dynamo is shown in Fig. 31.5. A plane rectangular coil rotates about an axis OP which is set at right-angles to a magnetic field, assumed uniform, of flux density B. In the position shown in Fig. 31.5 the sides WX and ZY are moving at right-angles to the flux, in opposite directions. Thus the e.m.f.'s induced in these sides reinforce each other, tending to push electric charge one way around the circuit. The sides XY and WZ, however, are not cutting lines of flux: nor will they do so at any stage during rotation of the coil. Figure 31.6 shows a section through the coil, viewed in a direction along its axis of rotation, at various stages during rotation of the coil.

During one half-revolution the current will flow one way around the circuit, during the next half-revolution it will flow the opposite way. Consider the coil when its plane makes an angle α with the magnetic field B (Fig. 31.7). Let v be the velocity of the sides WX and ZY, each of length y, then, as shown earlier, the e.m.f. induced in each side must be

$$Byv \sin (90 - \alpha) = Byv \cos \alpha.$$

If the angular velocity of the coil is ω, then $\alpha = \omega t$ and the linear velocity of the sides WX and ZY is given by

$$v = \omega \frac{x}{2},$$

where $XY = x$. Hence the net e.m.f., due to the two sides WX and ZY, can be written:

$$\mathscr{E} = 2 \times By\omega \frac{x}{2} \cos \omega t$$

or

$$\mathscr{E} = Bxy\omega \cos \omega t.$$

The area enclosed by the coil is xy, and we can call this A. Hence

$$\mathscr{E} = BA\omega \cos \omega t.$$

Writing $\omega = 2\pi f$, where f is the frequency, we get, for a coil of N turns,

$$\mathscr{E} = 2\pi NBAf \cos 2\pi ft. \qquad (31.2)$$

This is a sinusoidally varying quantity whose amplitude is given by $\mathscr{E}_0 = 2\pi NBAf$. This expression displays clearly the factors determining the amplitude or peak value of the e.m.f.—namely the number of turns in the coil, the flux density of the field, the area enclosed by the coil, and the speed of rotation. In designing a dynamo an engineer attempts to make all these parameters as large as possible, within reason; also he will attempt to have as low a coil resistance, R, as possible because this determines the internal resistance of the dynamo. It is obvious that, in practice, one cannot have the parameters B, N, A, large and R small at the same time. If N is increased, the coil must necessarily take up more space, and increase the size of the armature: there is bound to be

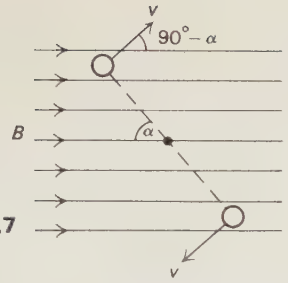


Fig. 31.7

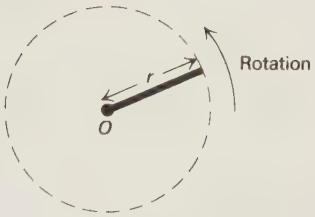


Fig. 31.8 A rod, length R, rotating about an axis at one end. A uniform magnetic field is directed perpendicularly into the plane of the paper.



Fig. 31.9 Faraday's ring. (Reproduced by permission of the Royal Institution of Great Britain.)

a practical upper limit set upon this size. Increasing A also means an increase in size, and increasing B means using larger electromagnets for the field which also necessitates an increase in size. Decreasing R means using thicker wire for the coils and/or a shorter total length of wire: this requirement conflicts with the need to maximize N and A. Thus, a design compromise has to be achieved. As in most electric motors, the magnetic field in which the armature rotates is nearly always provided by an electromagnet, rather than by permanent magnets.

In a typical a.c. generator the coils which constitute the two-pole field magnet rotate at 3000 revolutions per minute (50 Hz) as the rotor whilst the output voltage is developed in stationary coils (stator). This arrangement ensures that the output current does not have to flow through any system of slip-rings and brushes. The rotor is usually driven by a high speed turbine.

31.2 THE DISC DYNAMO

When a conducting disc is spun in a magnetic field, with its plane perpendicular to the flux, a p.d. appears between the axis of rotation and the rim. This is easily understood if we first find out what happens if a straight conducting rod is rotated about one end in a field. Consider the rod in Fig. 31.8. This is moving in the direction shown by the arrow, and it sweeps out, in each revolution, an area equal to the area of a circle radius r. Whatever the position of the rod during rotation, a steady p.d. is set up between the ends of the rod.

Now suppose we had a number of spokes, each of length r, and all in the same plane, mounted on the same axle. As these rotate the same p.d. will be set up in each of them, and their outer ends will all be at the same potential relative to the axis. The thickness of the spokes does not affect the result, and so if we imagine a large enough number of spokes, all thick enough so that we can see no daylight between them, then we have a solid conducting disc and there must be a p.d. between its axis and its rim.

31.3 INDUCTION WITHOUT MOVEMENT

So far we have seen only those examples in which an e.m.f. is induced by relative movement of a conductor

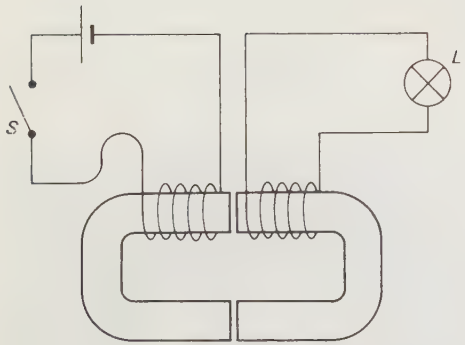


Fig. 31.10

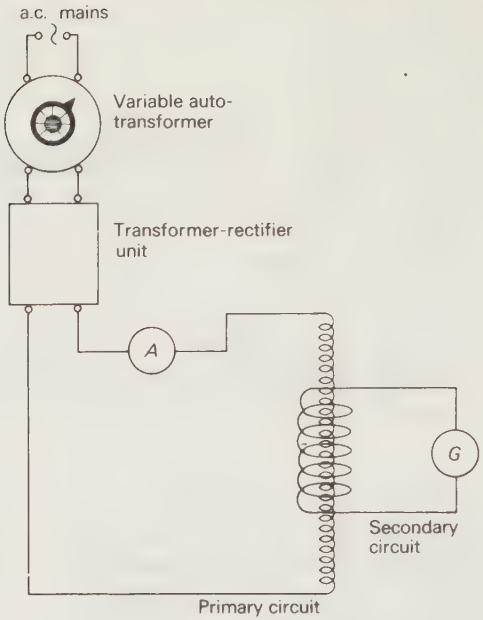


Fig. 31.11

and magnetic flux. Faraday demonstrated that electromagnetic induction does not necessarily involve movement: it can be brought about by the change in strength of a magnetic field.

Faraday's famous experiment used an iron ring (Fig. 31.9) with two separate coils of wire wound upon it (the wire was of copper and the insulation, of calico, laboriously wrapped around the wire by hand). To demonstrate the same phenomenon today we use basically the same arrangement, except that it is more convenient to have coils which can be quickly slipped on and off a magnetic core, as in Fig. 31.10. Provided that one chooses a pair of coils having a suitable number of turns, one can show the electromagnetic induction impressively by having a low-power, low-voltage lamp connected to one of the coils, and a battery and switch connected in series with the other coil. When the switch is closed, the lamp flashes briefly, and again when the switch is opened. With the switch closed, and thus a steady current flowing in the one coil, there is found to be no current in the other coil. Thus, an e.m.f. is generated only while the magnetic flux set up by the one coil is changing. One can refine the experiment of Fig. 31.10 a little by adding a variable resistor in series with the battery and one coil to vary the current, and hence also the magnetic flux, less suddenly than is done by a switch. With a suitable ammeter in series with the lamp it will then clearly be seen that there is current in the lamp-circuit so long as the current in the primary coil is actually changing. It will also be seen that when the current in the primary coil is increasing, the current in the secondary coil flows in one direction; and with primary current decreasing, the secondary current flows in the opposite direction.

What would one have to do in order to keep a continuous, steady current flowing in the secondary coil? Presumably it would be necessary to make the current in the primary increase continually; or decrease continually. This, obviously, is not physically possible for any great length of time. But if the current in the primary is made to increase, then decrease, alternately, a current will flow in the secondary, continually reversing direction: in other words, an alternating current will flow in the secondary.

Figure 31.11 shows a simple experiment for in-

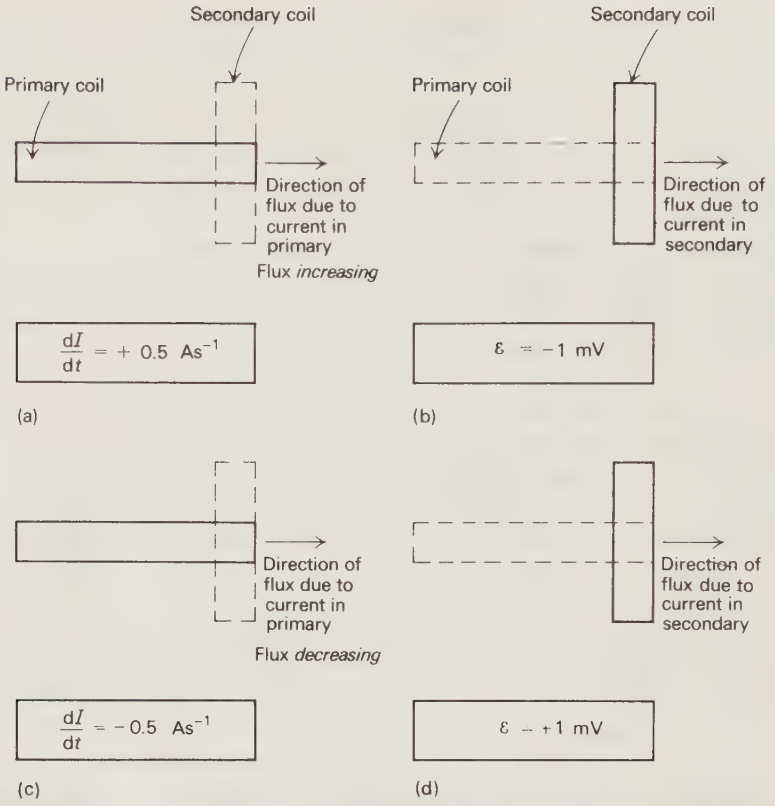


Fig. 31.12

vestigating how an e.m.f. induced in a coil by a changing magnetic flux depends upon the rate of change of flux.

The use of a "Variac" type variable auto-transformer and transformer-rectifier unit may seem a complicated way of getting a variable current for the primary coil, but it makes it much easier to increase the current in the primary at a steady rate than a simple battery and rheostat would. The primary coil is the long air-cored solenoid. The secondary is a short coil which can be slid loosely along the primary. It is found that, when the current in the primary is increased at a uniform rate (e.g. 0.5 ampere per second), the secondary current is constant. Choosing different rates of change of primary current demonstrates that the strength of the secondary current is directly proportional to the rate of change of primary current. But, as was found when investigating motional e.m.f., it is not the induced current which is the fundamental quantity, but the e.m.f. You could put this fact to the test, if desired, by altering the resistance of the secondary circuit, and using a constant rate of change of primary current (I_p). You would find that the secondary current is inversely proportional to the total circuit resistance, so that the product (current × resistance), which is equal to the e.m.f., is constant. Hence we can write, in symbols

$$\mathcal{E}_s \propto \frac{dI_p}{dt}.$$

We need now to make clear what is happening in terms of the magnetic flux in the coils. The direction of the flux is easily shown by means of a pivoted compass needle. We set a steady current in the primary, and show that the direction of flux in the solenoid is, for example, as shown in Fig. 31.12(a). When the current is increased, this flux becomes

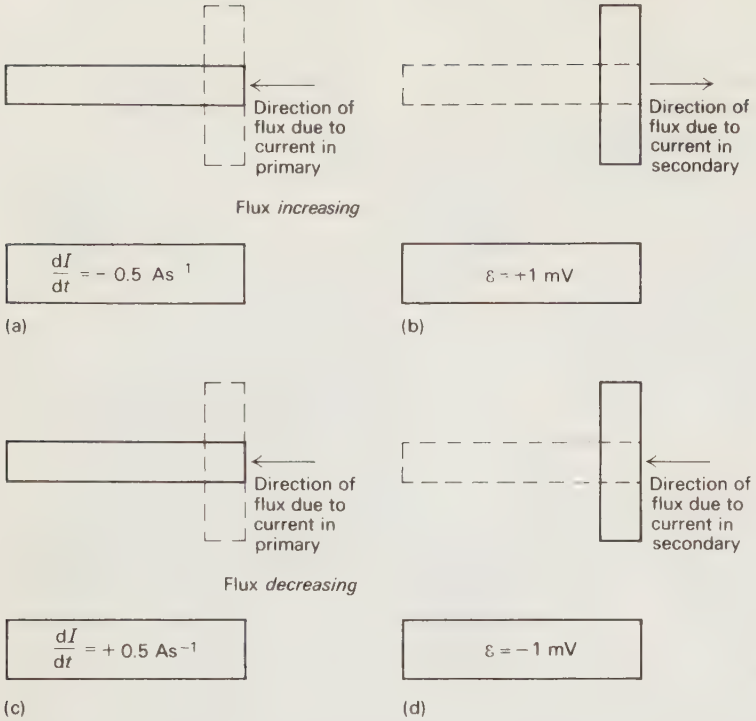


Fig. 31.13

stronger. But while it is increasing, a current is induced in the secondary coil. What about the magnetic flux set up by that induced current? We cannot investigate this directly, because its effects would be masked by the strong flux of the primary coil. But we can, by means of a battery and suitable resistances, cause a current to flow in the secondary which is in the same direction as the induced current was. When we do this (see Fig. 31.12b) with the primary switched off, we find that the flux due to the secondary current is in the opposite direction to the flux which we had previously in the primary. Now we know that when the primary current is decreasing (Fig. 31.12a), the secondary current changes direction, thus the flux due to the secondary current must be as shown in Fig. 31.12(d). So it appears that the flux due to the secondary current, by opposing the primary flux when it is increasing, and assisting the primary flux when it is decreasing, "strives to prevent" any changes in the primary flux. If the experiment is extended by interchanging the battery connections (Fig. 31.13) and thus reversing the direction of the primary current, the above suggestion will be found again applicable. These directional relationships are important, particularly in the theory of transformers, and they are most concisely expressed mathematically. If we use the symbols I_p and I_s for the primary and secondary current, we reckon these as having the same sign if they produce magnetic flux in the same direction. Thus, if $I_p = +3$ A it produces a flux towards the right as in Fig. 31.12(a). If it is increasing at a rate 0.5 amperes per second, we can write this

$$\frac{dI_p}{dt} = +0.5 \text{ A s}^{-1}.$$

If it is decreasing at the same rate, we write

$$\frac{dI_p}{dt} = -0.5 \text{ A s}^{-1}.$$

Also, if the primary flux is increasing in the reverse direction at 0.5 A s^{-1} we write $dI_p/dt = -0.5$ A s^{-1}, and if decreasing in the reverse direction, $dI_p/dt = +0.5$ A s^{-1}. It should be clear from all this that when

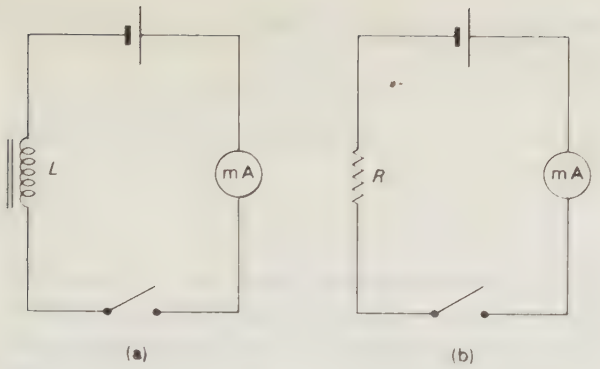


Fig. 31.14

we write the relation between I_p and I_s we must put it in the form

$$\mathscr{E}_s = -M \frac{dI_p}{dt}. \qquad (31.3)$$

With the numerical values we have used in the illustrations of Fig. 31.12 we find that the value of M must be given by

$$M = \frac{10^{-3}\ \text{V}}{0.5\ \text{A s}^{-1}} = 2 \times 10^{-3}\ \text{V A}^{-1}\ \text{s}.$$

The unit, VA^{-1}s, is usually called the henry (H). The value of M, a constant for the primary and secondary circuits, called the *mutual inductance*, depends on the number of turns on the two coils, their spatial dimensions and relative positions: also upon the magnetic properties of the medium inside and around the coils. A more thorough investigation of mutual induction will follow shortly, when we develop the theory of transformers. For the time being it is worth doing one further experiment with the apparatus of Fig. 31.11. If we keep a constant rate of change of primary current, but use different numbers of turns on the secondary, we find that the steady induced current, as shown by the galvanometer, is proportional to the number of turns used. It is important here, however, to ensure that the total resistance of the secondary circuit remains unchanged—or, at least, altered by only a negligible fractional amount. This can be ensured by arranging that the resistance of the galvanometer plus the extra resistance, R, in the circuit of Fig. 31.11, is very much greater than the total resistance of the tapped secondary coil. Thus we see that the induced secondary e.m.f. is proportional to the number of turns, N, on the secondary coil, and because all the magnetic flux produced by the primary solenoid passes through the secondary coil, or links the secondary circuit, we come naturally to the concept of *flux linkage* in a coil which is defined as the product of the flux passing through the coil and the number of turns in the coil.

So far we have seen how electromagnetic induction can be brought about by relative moving of a conductor and a magnetic field, and by changes in magnetic flux linking a circuit without any physical movement. It should not be surprising, therefore, that when a changing current flows in a conductor, the changing magnetic flux due to this current induces an e.m.f. in the self same conductor. This phenomenon is called *self induction*. It is most easily demonstrated with a very large inductor: a coil of many thousands of turns wound upon a laminated magnetic alloy core. Two circuits are set up, side by side, as in Fig. 31.14. The two circuits (a) and (b) are the same except that (a) has the inductor, L, and (b) has a resistor R (in the form of a resistance box) which has a resistance equal to that of the inductor. When the two circuits are switched on simultaneously the two meters register a current, but it is observed that the current in (a) rises much more slowly than the current in (b). There is, of course, a natural time-lag with a galvanometer, between switching on a current and the needle reaching a steady reading, but in spite of that, the inductor in (a) is clearly having a delaying effect upon the rise of current. What is more, when the two switches are opened a fat spark may often be observed between the contacts of the switch in (a) as they separate, but not in (b). (For this reason it is wise to use cheap switches, or better, old morse-keys where the contacts can be easily seen.)

Why does the presence of the inductor in the circuit of (a) delay the rise of current? If one has a suitable oscilloscope, with a long-persistence screen, one can display the time-trace of the current by having

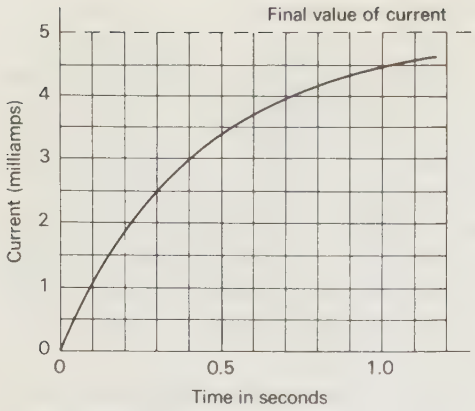


Fig. 31.15 The growth of current in an inductive circuit.

a small resistor (say, 1 Ω – 10 Ω) connected in series in circuit (a) and the input leads to the oscilloscope connected across it. The time-trace will be found to appear as in Fig. 31.15. The rate of rise of current becomes progressively less, and the current approaches a final steady value which is determined by the resistance of the circuit. The initial rate of rise can be shown to depend upon the size of the inductor used: the larger the inductor, the slower the initial rate of rise of the current. The behaviour just described can be explained if we suppose that an e.m.f. is being induced in the coil which opposes the rising current. If is reasonable to suppose that this e.m.f. is related to the rate of change of current in the same sort of way as it was in the mutual inductance experiment of Fig. 31.10. Thus we can write, for the p.d. across the inductor,

$$\mathscr{E} = -L\frac{dI}{dt}. \tag{31.4}$$

where L is a constant for the particular coil being used. The negative sign appears here, by convention, as it did in the earlier expressions, and it signifies that the flux set up by any current caused by the e.m.f. will be in such a direction as to "try to prevent" any changes in flux. L is known as the *self inductance* or simply *inductance* of the coil. It is measured in henrys (VA^{-1} s). Thus, if the e.m.f. of the battery in the circuit of Fig. 31.14(a) is E, and the circuit resistance is R, we can write

$$E + \mathscr{E} = IR$$

because $E + \mathscr{E}$ is the resultant e.m.f. in the circuit. Hence

$$E - L\frac{dI}{dt} = IR. \tag{31.5}$$

Using suitable numerical values for L, R, and E, we can carry out a step-by-step solution of Eq. (31.5). This is shown below. At the instant we close the switch in the circuit, the time $t = 0$, the current $I = 0$, and the p.d. across the resistance, R (we are assuming that we can treat the resistive part of the inductor as being separate from the inductive part), is also zero. The initial rate of change of current is therefore given by

$$E - L\frac{dI}{dt} = 0$$

hence

$$\frac{dI}{dt} = \frac{E}{L}.$$

Thus, as we have stated above, the initial rate of rise of current depends only upon the inductance, L, and is independent of the resistance. With the numerical values used in the example (L = 250 H, E = 2.5 V) below this gives

$$\frac{dI}{dt} = 10 \times 10^{-3}\ \text{A s}^{-1} = 10\ \text{mA s}^{-1}.$$

If we assume (though of course this is an approximation) that the rate of rise of current remains at this value for a short interval of time Δt, then we can compute the increment in the current, ΔI, by using

$$\Delta I = \frac{dI}{dt}\Delta t.$$

We have chosen Δt = 0.1 s in the example, so, after 0.1 s from the instant of switching on, the current has increased by an amount (10 mA s^{-1}) × (0.1 s) = 1 mA, and so this is the current flowing at time t = 0.1 s. Then, using this value of current, I, we can compute a new value for the rate of change of current, dI/dt, hence a new value for ΔI, and thus

Table 31.1 Step-by-step calculation for growth of current in an inductive circuit.

$R = 500\ \Omega$, $L = 250$ H, $E = 2.5$ V

$$E - L\frac{dI}{dt} = IR$$

$$\frac{dI}{dt} = \frac{E - IR}{L}$$

$\Delta t = 0.1$ s

t s	I A × 10^{-3}	IR V	$E - IR$ V	$\frac{dI}{dt} = \frac{E - IR}{L}$ A s^{-1} × 10^{-3}	ΔI A × 10^{-3}
0.0	0	0	2.5	10	1.0
0.1	1	0.5	2.0	8	0.8
0.2	1.8	0.9	1.6	6.4	0.64
0.3	2.44	1.22	1.28	5.12	0.512
0.4	2.952	1.476	1.024	4.096	0.410
0.5	3.362	1.681	0.819	3.276	0.328
0.6	3.690	1.845	0.655	2.620	0.262
0.7	3.952	1.976	0.524	2.096	0.210
0.8	4.162	2.081	0.419	1.676	0.168
0.9	4.330	2.165	0.335	1.340	0.134
1.0	4.464				

obtain the current flowing at time $t = 0.2$ s. In this way we can compute the current at successive instants in time, and the results are shown in Table 31.1 and also plotted in Fig. 31.15.

After a very long time (we can write $t = \infty$) the current is changing no more, that is, $dI/dt = 0$, hence Eq. (31.5) simplifies to $E = IR$, and so (as has been said earlier) the final current depends only upon the resistance and is independent of the inductance.

The exponential character of the graph of Fig. 31.15 should be noted. The final steady current in the inductor is given by

$$I = \frac{E}{R}$$

and hence, with the numerical values used here, this gives $I = 5$ mA. From the graph it can be seen that the time taken for the current to rise from zero to half the final value, namely 2.5 mA, is 0.31 s. After a further time interval of 0.31 s the current has risen to a value which is less than the final value by one quarter of the final value, namely 3.75 mA. After a further 0.31 s the current is less than the final value by one eighth of the final value, namely 4.375 mA. In other words, the difference between the instantaneous current and the final steady current is halved in successive equal intervals of time. In this respect the growth of current in an inductor resembles the growth of p.d. across a capacitor when being charged,

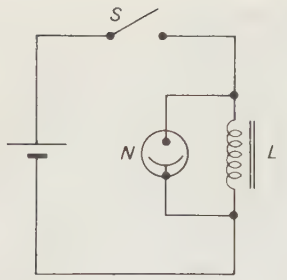


Fig. 31.16

and the decay of p.d. across a capacitor when being discharged.

Now we shall examine what happens when a current is switched off in an inductive circuit. If a large inductance has been used in the experiment of Fig. 31.14 we will probably have noticed that, when the switch is opened, a fat spark appears between the switch contacts as they separate. This observation suggests that there is, for a moment at least, a large voltage between the switch contacts. This may be surprising, because the voltage of the battery used in this experiment is not more than a few volts. The experiment of Fig. 31.16 demonstrates what is happening. *N* is a neon lamp designed to work at mains supply voltage (that is, about 250 V r.m.s.). With a high-voltage d.c. supply, or a suitable battery, we can show that a glowing discharge in the neon lamp does not strike until the voltage is about 100 V. In the circuit of Fig. 31.16 we close the switch and find (not surprisingly) that the lamp does not glow, for the battery voltage is far too small to cause a discharge in the lamp. When the switch is opened a momentary flash is observed, indicating that for a very brief time the voltage is about a hundred volts or greater. The demonstration can be made more convincing if one has an oscilloscope in which connection can be made directly to the *Y*-deflector plates. If the two deflector plates are connected to the two terminals of the inductor, one can display the very large voltage "spike" which occurs when the switch *S* is opened. If the sensitivity of the cathode ray tube is known (that is, the voltage needed to deflect the spot of light on the screen by one centimetre) it may be possible to measure the magnitude of this voltage spike. This very large but short-lived voltage is the e.m.f. induced in the inductor when the current decays very rapidly.

If any electrical circuit has a large inductance in it the problem of switching off a current in that circuit can be a serious one. It is perfectly possible for the spark which occurs between the switch contacts as they separate to persist and for there to be a continuous electric arc between the contacts, even when they are fully separated. In this condition the current will still be flowing: it will not have been switched off at all! Nearly all electrical machinery has considerable inductance, and so the design of suitable switches for use in the electrical supplies to factories and all industrial plant is of great importance. In some types of heavy-duty switchgear the electric arc which forms when the contacts separate is extinguished by a powerful blast of air, or quenched by having the whole switch mechanism immersed in a suitable oil insulating medium. Sometimes it is not possible to switch the current off rapidly: it has to be reduced gradually by switchgear of highly sophisticated design.

Chapter 32
ALTERNATING CURRENTS

32.1 Alternating current
32.2 Inductors in a.c. circuits
32.3 Capacitors in a.c. circuits
32.4 The transformer
32.5 Motors and alternating currents

32.1 ALTERNATING CURRENT

The alternating current (a.c.) from the mains, in Britain, has a frequency of 50 Hz. This is inconveniently high for us as we begin our study of alternating currents. We start therefore by considering an experiment to demonstrate a sinusoidal a.c. in which we use a very low frequency generator. There is no suitable dynamo available which gives a very low frequency a.c. which would be strong enough to be useful. We use, therefore, a special potentiometer, the principle of which is shown in Fig. 32.1, and this provides an alternating voltage of constant amplitude at any frequency we choose, and the voltage is sinusoidal provided that the shaft is rotated at a constant speed.

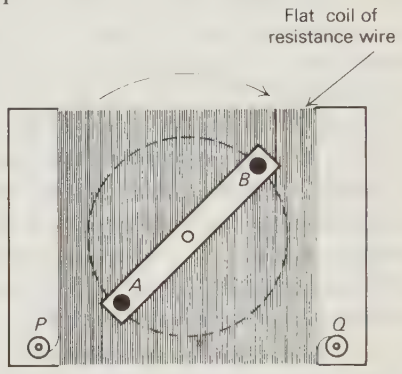


Fig. 32.1 A low frequency a.c. generator. Terminals *P* and *Q* are connected to a battery. *A* and *B* are sliding contacts mounted on an insulating, rotating arm. An alternating voltage appears between terminals *A* and *B*.

In the experiment we use this a.c. generator to light a lamp, as in Fig. 32.2, with the switch *S* in the left-hand position as shown. The lamp and ammeter are connected to a battery and variable resistor so that the lamp can be lit with d.c. instead of a.c. *A* is a centre-zero ammeter. When we rotate the shaft of the generator very slowly we observe that the lamp lights intermittently, reaching maximum brightness twice per revolution—once when the current flows one way, and again when the current flows the opposite way. If we spin the generator shaft faster we find that the lamp flickers more rapidly and that the variations in brightness become less; also, the ammeter cannot

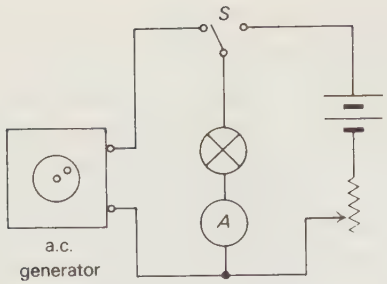


Fig. 32.2

respond adequately to the variations of current, so it oscillates just a small distance either side of the centre zero. We can run the generator at higher speed by using an electric motor to drive it, coupled by pulleys and a belt. Above a certain speed we find that the lamp appears to be steadily lit, and the needle scarcely moves at all, registering zero all the time. We have the slightly paradoxical situation of an ammeter showing zero current in a circuit with a lamp showing clearly the presence of a current. The problem immediately arises, how does one measure the strength of the current in this situation? An ordinary moving-coil meter is useless but the brightness of the lamp gives a clue: perhaps one could use the heating effect of the current as a measure of its strength. Suppose we arrange that the maximum current in the circuit (which, you will recall, is obtained at two instants during each revolution of the generator shaft) is 100 mA. We spin the shaft at high speed, and note the steady brightness of the lamp. Then we switch over to the battery and adjust the variable resistor until the lamp is at the same level of brightness, and we read the steady current on the ammeter. We will find that the ammeter reads about 70 mA. We deduce from these observations, therefore, that a sinusoidal alternating current whose peak value is 100 mA is equivalent to a steady d.c. of about 70 mA, if by equivalent we mean that the a.c. and the d.c. are carrying energy at the same average rate to the lamp bulb, where the energy is utilized as heat and light.

We now proceed to find a theoretical reason for this relationship between the peak value of a sinusoidal a.c. and the equivalent d.c. We can represent a sinusoidally varying a.c. by the expression

$$I = I_0 \sin \omega t. \qquad (32.1)$$

(Compare the expressions used for the displacement of a particle moving with simple harmonic motion, as in Unit 4.) As has been shown in Section 26.1, the rate of conversion of electrical energy into heat when a current I flows in a conductor whose resistance is R is I^2R, so at time t the instantaneous rate of conversion of energy by the a.c. is

$$I_0^2 R \sin^2 \omega t.$$

We are concerned with the average rate of conversion of energy by the a.c. because this is what determines the brightness of the lamp bulb. To find this we need only find the average value of the term $\sin^2 \omega t$ because the term $I_0^2 R$ is constant. We find this average value by taking the value of $\sin^2 \omega t$ at equally spaced instants in time during one complete cycle of variation, that is, from $\omega t = 0$ to $\omega t = 2\pi$ radians, or 360°. A suitable increment for ωt is 20°, and Table 32.1

Table 32.1

ωt	$\sin \omega t$	$\sin^2 \omega t$
0	0	0
20°	0.342	0.117
40°	0.643	0.413
60°	0.866	0.750
80°	0.985	0.970
100°	0.985	0.970
120°	0.866	0.750
140°	0.643	0.433
160°	0.342	0.117
180°	0	0
200°	−0.342	0.117
220°	−0.643	0.413
240°	−0.866	0.750
260°	−0.985	0.970
280°	−0.985	0.970
300°	−0.866	0.750
320°	−0.643	0.413
340°	−0.342	0.117
360°	0	0
	Sum:	9.000
	Mean value:	0.500

shows the values computed for one whole cycle. The whole cycle comprises 18 time intervals, so to find the mean value of $\sin^2 \omega t$ we divide the sum of all the values in the right-hand column by 18. This gives 0.500 as the mean value of $\sin^2 \omega t$. Thus the mean rate of conversion of energy per cycle can be written

$$0.500 I_0^2 R, \qquad \text{or} \quad \tfrac{1}{2} I_0^2 R.$$

If the values are computed to an accuracy of more than three significant figures the mean value of $\sin^2 \omega t$ will be not exactly $\frac{1}{2}$: but if smaller increments for ωt are used, then the result becomes nearer to the precise value of $\frac{1}{2}$. Using calculus to find the average value of $\sin^2 \omega t$ one can show that the precise value is, in fact, $\frac{1}{2}$, as follows.

The average value of $\sin^2 \omega t$ over one cycle is

$$\frac{\int_0^{2\pi/\omega} \sin^2 \omega t \, dt}{2\pi/\omega}$$

since $2\pi/\omega$ is the period (that is, the time for one cycle). Using the trigonometrical relationship $\sin^2 \omega t = \frac{1}{2}(1 - \cos 2\omega t)$ we can write the expression thus:

$$\frac{\frac{1}{2}\int_0^{2\pi/\omega} (1 - \cos 2\omega t) \, dt}{2\pi/\omega}.$$

After integration this becomes

$$\frac{\frac{1}{2}((2\pi/\omega) - 0)}{2\pi/\omega}$$

which equals $\frac{1}{2}$.

Now, if I_s is the value of the steady d.c. which produces a rate of conversion of energy equal to the average rate of conversion produced by the a.c., we can write

$$I_s^2 R = \tfrac{1}{2} I_0^2 R.$$

For the experimental value of 100 mA for I_0 we have

$$I_s = \frac{100}{\sqrt{2}} = \frac{100}{1.414} = 70.7 \text{ mA}$$

which agrees with the measured value of I_s. I_s is usually known as the *root mean square* value of the alternating current. It is equal to the peak value divided by $\sqrt{2}$ only if the pattern of time-variation is truly sinusoidal. In practice the root mean square (generally abbreviated to r.m.s.) value of the current is used as the measure of its strength, rather than the peak value.

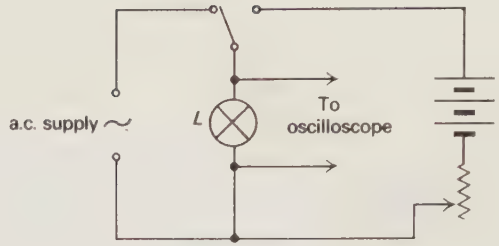


Fig. 32.3

The r.m.s. value and the peak value of an alternating potential difference are related in the same way, and the r.m.s. value is the one usually used to specify the magnitude of an alternating voltage. Thus, when we speak of the "250 volt a.c. mains" we mean that the r.m.s. value of the voltage is 250 V, and thus the peak value is $\sqrt{2} \times 250$, which equals 354 V.

An experiment which shows directly the relation between r.m.s. and peak voltage is represented in Fig. 32.3. Here an oscilloscope is used to measure the voltage; and the oscilloscope must be one which has a d.c. input so that when a steady voltage is applied to the input terminals the spot of light on the screen is displaced vertically and stays displaced, the displacement being proportional to the voltage. When the switch is in the left-hand position the lamp L is fed with a.c. from the low voltage terminals of a step-down transformer, and the peak value of the voltage is measured by reading off the amplitude of the sinusoidal trace on the oscilloscope. The switch is then put to the other position, the variable resistor adjusted until the lamp has the same brightness, and then the steady voltage is measured from the displacement of the trace on the oscilloscope. One cannot achieve high accuracy of measurement here, but it will be found that the peak value of the voltage is about 1.4 times the steady voltage used. Making the adjustment needed to get the lamp to exactly the same level of brightness with the d.c. as with the a.c. is a little

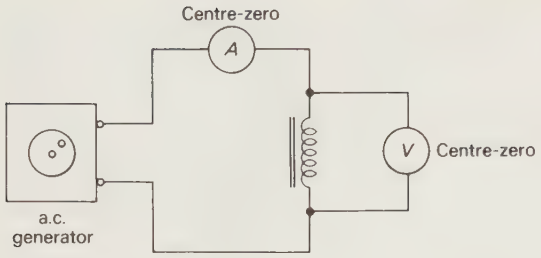


Fig. 32.4

difficult, but rapid switching of S back and forth repeatedly will help to provide the greatest sensitivity.

When an alternating voltage is applied to a conductor having some resistance, the current which flows is in phase with the voltage, and if the conductor obeys Ohm's law, then at every instant of time the potential difference V is related to the current I by the equation $V = IR$, and R is constant. In general, p.d. and current are not usually in phase in an a.c. circuit. Even in the case of a circuit with apparently pure resistance, p.d. and current are out of phase at very high frequencies. This is because the circuit does have a very small, but finite inductance. We proceed at once therefore to study those particular electrical devices which cause a phase-difference between voltage and current: inductors and capacitors.

Before making a theoretical analysis of what happens when an alternating p.d. is applied to an inductor we can demonstrate the phase-difference between voltage and current which is observable in a circuit which contains a very large inductance, using very low frequency a.c. supplied by the generator of Fig. 32.1. The two meters in Fig. 32.4 are adjusted so that they have a centre zero. When the generator shaft is rotated slowly the voltage is seen clearly to reach its maximum value at a different instant from the current. In fact the voltage will reach its maximum positive value before the current does, the phase-difference being usually less than, but never greater than, one quarter-cycle. The way in which this phase-difference arises can be appreciated if, having observed what happens when the generator shaft is rotated at steady speed, one rotates it in jerks, a quarter revolution at a time. The way in which the current lags behind the voltage will then become very apparent. That this phase difference is not due to some difference in the action of the two meters should be demonstrated. This is done by using a resistor instead of the inductor: then it will be clearly seen that the current and voltage are in phase.

To understand how this phase-difference comes about we study the graphs of Fig. 32.5. Here the phase-difference is exactly one quarter-cycle: this is the phase-difference we get whenever the effect of resistance in the circuit is negligible compared with the effect of inductance. We assume here that the current and voltage have been varying for some while and that a steady state of sinusoidal variation of both quantities has been reached. To simplify the argument we will assume that the resistance of the circuit is zero. We know already what happens if a steady voltage is applied to an inductor of zero resistance (Section 30.1): the current rises at a constant rate, and at such a rate that the back e.m.f. induced by the changing magnetic flux in the inductor is equal and opposite to the applied voltage. Consider now what is happening at the instant marked B in Fig. 32.5. The voltage at this instant has its maximum value, say, $+6$ V. The current is rising at a rate of, say, $+2$ A s^{-1}. The inductor has an inductance of 3 H (henrys), and so the back e.m.f., which equals the product: $-$(inductance) $\times$ (rate of change of current), is $(-3) \times (+2) = -6$ V.

If the applied voltage stayed at a steady value of $+6$ V, then the current would have to continue rising at the same rate, $+2$ A s^{-1}. But the applied voltage decreases, and at point C it is at half its maximum value, namely at $+3$ V. At this instant we see that the gradient of the current curve is less than at B: in fact it is just half what it was at B: it is $+1$ A s^{-1}. At point D the applied voltage is zero, and therefore the rate of change of current is zero, making the back e.m.f. zero. At E the voltage is -3 V: it has changed

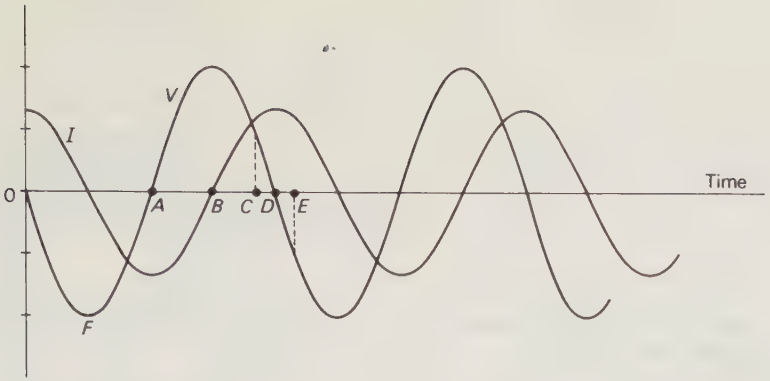


Fig. 32.5 Current and voltage in a circuit containing pure inductance.

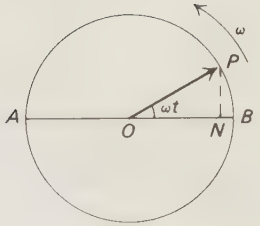


Fig. 32.6 Rotating vector representation of a sinusoidally varying quantity.

sign, so the rate of change of current must also have changed sign, and it is now -1 A s^{-1}: *decreasing* at a rate of one ampere per second, so that the back e.m.f. is $(-3) \times (-1) = +3$ V. At point F the voltage has its maximum negative value, -6 V, and the rate of change of current its greatest negative value, -2 A s^{-1}, so that the back e.m.f. is $(-3) \times (-2) = +6$ V. We need not continue this laborious discussion any further to make the point that the current changes in such a way that at every instant the back e.m.f. is equal and opposite to the applied voltage.

A very convenient method of representing sinusoidally varying quantities, especially when there are phase-differences between them, is the rotating vector technique. Unit 4 showed how sinusoidal motion could be derived from uniform motion in a circle: in a similar way the variation of any sinusoidally varying quantity can be derived from a vector rotating with uniform speed. In this context these rotating vectors are often referred to as *phasors*.

In Fig. 32.6 imagine OP to be a rod, like the spoke of a wheel, rotating anticlockwise about O, with uniform angular velocity ω. The imaginary clock which measures the time t is started at the instant when the end P is at B. At the instant shown, therefore, the angle $P\hat{O}N$ is ωt. The distance ON, which is the projection of OP upon the line AOB, is $OP \cos \omega t$. Thus ON is a sinusoidally varying quantity whose amplitude

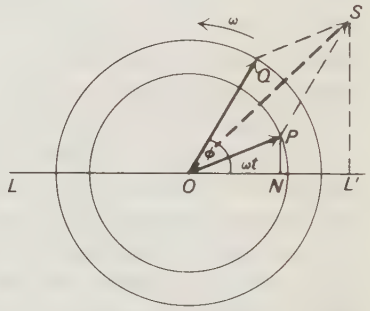


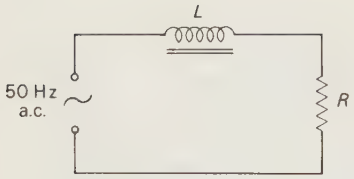
Fig. 32.7 Rotating vector representation of two sinusoidally varying quantities of different amplitude and phase.

is equal to the radius of the circle. In this way we can represent two sinusoidally varying quantities with the same frequency which have a phase-difference of ϕ between them:

$$x_1 = A_1 \cos \omega t$$

$$x_2 = A_2 \cos (\omega t + \phi)$$

by the diagram of Fig. 32.7. If these two quantities are ones which can be added (for instance, the voltages across a pair of electrical components connected in series), then their sum at any instant is represented by the projection of OS upon the horizontal LOL'. OS is the diagonal of the parallelogram which has OP and OQ as two of its sides: in other words, $\boldsymbol{OS}$ is the vector sum of the two vectors $\boldsymbol{OP}$ and $\boldsymbol{OQ}$.



L = 100 H, with a resistance of about 300 Ω
R = 33 kΩ

Fig. 32.8

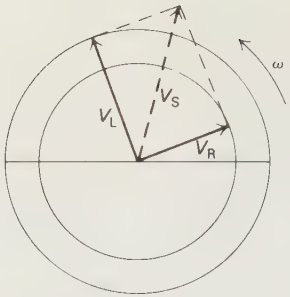


Fig. 32.9 Rotating vector diagram for *L–R* circuit.

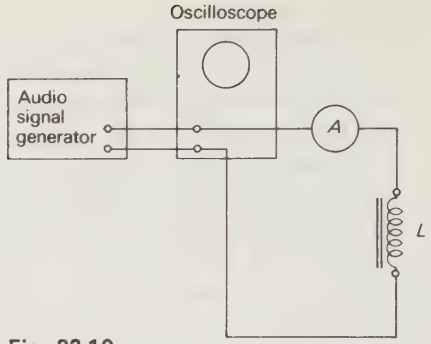


Fig. 32.10

32.2 INDUCTORS IN A.C. CIRCUITS

The rotating vector notation will prove helpful in analysing a.c. circuits, and we now go on to consider some very simple circuits.

Consider a circuit containing simply a resistor, R, and an inductor, L, connected in series with a supply of sinusoidally alternating voltage. The inductor, we shall assume, has negligible resistance. A circuit like this can easily be set up for demonstration, but the demonstration is convincing only if the effect of the finite resistance of the inductor *itself* is masked by the effect of its inductance. Suitable values are suggested in Fig. 32.8. If an oscilloscope is connected in turn across the terminals of the supply, across the inductor alone, and across the resistor alone, the surprising result is found that the sum of the voltage across the inductor and the voltage across the resistor is *greater* than the voltage across the terminals of the supply! The vector diagram of Fig. 32.9 shows how this comes about.

It must be remembered that the voltage across a pure inductance leads the current flowing through it with a phase-difference of one quarter-cycle. The voltage across a pure resistance, on the other hand, is in phase with the current through it. Thus the voltage across the inductor, whose peak value we shall call V_L, is a quarter-cycle ahead of the voltage across the resistor, whose peak value we call V_R. In the rotating vector diagram of Fig. 32.9 the vectors are rotating anticlockwise, and the angle between V_L and V_R is $\pi/2$ radians, or 90°, all the time. The voltage of the supply is represented by the vector V_S and this is the vector sum of V_L and V_R. The lengths of these vectors as shown in Fig. 32.9 are in the ratio 4:3. Let us suppose that V_L = 4 V and V_R = 3 V: then V_S is given by (Pythagoras' theorem)

$$V_S^2 = V_L^2 + V_R^2$$

hence

$$V_S = 5 \text{ V}.$$

So clearly, the arithmetical sum of V_L and V_R is not equal to V_S: four plus three does not equal five!

Calculations based on vector diagrams like that of Fig. 32.9 do not require the presence of the circles, or the horizontal diameter: so in future we shall omit these parts of the diagram and simply keep the vectors themselves.

An inductor has the important property of impeding the flow of an a.c., and we shall now discuss this property. The effect is convincingly shown in the experiment depicted in Fig. 32.10. An audio signal generator (with low impedance output) is used to provide an alternating voltage whose frequency can be varied as desired. The oscilloscope is used to monitor the output of the signal generator, simply to ensure that the peak value of the voltage remains constant when the frequency is varied. A is an a.c. ammeter of suitable range. L is an inductor consisting of a coil wound on a former which can be slipped on a magnetic core, and the inductance can be varied by

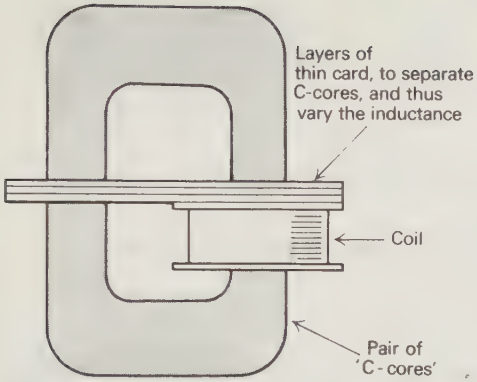


Fig. 32.11

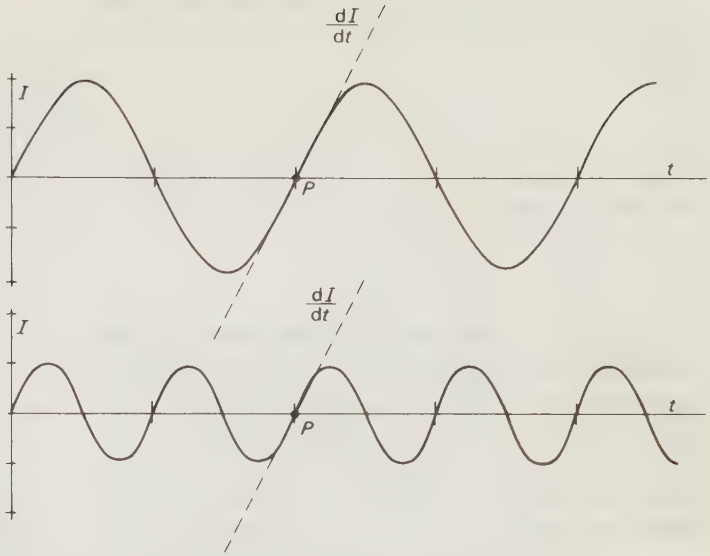


Fig. 32.12 The rate of change of current at point *P* is the same in both cases.

altering the distance between the two halves of the magnetic core, as shown in Fig. 32.11. The further apart the two halves, the smaller is the inductance. We find that, when the inductance is reduced, the current increases. When the frequency of the alternating voltage is increased the current is reduced, and a few measurements will suggest that, for a constant peak value of supply voltage, the current is inversely proportional to the frequency.

Why should the current (its peak value, let us say) be inversely proportional to the frequency? We can argue it in the following way, using imagined numerical values as we did in the argument of the last section. Referring back to that argument, suppose we start, as before, by considering the instant when the voltage applied to the inductor (assumed to have inductance only, and no resistance) is +6 V. The inductor has an inductance of 3 H. At this instant the current must be changing at a rate of $+2\ \text{A s}^{-1}$, so that the back e.m.f., given by $\mathscr{E} = -L(dI/dt)$, is $(-3) \times (+2) = -6$ V. Suppose now that we double the frequency of the supply voltage, keeping its peak value unchanged. If the peak value of the current remained unchanged, then the rate of change of current at this instant would be $+4\ \text{A s}^{-1}$, simply because the frequency has doubled and hence the rate of change of current at each point during a cycle is twice what it was before. But a rate of change of current of $+4\ \text{A s}^{-1}$ would mean that the back e.m.f. is $(-3) \times (+4) = -12$ V. But we know that the back e.m.f. must be equal and opposite to the applied voltage. This must mean that the peak value of the current is half what it was before. Figure 32.12 shows the time-traces of the current at the old frequency and the current at the new frequency, for comparison. It follows, therefore, that if the supply voltage is kept constant, the peak value of current (hence also the r.m.s. value) must be inversely proportional to the frequency.

A similar argument can be constructed to show that the current must also be inversely proportional to the inductance when the inductance is varied. Again, suppose we consider the instant when the voltage applied to the inductor is +6 V. If we double the inductance, so that it becomes 6 H, then back e.m.f. is $(-6) \times (+1) = -6$ V. This must be the case, if the frequency is not altered, and because the

frequency is not altered it implies that the peak value of the current must be halved. Thus we can write, in symbols:

$$I_0 \propto \frac{1}{f} \quad \text{and} \quad I_0 \propto \frac{1}{L},$$

and these expressions can be combined and written

$$I_0 \propto \frac{1}{fL}.$$

To find the constant of proportionality in this expression we use calculus, as follows. For the instantaneous current flowing in the inductor we write

$$I = I_0 \sin 2\pi ft.$$

Now, the e.m.f. induced by the changing current is given by

$$\mathscr{E} = -L\frac{dI}{dt}$$

hence, substituting for I, we get

$$\mathscr{E} = -L2\pi fI_0 \cos 2\pi ft.$$

The applied voltage V is given by

$$V = -\mathscr{E}$$

so that

$$V = 2\pi fLI_0 \cos 2\pi ft$$

which can be written

$$V = V_0 \cos 2\pi ft,$$

where

$$V_0 = 2\pi fLI_0$$

and is the amplitude of the applied p.d. Thus

$$I_0 = \frac{V_0}{2\pi fL}.$$

This expression can be rearranged and written

$$\frac{V_0}{I_0} = 2\pi fL, \qquad (32.1)$$

and here we see that it resembles the expression which defines the resistance of an electrical device, $V/I = R$. But the right-hand side of the expression, $2\pi fL$, is not

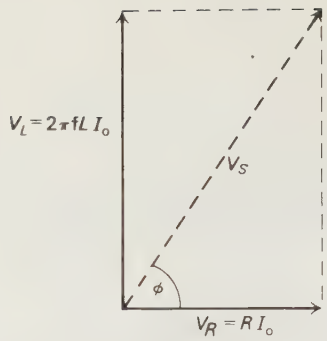


Fig. 32.13 Vector diagram for L–R circuit.

the resistance of the inductor (which we have assumed to have zero resistance anyway). It must be remembered that the voltage and current are quarter of a cycle out of phase. The quotient V/I only defines *resistance* when the voltage and current are in phase. Here this quotient defines a different quantity which is called the *reactance* of the inductor. It is, like resistance, measured in ohms but, unlike what happens in a resistance, no power is dissipated when a current flows in a pure inductance. It is not immediately obvious that the quantity $2\pi fL$ should be measurable in ohms, so let us see how this comes to be so. Frequency, f, has units: s^{-1}. Inductance has units: $V\ s\ A^{-1}$. So the product fL has units: $(s^{-1}) \times (V\ s\ A^{-1}) = V\ A^{-1}$ or ohms.

So far our arguments have been based on the assumption that the inductor has no resistance. We are now equipped to understand the function of an inductor which does have appreciable resistance. Such an inductor can be treated simply as if it were a pure inductance in series with a pure resistance, as in Fig. 32.11 earlier. That vector diagram therefore applies to this situation, and we reproduce this diagram here in Fig. 32.13, without the circles, as we promised we would do. The peak value of the voltage across the series combination of inductance and resistance is the vector V_S in the diagram. Now, the peak voltage across the inductance is V_L and we have already shown that V/I for a pure inductance equals $2\pi fL$,

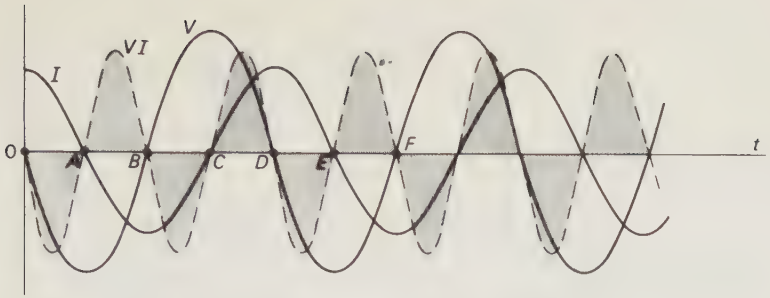


Fig. 32.14 Power in a purely inductive circuit.

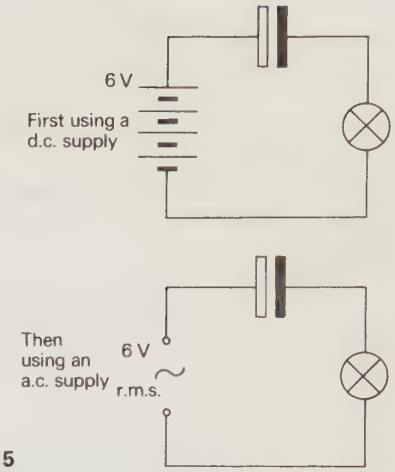


Fig. 32.15

and so $V_L = 2\pi fLI_0$. The peak voltage across the resistance R is given by $V_R = RI_0$. Hence, by Pythagoras' theorem,

$$V_S^2 = V_L^2 + V_R^2$$

$$= (2\pi fLI_0)^2 + (RI_0)^2$$

hence

$$\frac{V_S}{I_0} = \sqrt{(2\pi fL)^2 + R^2}. \qquad (32.2)$$

Now this quotient, voltage divided by current, is neither a reactance (where V and I are a quarter-cycle out of phase) nor a resistance (where V and I are in phase). We call it the *impedance*, and note that the voltage V_S across the inductor, and the current (which is in phase with V_R), have a phase-difference of ϕ between them. The impedance (Z) of an electrical device, or a whole circuit, is measured in ohms.

In the experiments described earlier, to demonstrate the properties of a pure inductance, we had to cheat a little by choosing suitable inductors and frequencies so that the effect of any resistance in the inductor itself was negligible, being masked by the effect of the inductance. What we did, in fact, was to choose values of L and f such that, in the vector diagram of Fig. 32.13, V_L was much greater than V_R and so the phase-angle ϕ was very close to $\pi/2$ radians, or 90°.

It was said earlier, without any reasons given, that in a pure inductance no power is dissipated. This will now be explained.

We base our argument upon the graphs of voltage and current against time which were shown earlier, and are shown again here in Fig. 32.14. The shaded curve shows how the product of voltage and current varies with time.

Between points O and A on the time-scale, I is positive and V is negative, and so the product VI is negative. Between A and B both I and V have negative values, thus their product is positive. Between B and C, V is positive but I is negative, and thus the product is negative. Between C and D the product is positive. When the product VI is positive, energy is being fed from the source into the inductor and being stored in the magnetic field of the inductor. (This process of energy storage is described later in Unit 7.) When VI is negative, energy is being returned from the inductor to the source. During one complete cycle, therefore, it is clear that there is no net energy transfer from source to inductor, or in the opposite direction.

32.3 CAPACITORS IN A.C. CIRCUITS

One basic fact about capacitors and alternating currents is demonstrated strikingly by the experiment of Fig. 32.15. A capacitor whose capacitance is about

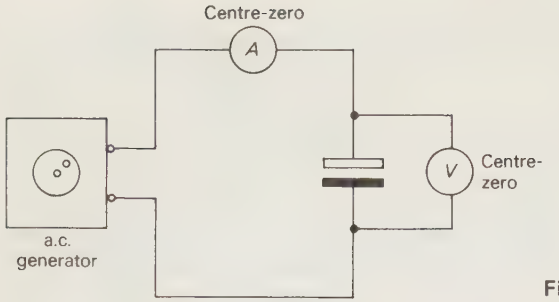


Fig. 32.16

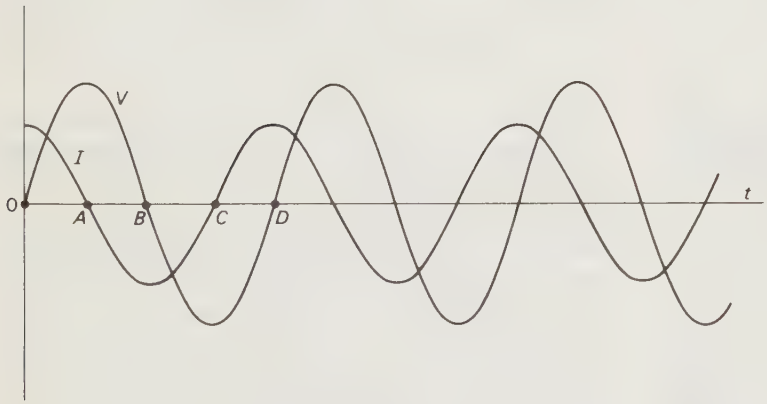


Fig. 32.17 Current and voltage in a circuit containing pure capacitance.

1000 μF is connected in series with a 6 V battery and a 6 V 0.06 A lamp. (Care is taken to connect the capacitor with the correct polarity.) The lamp does not light, although it lights perfectly well when connected directly to the battery. This is as expected. Then the battery is replaced by a 6 V r.m.s. a.c. supply, and the lamp is observed to light to almost full brightness. (Although the capacitor is here being subjected to a voltage of about $6 \times \sqrt{2} = 8.5$ V, of the wrong polarity, every other half-cycle, it usually survives the treatment, provided that its working voltage is 15 V or greater.) The experiment shows that whereas the capacitor does not conduct a d.c., it does effectively conduct an a.c. The capacitor is, of course, being charged, discharged, and then charged with the opposite polarity, and discharged again, fifty times per second, and it is the charging current flowing in the wires of the circuit which lights the lamp.

To investigate the phase relationship between the voltage across a capacitor and the current in the circuit we can use the arrangement of Fig. 32.16. The voltage is found to lag behind the current (unlike the case of an inductor), reaching its maximum value about one quarter-cycle after the current reaches its maximum. The variations of voltage and current are shown in Fig. 32.17. To understand why the voltage and current vary in this way we must remember that current is the rate of flow of electric charge. At the instant A on the time-scale of the graph (Fig. 32.17), the voltage, and hence the amount of charge on the capacitor plates, is a maximum: the charge at this instant has ceased flowing into the plates and is about to flow away from them. Between A and B the voltage

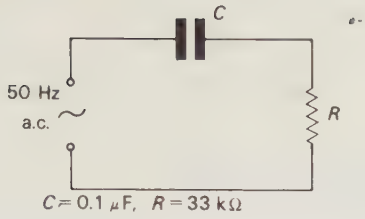


Fig. 32.18

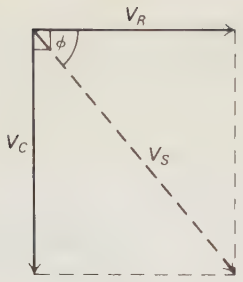


Fig. 32.19 Vector diagram for *R–C* circuit.

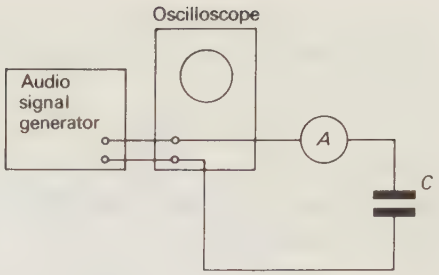


Fig. 32.20

is decreasing which means that the amount of charge on the plates is decreasing also: the rate of decrease is a maximum at *B*, where the voltage curve has its greatest negative slope. At this instant therefore the current—the rate of flow of charge—has its greatest negative value. The rate of flow of charge then becomes less, as charge builds up on the capacitor plates, charging them with opposite polarity. At *C* the charge on the plates is again a maximum, and the rate of flow of charge, the current, is zero at this instant. You can continue the argument in this way, referring to the graphs.

Consider now a circuit containing capacitance and resistance, as in Fig. 32.18. An experiment similar to the one with an inductor, and shown in Fig. 32.18, is well worth doing. Suitable values for *C* and *R* are shown in Fig. 32.18. The peak value of the voltages across *C* and *R* are measured with an oscilloscope and compared with the peak value of the voltage across the supply terminals. The sum of the first two, $V_C + V_R$, is found to be greater than the supply voltage, V_S. The vector diagram of Fig. 32.19 explains this situation. Remember that these vectors are imagined rotating anticlockwise: thus V_C is lagging behind V_R by one quarter-cycle (V_R is in phase with the current flowing through *R*, and we have shown already that the voltage across a capacitor lags behind the current). V_S is the vector sum of V_C and V_R and its magnitude is clearly less than the arithmetical sum of the magnitudes of V_C and V_R.

It is easy to get the phase relations between current and voltage for a capacitor confused with those for an inductor, and a useful mnemonic is the word *CIVIL*. Take the first three letters, *CIV*, and interpret them thus: in a capacitor (*C*) the current (*I*) leads the voltage (*V*). Take the last three letters, *VIL*, and interpret them: the voltage (*V*) leads the current (*I*) in an inductor (*L*).

When a capacitor is present in an a.c. circuit, no electrical energy is dissipated in the capacitor, unless its dielectric has finite resistance. In an ideal capacitor the dielectric has infinite resistance. Using an argument similar to that for an inductor, based on graphs of current and voltage as in Fig. 32.17, you should show that, during one quarter-cycle, energy is fed from the source to the capacitor, and that this energy is returned to the source during the next quarter-cycle.

The reactance of a capacitor determines the relation between the magnitude of the voltage across it and the current in the circuit. The way in which the reactance depends on the capacitance and the frequency can be demonstrated by the experiment of Fig. 32.20. The capacitance can be varied by connecting other capacitors of the same value in parallel with *C*. In this way it is found that the current registered by the a.c. ammeter, *A*, is directly proportional to the capacitance: $I \propto C$. If the frequency is varied, making sure that the output voltage of the signal generator is kept constant (this is what the oscilloscope is for), it is seen that the current is also directly proportional to the frequency: $I \propto f$. Thus $I \propto fC$. Also (although this is perhaps obvious and scarcely needs testing), if *f* and *C* are kept constant, the current is found to be directly proportional to the

supply voltage, V. Thus the reactance, if we define it in the same way as for an inductor, is related to f and C thus:

$$\frac{V}{I} \propto \frac{1}{fC}.$$

Using calculus one can show that the complete expression should be:

$$\frac{V}{I} = \frac{1}{2\pi fC}.$$

This is done as follows. We can write an expression for the applied voltage thus:

$$V = V_0 \sin 2\pi ft,$$

where V_0 is the peak value of this voltage. The charge on the capacitor is given by

$$Q = CV = CV_0 \sin 2\pi ft.$$

The current flowing in the circuit is given by

$$I = \frac{dQ}{dt}$$

so that, substituting for Q in this expression, we obtain

$$I = 2\pi fCV_0 \cos 2\pi ft$$

which can be written

$$I = I_0 \cos 2\pi ft,$$

where I_0 is the peak value of the current, and

$$I_0 = 2\pi fCV_0.$$

Hence

$$\frac{V_0}{I_0} = \frac{1}{2\pi fC}. \qquad (32.3)$$

In Fig. 32.21 we see again the vector diagram for a circuit containing capacitance and resistance. V_C, the peak value of the p.d. across the capacitor, is, as we have just explained, equal to $I_0/2\pi fC$. V_R, the p.d. across the resistor, is RI_0. Thus, by Pythagoras' theorem,

$$V_S^2 = V_c^2 + V_R^2$$
$$= \left(\frac{I_0}{2\pi fC}\right)^2 + (RI_0)^2,$$

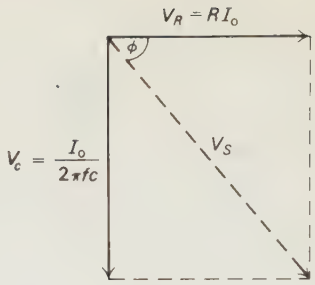


Fig. 32.21 Vector diagram for R–C circuit.

hence

$$\frac{V_S}{I_0} = \sqrt{\left(\frac{1}{2\pi fC}\right)^2 + R^2}. \qquad (32.4)$$

This quotient is the impedance of the circuit. Note that in this case the supply voltage, V_S, lags behind the current in the circuit (which is in phase with V_R) by an amount ϕ. In an inductive circuit the voltage leads the current.

In the circuit of Fig. 32.18 power is dissipated only in the resistance, R. The peak value of the p.d. across R is V_R and

$$V_R = V_S \cos \phi$$

and if I_0 is the peak value of the current, then the peak power dissipation is the product $V_R I_0$. But the mean power dissipation is the product of the r.m.s. values of p.d. and current, and so the mean power equals

$$\frac{V_R}{\sqrt{2}} \times \frac{I_0}{\sqrt{2}}$$

which equals

$$\tfrac{1}{2} V_S I_0 \cos \phi. \qquad (32.5)$$

The term $\cos \phi$ is often known as the *power factor* for the circuit. If the circuit were purely resistive ϕ would be zero and $\cos \phi$ would be unity. If the circuit were purely capacitative ϕ would be $\pi/2$ radians and $\cos \phi$ equal to zero, and, as we already know, no power is dissipated in a purely capacitative circuit.

Capacitors and inductors behave in a.c. circuits in ways which are similar but quite distinct. One of the most important aspects of their contrasting be-

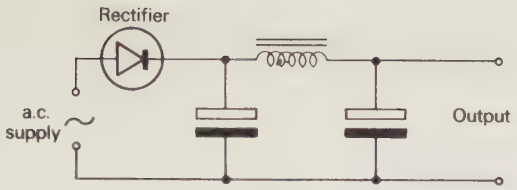


Fig. 32.22 An L–C smoothing circuit.

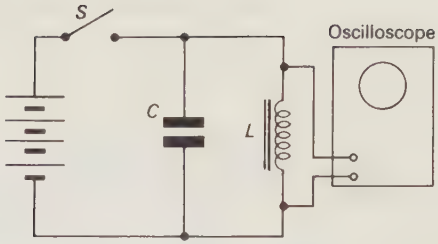


Fig. 32.23

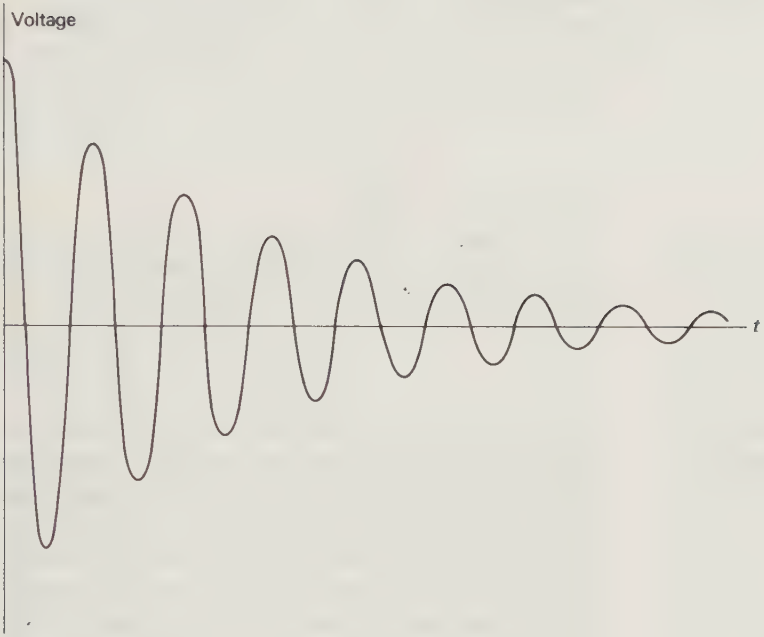


Fig. 32.24 The oscillations in the voltage across L and C in the circuit of Fig. 32.23.

haviour is the fact that the higher the frequency the better a capacitor conducts an a.c. (i.e. the lower its reactance) and the worse an inductor conducts. These properties enable one to design a smoothing circuit—one which reduces fluctuations in rectified supply of voltage and current. Figure 32.22 shows a typical smoothing circuit. The two capacitors can be thought of as tending to short-circuit any varying currents from the supply while the inductor (usually called a choke when used for this particular function) behaves like an open-circuit connection for varying currents. Often an inductor is not needed, and a resistor can be substituted between the two capacitors.

We have finally to discuss what happens in a circuit which contains both inductance and capacitance. It is worth speculating on what one might expect to happen if one connected a pure inductance to a pure capacitance and then by some means started an a.c. flowing in the circuit. If, as has already been stated, no power is dissipated in either of these devices, what would happen to the energy in the circuit? A first clue to answering this speculative question is provided by the experiment of Fig. 32.23. The oscilloscope, with its time-base set to give a very slow sweep rate (about 1 trace per second is suitable), records the voltage across the inductor and capacitor. Closing the switch S charges the capacitor. When the switch is opened, a trace like that shown in Fig. 32.24 is seen on the screen. We have here an oscillating voltage which decays with what looks suspiciously

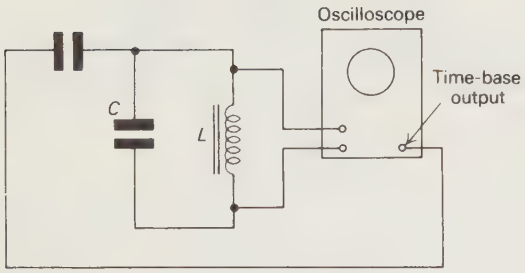


Fig. 32.25

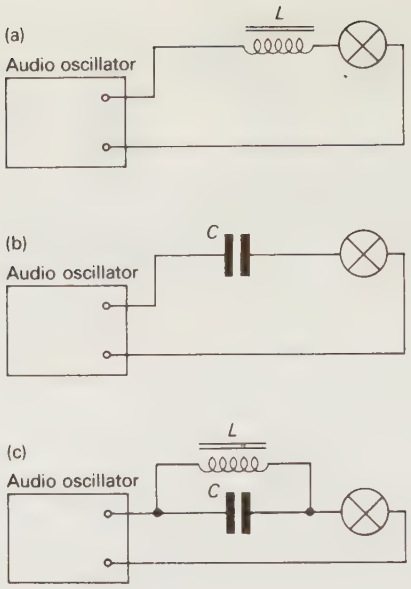


Fig. 32.26

like an exponential pattern. If a resistor is connected across L and C the decay of oscillation is found to be more rapid: we have heavier damping. This suggests that the decay of oscillation may be due, in part at least, to the loss of energy being dissipated in the resistance in the circuit.

It is convenient, for further experiments, to arrange that a continuously repeating trace on the oscilloscope screen can be obtained. Figure 32.25 shows how this can be done. A sharp voltage pulse appears at the time-base output socket every time the spot flies back to the left-hand side, and this can be used to charge the capacitor which then discharges through the inductor as the spot traces from left to right.

We can now investigate experimentally what factors determine the frequency of the oscillations. The inductance can be varied by using the technique described in Section 32.2 and illustrated in Fig. 32.11. It is found that increasing the inductance reduces the frequency of oscillation. The capacitance can be altered by substituting other capacitors for C, or by adding further capacitors connected in parallel with C: this will show that increasing the capacitance also reduces the frequency of oscillation. Thus we see that this kind of circuit has a characteristic frequency which depends upon the inductance and the capacitance in it. Suppose we have used an inductance of about 0.1 H and a capacitor of about 1 μF, we will find that this characteristic frequency is about 500 Hz. This frequency can be measured in practice on the oscilloscope by comparing the trace of the oscillation with the trace produced by a signal derived from the a.c. mains (50 Hz).

A further experiment, using the same values of L and C as were used in the previous experiment, is shown in Fig. 32.26. The audio oscillator must be one which has a low-impedance output so that it can light a low-voltage lamp to full brightness even when its amplitude control is at less than its maximum setting. With the inductor alone, as in (a), the lamp lights at low frequencies but goes out when the frequency is increased: this is as expected. With the capacitor alone, as in (b), the reverse situation is found, and the lamp only lights when the frequency is high. With L and C in parallel, as in (c), the lamp is

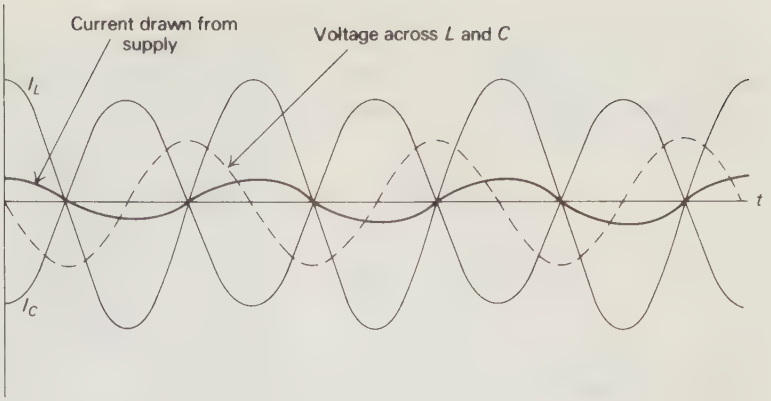


Fig. 32.27 Current and voltage in a resonant circuit.

found to light at both high and low frequencies, but over a small intermediate range of frequencies it is off. Not surprisingly, perhaps, these frequencies are close to the characteristic frequency of oscillation discovered earlier.

The demonstration of Fig. 32.26(c) is improved by having two other lamps, one connected next to L and in series with it, and one next to C in series with it. Then one finds a paradoxical situation: when the signal generator is set at the characteristic oscillation frequency of the L–C circuit, the two lamps next to L and C are both on while the original lamp is off. Using an a.c. ammeter instead of the lamps reinforces these observations. At the characteristic frequency the current in the L-branch and the current in the C-branch of the circuit are both greater than the current drawn from the signal generator. This result seems a gross contradiction of Kirchhoff's first law (Section 25.1). But what we know about the phase-relationships of current and voltage in inductors and capacitors should resolve the paradox. We know that in a pure inductor the voltage across it leads the current by quarter of a cycle, and that in a pure capacitor the current leads the voltage, also by a quarter-cycle. Now, in our circuit, the capacitor and inductor are connected in parallel, so that the voltage across each must be the same in magnitude and phase. Thus the current in the inductor branch must be one half-cycle ahead of the current in the capacitor branch, that is, these two currents have opposite phase. The situation is shown in Fig. 32.27.

The way the currents go, as shown in Fig. 32.27, implies that nearly all the current is surging around the L–C circuit, first clockwise, then anticlockwise. It is the electric charge therefore that is in a state of oscillation, sloshing back and forth, as it were, like water in a bath. At the instant when the capacitor carries maximum charge (when the voltage is a maximum), the current is zero, hence the magnetic flux in the inductor is zero. This is the instant when all the energy in the oscillatory circuit is stored in the capacitor. A quarter of a cycle later the capacitor is uncharged, and the current is a maximum: this is the instant when the energy is all stored in the inductor. Thus, neglecting any energy dissipation in the resistance of the circuit, the energy is transferred back and forth from inductor to capacitor, from being associated with a magnetic field to being associated with an electric field, alternately.

We said earlier that the characteristic frequency of the L–C circuit depended upon the inductance and the capacitance, and that increasing either of these quantities caused a reduction of the frequency. We shall now analyse the circuit theoretically to find precisely the relation between the frequency and L and C. Suppose the charge on the capacitor at some instant of time is Q, then the p.d. across the capacitor V_C is given by: $V_C = Q/C$. The p.d. across the inductor

V_L is related to the rate of change of current through it, the relation being $V_L = -L(dI/dt)$. Now the current, I, is the rate of flow of charge, and hence is equal to the rate of change of charge on the capacitor. In symbols,

$$I = \frac{dQ}{dt}.$$

Hence

$$\frac{dI}{dt} = \frac{d}{dt}\left(\frac{dQ}{dt}\right) = \frac{d^2Q}{dt^2}.$$

Since the inductor and capacitor are connected in parallel, we must have $V_C = V_L$. Therefore we can write

$$\frac{Q}{C} = -L\frac{d^2Q}{dt^2}$$

hence

$$\frac{d^2Q}{dt^2} = -\frac{1}{LC}Q.$$

This equation is of the same kind as one used in Unit 4, and which was found to represent simple harmonic motion, namely:

$$\frac{d^2x}{dt^2} = -\omega^2 x.$$

The solution to this was $x = a_0 \cos \omega t$ where $\omega = 2\pi f$, and the quantity x varies sinusoidally with time, the frequency being f. Thus the solution to this new equation can presumably be written

$$Q = Q_0 \cos \omega t$$

implying that the charge on the capacitor varies sinusoidally, and the frequency of oscillation can be derived from the equation

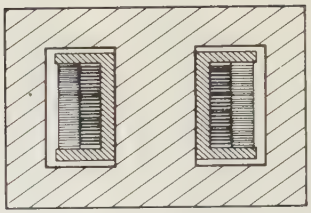
$$\omega^2 = \frac{1}{LC}$$

hence

$$f = \frac{1}{2\pi\sqrt{LC}}. \qquad (32.6)$$

32.4 THE TRANSFORMER

The phenomenon of mutual induction was investigated in Section 31.3. The alternating current transformer is a device which makes use of mutual induction to step-up or step-down a voltage.



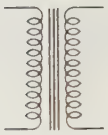


Fig. 32.28 A section through a transformer and the conventional symbol for a transformer.

The two coils, or windings, of a simple transformer are generally wound concentrically on a common core made of soft magnetic alloy. For reasons which will be explained later this core is usually built up of thin layers of metal, called *laminations*, which are electrically insulated from each other by very thin layers of paper. Figure 32.28 shows a cross-sectional view of such a transformer, and also the conventional circuit-symbol for a transformer. The form of the magnetic core is such that it provides two closed paths for the magnetic flux produced by currents in the windings: in a well-designed transformer there is negligible leakage of magnetic flux outside the core. Having the two windings concentric upon a common core ensures good flux linkage between the two windings, that is to say, nearly all the flux produced by current in the primary winding threads through, or links, the secondary winding.

Before attempting to analyse the action of a transformer theoretically we will look at the results of some simple experiments with transformers. For the first experiment we can use a "do-it-yourself" transformer in which we make our own secondary winding. Figure 32.29 shows two kinds of design suitable for this purpose. In each of these the primary winding has a fixed number of turns and can be connected directly to the 250 V a.c. mains. We take

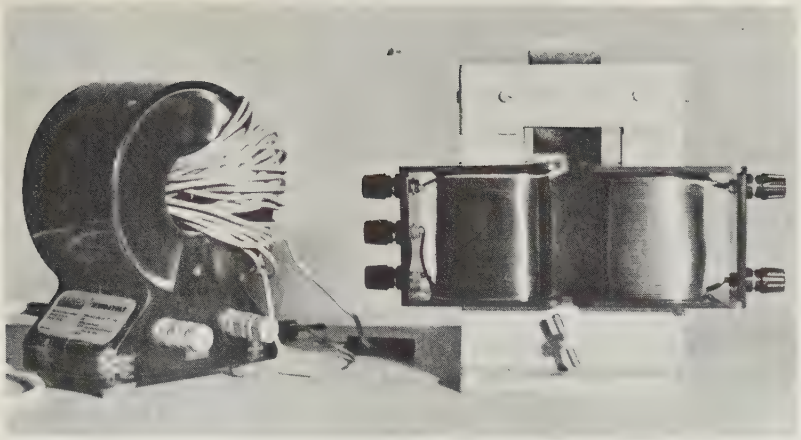


Fig. 32.29 Examples of laboratory transformers. (Photograph by E.J.W.)

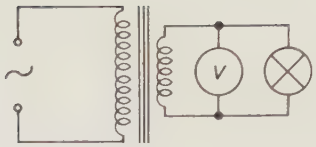


Fig. 32.30

a length of insulated copper wire and wind first one, then two, then three, and more, turns around the transformer core. For each number of turns used we connect the ends of the wire to a suitable low-voltage lamp with an a.c. voltmeter connected in parallel with it to measure the secondary voltage of the transformer. The arrangement is thus as shown in Fig. 32.30.

When we plot a graph of secondary voltage against n_s, the number of turns in the secondary winding, we find that it is a straight line. Thus, in principle, it is a simple matter to find how many turns we need to obtain any other chosen value of secondary voltage. For the primary winding the number of turns is determined by the primary voltage and, for a transformer with good flux-linkage between the two windings, the relationship

$$\frac{\text{primary voltage}}{\text{secondary voltage}} = \frac{\text{number of primary turns}}{\text{number of secondary turns}}$$

is very nearly obeyed. Thus, by choosing suitable numbers of turns for the two windings we can convert the voltage of an alternating supply into a higher voltage (as in a step-up transformer) or into a lower voltage (step-down transformer).

Now we take a closer look at the voltages and currents in a transformer when it is working. The circuit of Fig. 32.31 is used to apply different loads to a transformer, and to measure voltage and current in primary and secondary.

The results of such an experiment are shown in Table 32.2. The first set of readings was taken with the secondary on open circuit (that is, with nothing at all connected to it apart from the voltmeter which we can assume takes negligible current). The different loads were simply different numbers of 12-volt lamps connected in parallel. Also, the resistance of the primary and secondary windings was measured with an ohm-meter. The results were: primary 9.0 Ω, secondary 0.2 Ω.

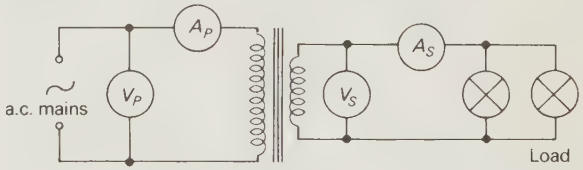


Fig. 32.31

Table 32.2

Primary voltage V_P	Primary current I_P (amps)	Secondary voltage V_S	Secondary current I_S (amps)
250	0.100	13.0	0
250	0.175	12.9	1.76
250	0.250	12.7	3.30
250	0.388	12.5	6.20
250	0.490	12.4	8.20

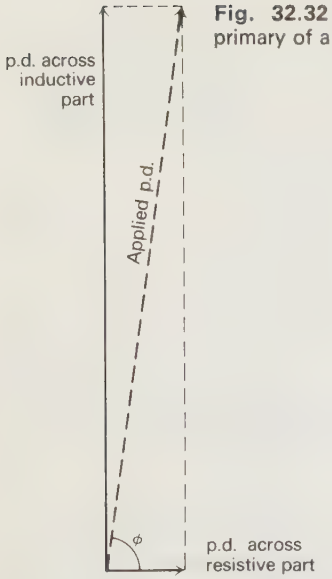


Fig. 32.32 Vector diagram for the primary of a transformer.

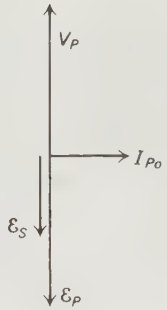


Fig. 32.33 Vector diagram for the secondary of a transformer (no load).

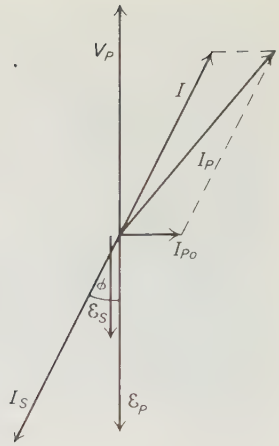


Fig. 32.34 Vector diagram for a transformer on load.

Table 32.2 needs to be studied carefully. Consider, first, the situation when there is no load. The primary current is 0.100 A. (This is shown not to be due to the presence of the voltmeter connected to the secondary simply by temporarily disconnecting that voltmeter.) The resistance of the primary winding was found to be 9.0 Ω, so if we apply the relation $V = IR$ to the primary, we should get, for the primary current:

$$I_P = \frac{250}{9.0} = 28 \text{ A}$$

approximately. But we found the primary current to be 0.100 A. Why the huge discrepancy? Obviously, because the primary has appreciable inductance (in fact, the inductance is approximately 8 H as you can verify, if you wish, from the data). The inductive reactance (see Section 32.2) is very much greater than the resistance, and so the vector diagram for the primary winding, relating the applied voltage with the p.d. across the inductive part and the p.d. across the resistive part (as in Fig. 32.13) would be as shown in Fig. 32.32. The phase-angle ϕ, which, of course, is the phase-difference between the applied p.d. and the current, is very nearly $\pi/2$ radians (90°), and the applied p.d. very nearly equal in magnitude to the p.d. across the inductive part of the winding.

Thus the primary behaves very nearly as a pure inductance, and, as explained in Section 32.2, no power is dissipated in a pure inductance. But, as we see from the table above, when a load is connected to the secondary and a current is drawn from it, the primary current increases in magnitude. Since power is now being fed from the secondary to the load, power must be going from the mains into the primary (if the principle of conservation of energy applies, as obviously it must). Let us consider the case in which a purely resistive load is connected to the secondary. The secondary current flowing will be in phase with the secondary voltage. This current creates a magnetic flux which tends to reduce the total flux in the core due to the primary current, and if this flux is reduced then the back e.m.f. in the primary is reduced, and if so then the primary current increases. To enable you to understand this better it is helpful to build up a vector diagram which includes voltages and currents. Figure 32.33 shows the situation when there is no load connected to the secondary. Assuming that the primary has negligible resistance, then the back e.m.f. $\mathcal{E}_P$ in the primary is equal and opposite to the applied

voltage, V_P. The primary current in this condition is I_{P_0}, lagging behind V_P with a phase-angle of $\pi/2$ radians. The e.m.f. induced in the secondary is $\mathscr{E}_S$, and has the same phase as the back e.m.f. in the primary (because they are both e.m.f.'s induced in coils by the same changing magnetic flux). Since there is no load connected to the secondary, no secondary current flows, and therefore the figure is a complete representation of the situation. But when a load is connected to the secondary, current flows. We call this current I_S. Since, even with a purely resistive load, the secondary circuit is partly inductive (simply because the secondary winding has self-inductance), the secondary current will not be in phase with the secondary voltage, but will lag behind it, with a phase-angle ϕ, as shown in Fig. 32.34. As has been said already, this current tends to reduce the magnetic flux in the core, and thus the primary current increases by an amount I. I is the current needed to produce a magnetic flux equal to the reverse flux produced by I_S. The current drawn from the supply is I_P, the vector sum of I and I_{P_0}.

In order to calculate correctly the power input to the primary and the power output from the secondary one has to take into account the phase-relationships of V_P, I_P, etc. To do this in detail is beyond the scope of this book, but these facts explain why you cannot simply take the values of voltage and current in Table 32.2 (which are r.m.s. values, of course) and multiply V_P by I_P to get the power input, and V_S by I_S to get the power output. As a rough rule, however, when a transformer is working with the load for which it is designed, one can use the following approximate relation:

$$\frac{I_p}{I_s} = \frac{V_s}{V_p} = \frac{n_s}{n_p},$$

where n_P and n_S are the numbers of turns in the primary and secondary windings, respectively.

As has been already explained, if more current is taken from the secondary, the primary current immediately increases. Presumably, there must be a limit to the strength of secondary current which can be taken. It can be seen in Table 32.2 that the effect of taking more current from the secondary is that the secondary voltage decreases slightly. This happens because the secondary winding has resistance, and of course heat is dissipated in this resistance. Likewise, some heat is dissipated in the primary when any current is flowing. This dissipation of heat in the windings is what electrical engineers often refer to as *copper losses* and these determine the maximum current which can be safely drawn from a transformer.

The laminated form of the transformer's magnetic core, mentioned earlier, is designed to minimize eddy-current losses. If the core were made of a single solid piece of magnetic alloy the current induced in it by the changing magnetic flux could be sufficient to cause excessive heating of the metal. The paths of these eddy currents are in planes perpendicular to the magnetic flux so that if the laminations (which, remember, are insulated from each other) are parallel to the magnetic flux, the eddy-current paths are interrupted by the layers of insulator. It is found that the rate of energy loss due to eddy-currents is roughly proportional to the square of the lamination thickness, and it is easy in practice to use laminations thin enough to reduce eddy-current loss to a negligible amount. Although this method of transformer core construction is adequate at low frequencies (especially the frequency of the a.c. mains, 50 Hz), at higher frequencies, in the megahertz region, for instance, magnetic alloys are not suitable and ferrites are used instead nowadays. Ferrites are ceramic materials, not metals, and have extremely high resistivities so that they can be classed as insulating materials, and eddy-current loss in them can be negligibly small at high frequencies.

At very high frequencies, around 100 MHz and above, it is often not necessary for a transformer to have a magnetic core at all. In such air-cored transformers the magnetic flux linkage between a pair of coils wound concentrically, one inside the other, can often be quite enough for efficient power-transfer from primary to secondary.

32.5 MOTORS AND ALTERNATING CURRENTS

By far the majority of the electric motors in use operate on alternating currents and most of them are

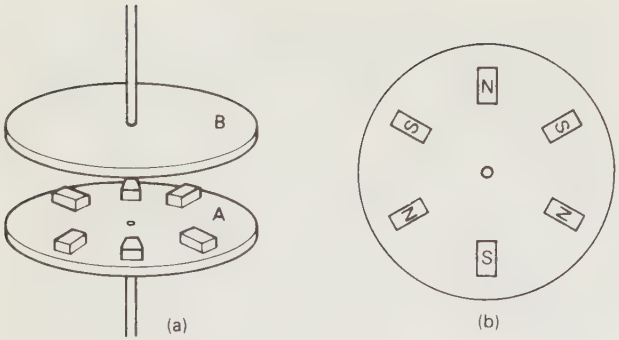


Fig. 32.35

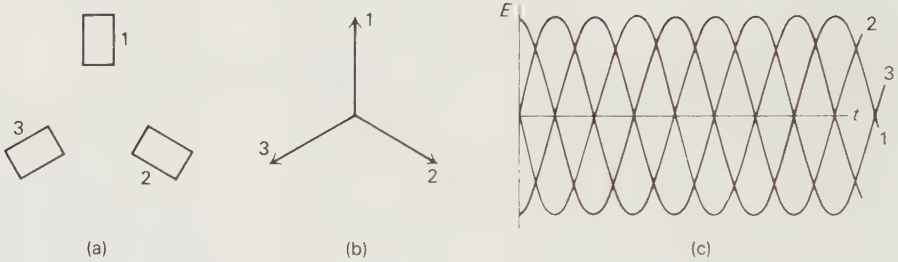


Fig. 32.36 Three phase a.c.

induction motors. They depend on the observation that a moving magnetic field can cause a nearby conductor to move as well. Consider the simple device shown in Fig. 32.35. Six ceramic magnets have their pole faces as shown (Fig. 32.35b) and are secured symmetrically to the wooden disc *A*. Above this a metal disc, *B*, is mounted. Both discs are free to rotate. When the magnet array is set in rotation it is found that the upper disc will follow suit. It will rotate in the same direction as the magnet disc but will constantly slip behind it. Evidently a torque acts on the upper, conducting disc and rotation results. The cause is the movement of the magnetic field nearby.

Such a simple observation immediately raises questions. How best can we produce a moving magnetic field? How can we bring the field and the conductor together so that the action is efficient? What is the best shape for the conductor? These must be answered if an efficient motor is to be produced.

One solution to the moving field problem which is also simple to relate to the experiment just described is to replace the three pairs of rotating permanent magnets by three stationary pairs of stationary coils (electromagnets). If now the coils are fed with alternating currents which are out of phase with one another the fields will rise and fall in such a way that the total field can be regarded as a rotating one. Three-phase electrical supplies provide exactly this with the three alternating voltages out of phase with one another by 120° (see Fig. 32.36). Figure 32.36 shows at 1, 2, and 3 three such voltages with a phase difference of 120°. Voltages, such as these, are produced commercially by rotating a very powerful electromagnet (the rotor) in the annular space left between three pairs of coils (the stator). An alternating voltage is induced in each of the three pairs of coils.

Let us return to the application of such supplies to the motor. The rotor of this kind of motor is usually of the squirrel-cage pattern (see Fig. 32.37). The

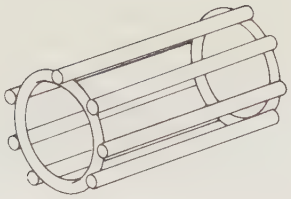


Fig. 32.37 Squirrel cage pattern rotor.

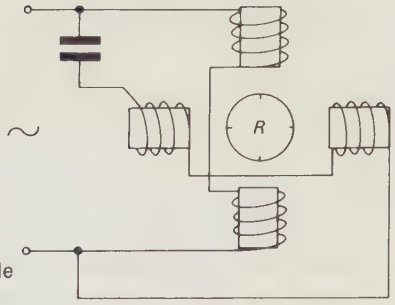


Fig. 32.38 A simple induction motor.

copper rods are welded together and are embedded in a core of laminated iron which increases the field.

These motors are usually powerful and their use is limited to factories and workshops where three-phase supplies are available. In the home, single-phase supplies are the rule. Other techniques must be used to produce the moving magnetic fields. One way is to use two pairs of coils only. Arranged in parallel, with a suitable capacitor in series with one coil (Fig. 32.38), out of phase currents will flow and an effective moving magnetic field will result. The conducting rotor R will experience a torque and rotate.

Either rotary motion or linear motion may result from the application of this observation and research on linear induction motors is being undertaken widely (Laithwaite, 1966).

PROBLEMS: Unit Six

6.1 a) A certain American electric toaster has a power-rating of 1 kW and is designed to work on a 110 V supply (as in the USA and Canada). It is brought by a visitor to Britain who fits it with a 13 A fused plug and connects it to the 240 V mains. Discuss whether the toaster can be expected to work satisfactorily or not. If not, what will be the consequences?

b) A British electric kettle, rated at 2.75 kW and designed for use on the British mains is taken to the USA and plugged into the mains there. Discuss whether the kettle will work satisfactorily in this situation.

c) Suppose you are arranging strings of electric lamps to provide lighting for a garden party at night. Calculate how many 240 V 60 W lamps you can safely connect to the 240 V mains using a plug containing a 13 A fuse.

d) A British fast-boiling kettle, rated at 2.75 kW, is found on a certain day to take 4 min 40 s to heat 2 litres of water (mass: 2 kg) from 15° C to boiling point (100° C). Assuming that the kettle actually is working at its rated power, calculate the percentage of electrical energy which is converted into heat to boil the water. (It takes 4200 J to raise the temperature of 1 kg of water by 1° C.)

6.2 a) Given three resistors, each 1.2 kΩ, what values of resistance can you obtain by using them in combination, either two of them, or all three together? You may connect them together any way you wish. Sketch each combination and label it with its overall resistance.

b) For a particular function, in a certain electronic circuit, a resistance of exactly 100 Ω is required. Suppose you have a carbon resistor marked with its nominal value of 100 Ω, but you have found that its actual resistance is a little too high, say 102 Ω. (This is perfectly possible if the resistor's tolerance is ±10%). You decide to reduce its resistance slightly by connecting another resistor, of high value, in parallel with it. Among your collection of resistors you have the following preferred values: 470 Ω, 820 Ω, 1 kΩ, 4.7 kΩ, 8.2 kΩ, 10 kΩ, 47 kΩ, 100 kΩ. Which one of these, in parallel with the 102 Ω resistor, will bring you nearest to the desired value? If you want the resistance to be within ±1% of 100 Ω, will this be satisfactory?

6.3 A certain bicycle dynamo has an internal resistance of 0.1 Ω. When running at normal speed it lights the bicycle's front and rear lamps, each containing a bulb rated at 6 V 6 W, to normal brightness. The lamps are connected in parallel to the dynamo.

a) Assuming the p.d. across the dynamo terminals is 6 V exactly, calculate the e.m.f. of the dynamo.

b) Of the electrical energy generated by the dynamo, what percentage is dissipated as heat in its internal resistance?

6.4 The circuit shown (Fig. P6.1) contains a transistor. The value of the resistor R is to be chosen to provide the correct working conditions for the transistor: it is called the base bias resistor since it is connected to B, the base of

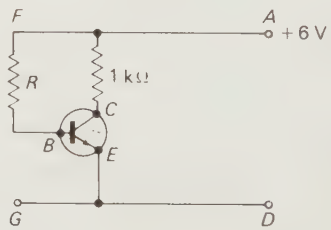


Fig. P6.1

the transistor. The supply voltage is 6 V; the negative rail *GD* is at a potential of zero and the positive rail *FA* at +6.0 V. The p.d. between *C* and *E* (the collector and emitter of the transistor) is 3.0 V. The transistor's current amplification factor is 50: that is to say, the current flowing from the positive rail into the collector, C, is 50 times the current flowing into the base, *B*. The potential difference between *B* and *E* is small enough to be neglected.

From this information, find the current flowing into the transistor at *B* (the base current) and hence calculate a suitable value for *R*.

6.5 Capacitors of small values in a micro-circuit are often made of a thin layer of gold deposited on the flat substrate, a thin layer of silicon monoxide dielectric deposited on the gold, and another thin layer of gold. Assuming that this construction constitutes a parallel-plate capacitor, estimate the capacitance per square millimetre of surface which can be obtained, if the relative permittivity of silicon monoxide is 4.5 and its thickness is 400 nm.

For higher values of capacitance TaO_5 can be used as a dielectric: its relative permittivity is about 22. Assuming that a 400 nm thickness of this dielectric is used, deduce whether or not it would be practicable to make a 1 μF capacitor by this technique in a micro-circuit.

6.6 In various electronic circuits one requires a voltage which increases at a steady rate, for instance, in an analogue computer where voltages are used to represent numerical quantities. A way of achieving this is to use a large capacitor in conjuction with a large resistance.

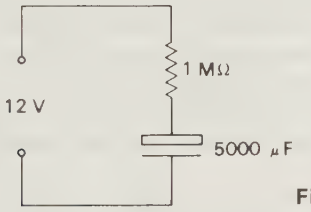


Fig. P6.2

Suppose we have a 5000 μF capacitor connected in series with a 1 MΩ resistor, and that this combination is connected to a 12 V d.c. supply. Initially the capacitor is uncharged (Fig. P6.2).

a) Calculate the initial current flowing in the resistor, when the p.d. across the capacitor is zero.
b) Assuming that this current gives the rate of flow of charge into the capacitor for the first second, by how much will the p.d. across the capacitor increase in the first second?
c) Using the result of (b), write down the p.d. across the resistor after the first second.
d) By what percentage does the current in the resistor after one second differ from the initial current? This will also be the percentage change in the rate of rise of voltage across the capacitor.
e) Estimate (roughly) how long the capacitor can go on being charged in this way before the rate of rise of voltage differs by 1% from its initial value.

6.7 In an electronic flash-gun, for use with a camera, a large capacitor is charged to a high voltage and then discharged through a gas discharge lamp, usually containing Xenon, which gives a very intense flash of white light of extremely short duration.

In a typical flash-gun, a 500 μF capacitor is charged up to 400 V. After the flash has been fired the capacitor is not completely discharged: there is a p.d. across it, usually about 50 V. Use this information to calculate how much energy is obtained from the capacitor in this process.

6.8 A semiconductor diode is one example of what is sometimes called a *non-linear* device, because its voltage-current characteristic graph is not a straight line, as it would be for a device which obeys Ohm's law.

Suppose we have a diode connected in series with a resistor to a d.c. voltage supply. How can one predict the current flowing in the diode and resistor? A graphical method is described here. Calling the supply voltage V_S, the voltage across the diode V_D, then

$$V_S = V_D + IR$$

or

$$V_D = V_S - IR.$$

The voltage-current characteristic for a germanium diode, Fig. 29.1, will be used here. This characteristic gives I as a function of V_D. On the same graph one can plot the line representing the function $V_S - IR$, if V_S and R are known, and where this line intersects the curve for the diode, $V_D = V_S - IR$, and the value of I at the point of intersection gives the current flowing in the diode and in R under these conditions.

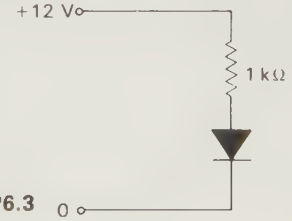


Fig. P6.3

Use Fig. 29.1 to find the current in the situation depicted in Fig. P6.3.

6.9 The current, in ampères, through a certain type of non-linear resistor is given by $I = 0.20\ V^3$ where V is the p.d. across the resistor. This resistor is connected in series with an ordinary carbon resistor (which obeys Ohm's law, that is, it is linear) to a constant d.c. voltage source of 6.0 V. What value of resistance should the carbon resistor have so that the current in the circuit is 0.40 A?

6.10 The circuit shown in Fig. P6.4 contains an ordinary tungsten filament lamp in series with a thermistor. It is observed that, when the switch S is closed, the lamp glows dimly at first then slowly brightens, until it reaches constant, nearly full, brightness. Knowing how the resistance of the two devices (see Chapters 25 and 29) depends upon temperature, explain the observed behaviour.

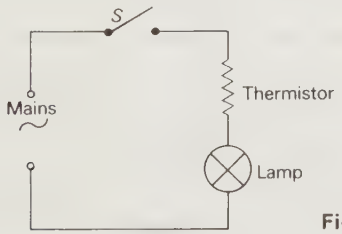


Fig. P6.4

Why might it be desirable to have a thermistor connected in series with a lamp in some cases, rather than connect the lamp directly to the mains?

6.11 The following tests were made on a "black box," with results as described. Try to deduce from these what might be inside the box. You can be assured that every component inside the box obeys Ohm's law (Fig. P6.5).

a) When a voltage V is applied to the terminals AB, the voltage between the terminals C and D is always found to be equal to V.
b) When a voltage V is applied to C and D, the voltage appearing between A and B is always found to be $\frac{3}{4}V$.
c) Whatever voltage is applied between the terminals A and C, the voltage between B and D is zero.
d) The resistance between terminals C and D is 800 Ω.
e) The resistance between A and B is 600 Ω.



Fig. P6.5

6.12 The spacing between adjacent conductors on the 132 kV Grid is 24 feet—say 8 metres, as the nearest metric equivalent value. The maximum allowable current in a conductor of diameter just under an inch is 400 A. Estimate the force between two conductors having this spacing and current, of length 300 m. The mass of this length of conductor is about 400 kg. Do you think that the magnetic forces need to be allowed for?

6.13 A shunt-wound electric motor is one in which the field is provided by an electromagnet, and the windings of the electromagnet are connected in parallel with the armature. It can be represented symbolically as shown in Fig. P6.6.

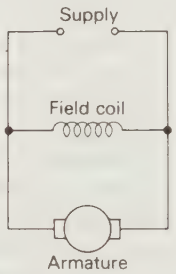


Fig. P6.6

Suppose such a motor has the following specifications —resistance of armature winding, 1 Ω; resistance of field winding, 100 Ω; current taken from 100 V d.c. supply when running at full speed, 10 A.

We can find the power which is converted from electrical into mechanical form as follows.

a) Calculate the current in the field winding.
b) Hence find the current in the armature winding.
c) Deduce the back e.m.f. in the armature.
d) Hence calculate the mechanical power of the motor.
e) Find the overall efficiency of the motor, that is, the ratio

$$\frac{\text{mechanical power output}}{\text{total electrical power input}}$$

and express it as a percentage.

6.14 It is possible to run a 110 V lamp (an ordinary, tungsten filament lamp) from the 240 V a.c. mains by using a capacitor in series with it as a voltage-dropper, instead of using a resistor. Suppose the lamp is rated at 60 W.

a) Calculate the value of the capacitance required, and the r.m.s. voltage across it when in use.

b) If a resistor were used instead of the capacitor, what value of resistance would be needed?
c) What power would be dissipated in the resistor?
d) What would be the advantage of using a capacitor instead of a resistor?

6.15 In the circuit shown in Fig. P6.7 the neon lamp is found to flash at regular intervals. A typical neon lamp does not conduct any current until the voltage across it rises to 110 V, the striking voltage, and then it keeps glowing until the voltage across it falls to 80 V. When it is glowing its resistance can be considered to be negligibly small in comparison with R.

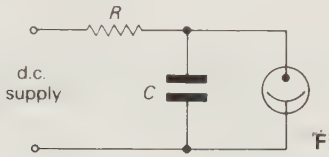


Fig. P6.7

a) Explain why the neon flashes repeatedly.
b) What would happen to the rate of flashing if R were increased?
c) What would happen to the rate of flashing if C were increased?

6.16 The tuned circuit of a radio receiver often consists of a fixed-value inductor connected in parallel with a variable capacitor. Inevitably there is stray capacitance in the circuit which is effectively an extra capacitance connected in parallel with the variable capacitor. In a typical tuned circuit the stray capacitance amounts to about 50 pF. Suppose such a circuit has an inductance of 120 μH, and is to be tunable over a frequency range from 0.6 MHz to 1.5 MHz (the medium wave band). What must be the maximum and minimum values of the variable capacitor?

6.17 A thin wire is stretched tightly between a pair of supports. When plucked it is found to have a natural frequency of oscillation of 400 Hz. A permanent horseshoe magnet is placed so that the poles are either side of the wire and near to its centre. An alternating current from a signal generator is passed down the wire.

a) Indicate on a diagram the direction in which the wire will oscillate relative to its own orientation and the direction of the magnetic field.
b) Describe, using graphs where appropriate, the way you think the amplitude of these oscillations will change as the frequency of the signal generator is varied from 200 to 600 Hz.
c) What would be the effect of tightening the wire?
d) With the wire still at its original tension, what, qualitatively, would you observe if the frequency of the applied a.c. was further raised from 600 to 1500 Hz?

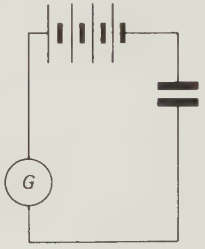


Fig. P6.8

6.18 The circuit shown in Fig. P6.8 might be used to study the storage of electric charge on a pair of parallel plates. The meter is a centre zero galvanometer. Describe with a short explanation in each case what would happen to the meter needle if

a) the top plate is moved sideways so that the overlap is reduced but the separation remains constant,
b) the top plate is returned to its original position and the plates pushed closer together,
c) the p.d. applied to the plates is increased after they have been restored to their original positions,
d) a sheet of polythene is slipped between the plates.

BIBLIOGRAPHY: Unit Six

Amos, S. W. (1965). *Principles of Transistor Circuits* (3rd ed.), Iliffe.
A sound introduction to transistors and how they are used, suitable for readers who are studying physics to A-level.

(Central Office of Information) (1970). *Britain 1970*, H.M.S.O.
This yearbook contains statistics of the electricity supply industry.

Feather, N. (1968). *Electricity and Matter*, Edinburgh University Press.
A scholarly treatise on the historical development of the concepts of electricity.

Gee, B., (1969) "G. S. Ohm", *Physics Education*, **4**, 106–113.

Goodier, J., and Meynell, J. W. (1964). *Electricity*, Mills and Boon.
This book, suitable for A-level students, contains a particularly thorough treatment of electromagnetism, especially Chapter 8: "Dynamo-electric Machinery."

Handel, S. (1967). *The Electronic Revolution*, Pelican.
A book, written for the layman, on the impact of new developments in the field of electronics.

Hawker, J. P., and Reddihough, J. A. (1967). *Electronics Pocket Book* (2nd ed.), Newnes-Butterworth.
A handy, elementary compendium of information on electronics well-suited to the needs of the amateur.

Hughes, L. E. C., and Holland, F. W. (Editors) (1967). *Electronic Engineers' Reference Book*, Iliffe.
A copious and compact source of information on all aspects of electronics. Further information on microelectronics can be found in Section 365 "Microelectronics." This is a professional engineer's handbook.

Hughes, J., and Johnston, T. M. (1970). *Using Semiconductors*, Heinemann.
An introduction to the practical use of diodes, transistors, thermistors, light-dependent resistors, and other semiconductor devices.

Kaye, G. W. C., and Laby, T. H. (1959). *Tables of Physical and Chemical Constants* (12th ed.), Longmans.
This is the source of the data on resistivities of materials temperature coefficients of resistivity, and dielectric constants.

Laithwaite, E. R. (1966) "Propulsion without wheels." English Universities Press.

Magie, W. F. (1963). *A Source Book in Physics*, Havard University Press.
Excerpts from original writings. In the section "Magnetism and Electricity" there are chapters on: Coulomb, Volta, Oersted, Arago, Ampère, Ohm, Faraday, Lenz, Henry, Hertz.

Matthias, B. T. (1970). *Physics Today*. **24**, 23–27.

Tomalin, M. (1968). *The Energy Men*, (Zenith Books) Hodder and Stoughton.
A survey, written for the layman, of sources of energy, including electrical energy.

Williams, L. Pearce (1965). *Michael Faraday*, Chapman and Hall.
A detailed biography. Chapter 4: "The Discovery of Electromagnetic Induction" and Chapter 5: "The Nature of Electricity (I)" are especially relevant.

UNIT SEVEN
Forces, Fields, and Radiation

33 Fields of Force
34 The Inverse Square Law
35 The Electric Field
36 The Gravitational Field
37 The Magnetic Field
38 Potential Energy in Fields of Force
39 Potential in the Electric Field
40 Electromagnetic Induction
41 Maxwell's Model of the Electromagnetic Field

Chapter 33
FIELDS OF FORCE

33.1 Introduction
33.2 Fields of force
33.3 The gravitational field
33.4 The electrostatic field
33.5 The magnetic field
33.6 Field strength directions
33.7 Mapping other fields
33.8 Lines of force
33.9 Faraday and lines of force
33.10 The magnetic field and moving charges
33.11 A unit for current
33.12 The behaviour of bodies under the influence of a field of force

33.1 INTRODUCTION

When, in Unit 2, we looked at the way forces set things into motion, or failed to move them, or changed the direction in which they moved, we did not consider how these forces were produced. They were just "pushes" or "pulls"—the result usually of a direct contact between something that moved and whatever was trying to produce the movement.

By looking at the effect these forces have when they act, we learn to define a force as something "trying to change motion." This is an important generalization of many happenings in the physical world. Once this generalization has been made, we are confident we can always recognize when a force acts. So when we bring a magnet up towards an iron nail, and we see the nail move even when the magnet is some distance away, we say that the magnet exerts a *force* on the nail. We let go a plate we are carrying and it crashes to the floor. The motion of the plate has been changed. We say that the Earth exerts a *force* on the plate.

We do not always have to see a change in motion to assert confidently that forces are acting. We demonstrate our knowledge of physics by rubbing a blown-up balloon on a jersey and sticking the balloon to the ceiling. We think, ordinarily, it ought to fall—because of the pull of the Earth on it. So we say another force must pull the balloon towards the ceiling. We learn eventually to call this an electric force.

All these forces, however, seem rather mysterious. They are not like the pushes and pulls we use to generalize our idea of a force. They are not forces of contact. The thing which causes the force is not in direct contact with the thing which moves. The force acts—that is undeniable—but nothing seems to be pushing or pulling. This mystified scientists for a long time; they felt that something ought to be visibly pushing or pulling and went to great lengths to prove that invisible mechanisms were involved.

J. C. Maxwell was one of the first scientists to point out that far from being mysterious, these sorts of forces were probably the only ones which exist. As an example, think of the following: from a cricketer's point of view, a bat hitting a ball involves a force of

contact. From the ball's point of view, the force is also one of direct contact—direct contact between ball and bat. But from the point of view of an atom on the surface of the cricket ball, it is not direct contact at all. The atom's outer electrons experience a repulsive electric force from the outer electrons of an atom on the surface of the bat as this approaches. This electric force pushes the ball-atom away. Other forces from neighbouring atoms help keep all the atoms in their same relative position within the ball. What has become of the force of contact?

Today, rather than try to explain how these forces act, as soon as we meet them, we accept their existence and explore their properties. Newton worked in this way when he made his important discoveries about gravitational forces. He was not the first to give serious attention to forces such as these, but his approach was much the most successful. You must agree with this, if success is measured by the ends to which we can put his laws, in trips to the Moon and launching Earth satellites. Yet science is no nearer explaining the force of gravity in terms of other fundamental properties. Before turning to an exploration of the properties of forces which act at a distance let us recall what Maxwell said, about two hundred years after Newton's work:

> "But if we leave out of account for the present the development of the ideas of science, and confine our attention to the extension of its boundaries, we shall see that it was most essential that Newton's method should be extended to every branch of science to which it was applicable—that we should investigate the forces with which bodies act on each other in the first place, before attempting to explain *how* that force is transmitted. No men could be better fitted to apply themselves exclusively to the first part of the problem, than those who considered the second part quite unnecessary."
>
> [*Reprinted from* Proc. Royal Inst. **7**, 1873.]

33.2 FIELDS OF FORCE

The gravitational, electrostatic, and magnetic forces with which we are concerned in this Unit are detectable in extensive regions of space around their source. The region around a source of one of these forces in which the force can be detected is called a *force field.* Science uses "field" with much the same understanding that it is used in everyday speech. When we talk of a field of grass, we refer to a region in space where grass is to be found. Similarly, when we talk of an electric field, we refer to a region in space in which an electric force can be found. (You will notice the scientist's shorthand here—we ought really to say "electric force field" and not just "electric field." The word "force" is invariably omitted and must be understood as being meant.)

In this Unit we shall explore some of the properties of gravitational, magnetic and electrostatic fields. We shall try to see the ways in which different force fields hold properties in common and the ways in which they differ from each other. We shall find it useful to develop models of these fields which will help us to understand their properties. Although we shall confine our attention to these three fields, science now recognizes two more of at least equal importance. These are the two forces which bind the nuclear particles together into an atomic nucleus. How little is yet understood about them is probably indicated by their names—*the force of strong interaction* which binds neutrons and protons together (amongst other things) and *the force of weak interaction* which, amongst other things, is responsible for beta decay.

It is probably true to say that less is understood about the quantitative properties of these two force fields than Newton understood about gravity, so having mentioned them, we shall not discuss them again.

One final point must be cleared up before we undertake a detailed survey of force fields. Forces within these fields occur as action/reaction pairs. The force of attraction of the Earth for a ball is, we believe, matched by an equal force of attraction of the ball for the Earth. The force of attraction of the charged plates in a Millikan cell for an oil drop is, we believe, matched by an equal force of attraction of the oil drop for the charged plates.

However, in almost all particular cases, it is the force on only one of these pairs of bodies which is of



Fig. 33.1 A freely falling ball. (Photograph by E.J.W.)

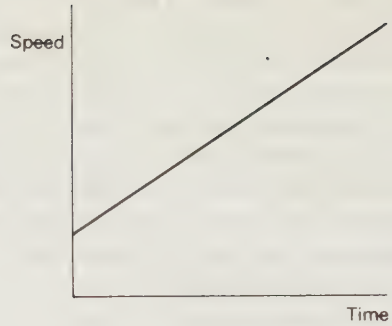


Fig. 33.2

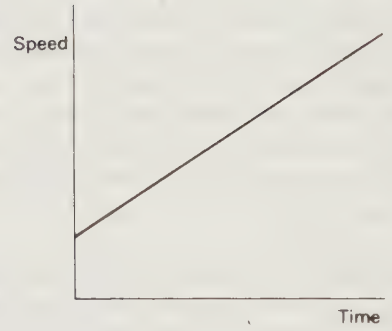


Fig. 33.3

interest. In Millikan's experiment the oil drop moves, so we are only interested in the force the plates exert on the oil drop. Thus the plates are looked upon as the source of the field which acts on the drop as receiver of this force. In this way a complex particular situation can be resolved into two parts:

a) the general case of the force field provided by the plates, and
b) the particular case of the force this field exerts on a particular drop.

33.3 THE GRAVITATIONAL FIELD

The mutual force of attraction between the Earth and any other body on it is now recognized to be only a particular example of an attractive force which exists between any two masses. This force is always one of attraction, and was shown by Newton to be responsible for the motions of the planets acted upon by the gravitational influence of the sun. Cavendish showed that a measurable force was exerted between two spheres, one of a few grammes mass, the other of a few kilogrammes, and lent support to Newton's hypothesis as to the universal nature of the gravitational force.

The gravitational field near the Earth's surface is of convenient size and accessibility for some initial experimental observations. One convenient method of starting this investigation is to measure the change in motion of a ball as it falls freely in a gravitational field. Figure 33.1 shows a freely-falling ball photographed stroboscopically using the technique described in Unit 2. A speed-time graph (Fig. 33.2) can be plotted for the motion.

The graph is straight and it follows that the ball falls with constant acceleration. The gravitational

force on the ball and the acceleration it acquires can be linked through Newton's second law of motion:

$$F_G = m_i a, \tag{33.1}$$

where F_G is the gravitational force on the ball, a its acceleration and m_i its inertial mass. Since the ball falls with constant acceleration, it follows that the force of gravity is constant over the region of the ball's fall.

We would find the acceleration unchanged if the experiment were repeated in any other part of the laboratory. The ball would also fall in the same direction so the region of the laboratory is a uniform force field.

It is important to note that the uniformity is not simply one of size (or magnitude), but of direction as well. Force fields are vector fields. The forces which make up the field have direction as well as magnitude.

Now let us imagine that we repeat the experiment with a different mass falling under gravity. The stroboscopic photograph (Fig. 33.1) was made using a 2.5 cm diameter ball. The graph (Fig. 33.3) was plotted from data obtained from a falling ball of 7.5 cm diameter. The two masses thus differed by a factor of 27. The graph is again straight and (within the limits of error set by the experiment) its slope is the same as the slope of the graph of Fig. 33.2. This ball has fallen with the same acceleration as the smaller ball, despite the large difference in mass. Thus for the two masses

$$F'_G = m'_i a,$$
$$F''_G = m''_i a.$$

Since a does not change

$$F_G \propto m_i.$$

The force of gravity is proportional to the inertial mass of a body. This experiment gives the reason why we can find the mass of a body by weighing it; a far more convenient process than by pulling it along with a recognized force and measuring the acceleration.

The force of gravity on any body within the region of the experiment is given by the formula

$$F_G = g \times m_i, \tag{33.2}$$

where g is a constant, the force of gravity per unit inertial mass: g gives a measure of the strength of the force field which is independent of the particular mass on which the force acts. It is called the *gravitational field strength* and is defined by the equation

$$g = \frac{F_G}{m}. \tag{33.3}$$

It should be noted here and in what follows the symbol g is used to denote any gravitational field strength and not just the particular gravitational field close to the surface of the earth. However, for all regions close to the Earth itself we find that, approximately,

$$g = 9.8 \text{ N kg}^{-1},$$

g is also the acceleration of free fall which, in the region where the field strength is 9.8 N kg^{-1}, is 9.8 m s^{-2}. (See Problem 2.12.)

33.4 THE ELECTROSTATIC FIELD

It is common knowledge that certain materials may become charged electrically, for the effect is readily observed in nylon clothing, plastics and man-made fibre carpets, for example. At first sight, it seems that such effects are associated with a class of substance which we have learned to call insulating as other materials, metals, for example, seem always to be electrically neutral. However, we may recall that those materials which do show acquired charges are neutral under most circumstances. This neutrality of matter is recognized to be the result of an equality in the number of positively and negatively charged particles which make it up.

Elementary experiments in which electrically neutral bodies are rubbed together reveal the existence of two (and no more than two) types of charge, and also that forces of attraction and of repulsion exist between charged bodies. Such charges are revealed on insulators merely because the property by which we recognize an insulator is that electric charges do not flow at all readily through them.

It is of interest, in this respect, to place two discharged insulating rods of different materials (e.g., polythene and acetate) into a conducting can which is mounted on the input terminal of a d.c. electrometer/

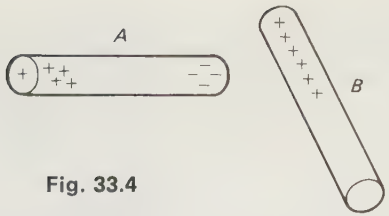


Fig. 33.4

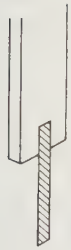


Fig. 33.5 A foil indicator.

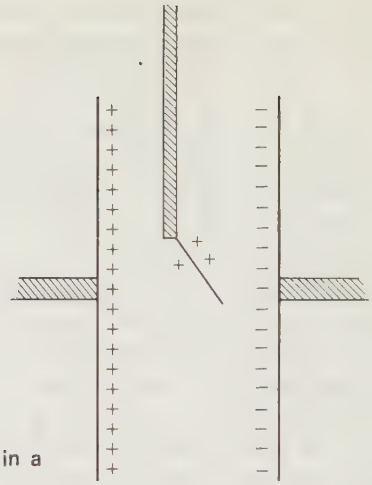


Fig. 33.6 A foil indicator in a uniform electric field.

amplifier (see Section 35.1). The electrometer should be connected to a suitable centre-zero galvanometer. If the rods are truly discharged (and this may be achieved by passing them quickly through a gas flame which is rich in ions) no charge will be indicated by the instrument. If the rods are removed from the can and the ends rubbed together for a moment, they will become charged. Placed carefully into the can one at a time, one rod will show a positive charge and the other a negative one. Place them together into the can and, once again, no charge will be indicated. The charges resulting from the process of rubbing the rods together are evidently equal and opposite in effect.

If a charged rod is brought near to an insulated conductor a re-distribution of charges will occur in that conductor (see Fig. 33.4). Naturally the negative charge induced in the right-hand end of the conductor *A* by the charge on the rod *B* will be attracted towards *B* whilst the positive charge at the remote end will be repelled by the similar charge on *B*. The net result is that the body in which charges are induced and the body carrying the inducing charge are attracted to one another. This is the mechanism which operates when a charged rod attracts a light object towards it.

This re-distribution of the charges in a body is called *polarization* and, although occurring in both conductors and insulators, is very marked in the former. As we shall see this makes measurements of electric field strengths difficult to perform.

A simple indicator of field strength can, however, utilize the effect. A very thin, flexible metal foil probe about 2 cm long and 0.5 cm wide is mounted on the end of a long insulating rod. (See Fig. 33.5.) If this probe is placed between two vertical charged plates as shown in Fig. 33.6, polarization of charge will occur. If now the probe is touched against one of the plates it will acquire an overall charge (positive in Fig. 33.6) and be attracted to the appropriate plate. The force exerted on the charge by the field and the force developed by the flexing of the foil will result in the foil bending through a definite amount. This provides a rough indication of the field strength in the space between the plates.

If the indicator is now moved around in this space between the plates it will be seen that the force must be the same at all points within the space bounded by the plates, except near to the edges. The space where the foil indicator does indicate a constant force contains a uniform electrostatic field.

To measure directly the size of the force field between the plates it is necessary to introduce a charged object into the space. Such a charge is often

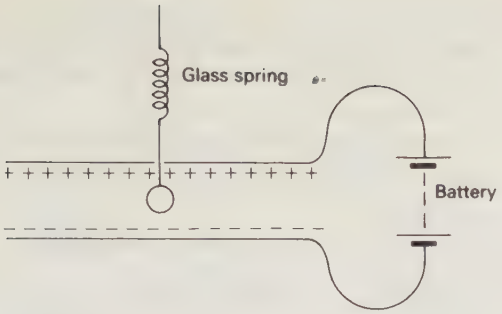


Fig. 33.7 A light, charged conductor in a uniform field.

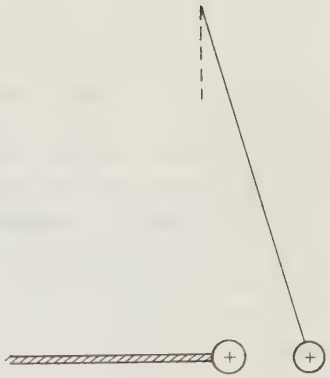


Fig. 33.8

referred to as a *test charge*. Unfortunately as soon as we do this with a test charge big enough to give measurable forces, we find that the additional charges induced in the plates themselves, by the test charge, interfere severely with any measurements we may wish to make. For example, a small charged sphere (Fig. 33.7) supported by the weak spring shown will induce polarization of the charge in both the upper and the lower plates. This changes the total distribution of charge and therefore the very force which we had hoped to measure. So, for the moment we will consider a different experiment for measuring the force experienced by a test charge in an electrostatic field.

The sphere on the insulated support (Fig. 33.8) is given a charge. This is conveniently done by connecting it momentarily to one terminal of a power pack providing several kilovolts. The light test sphere made of expanded polystyrene coated with a conducting paint, is suspended from a fine nylon thread and given a charge of the same sign as that on the fixed sphere. The charges repel one another and the force can be estimated from the inclination of the nylon thread. The force on the test charge, as we shall see when we consider this experiment in more detail in Section 35.1, is proportional to the magnitude of the test charge provided that the distance between the fixed charge and the test charge does not change during the experiment. This is important because, in this case, we are not dealing with a uniform field.

Problem 33.1

a) How could the charge on the test charge carrying sphere be changed by known amounts even if the actual value of the charge could not be measured?

b) What sources of error would one have to take special care about in the experiment just described?

We can write the result just quoted as

$$F_E \propto q,$$

where F_E represents the electrostatic force acting on the test charge q. If we compare this result with the case of the gravitational field (where $F_g \propto m_i$) we may write

$$F_E = Eq. \tag{33.4}$$

E is a measure of the electric field strength and is independent of the particular charge on which it acts. It is defined by the equation

$$E = \frac{F_E}{q}. \tag{33.5}$$

The units in which the electric field strength or intensity is quoted will be

$$\text{newton coulomb}^{-1}.$$

The units of g, it will be recalled, were newton kilogramme^{-1}.

33.5 THE MAGNETIC FIELD

The history of ancient Greece records the discovery of some ores of iron which could attract to them other pieces of iron. The force fields surrounding these ores are called *magnetic* fields. In the nineteenth century, Oersted showed that an electric current was also surrounded by a magnetic field. The field only exists when an electric current flows through the wire. In the case of the magnetized iron, the field is always there, so these pieces of iron ore are called permanent magnets.

A small pivoted permanent magnet always aligns itself approximately parallel to the Earth's north-south axis in the absence of any other source of magnetic field. From this, we conclude that the Earth has its own magnetic field and behaves like a large permanent magnet. A small magnet designed to reveal this field is called a compass needle.

Compass needles will be deflected by other magnetic fields and they are a very useful means of detecting such fields. The Earth's magnetic field is very weak, and it is not too difficult to make the fields we want to investigate strong enough to allow us to ignore the Earth's field, or even to cancel it out.

The work of Oersted and Ampère in the early nineteenth century revealed that an electric current could exert a force on a nearby compass needle and also on a second current carrying conductor. Magnetic force fields exist in the space around all current carrying conductors as well as in the vicinity of permanent magnets. Indeed there is good evidence for believing that the magnetism of permanent magnets is produced by currents of electric charge within the atoms themselves. Magnetic fields are fundamental properties of moving electric charges or currents and measurements of the strength of such fields are usually related to the forces between current-carrying wires rather than the forces between permanent magnets.

The basic tool for this purpose is the current balance. We have already seen in Section 30.1 how experiments with such balances lead to the conclusion that

$$F_M = BIl, \qquad (33.6)$$

where F_M is the force experienced by a conductor of length l carrying a current I normal to a field of magnetic flux density B. In the international system B is measured in tesla (N $A^{-1}m^{-1}$).

We may tabulate the main features of the measurement of the field strengths as in Table 33.1.

Table 33.1

Force field	Field strength symbol	Unit
Gravitational	g (Earth's field)	newton kilogramme^{-1}
Electrostatic	E	newton coulomb^{-1}
Magnetic	B	newton ampere^{-1} metre^{-1} (or tesla)

33.6 FIELD STRENGTH DIRECTIONS

a) Electric and gravitational

In the last section we considered how gravitational, electrostatic, and magnetic field strengths are defined and measured. Further experiment will give us some simple and important laws which relate these field strengths to their sources and the distances within the field. But to know the field strength is not enough; to be able to use these laws fully we need to know the direction in which the forces act as well.

Let us first consider *electrostatic* and *gravitational* fields, as the relationship between the forces which act and the field strengths is similar. We shall find that these two fields behave in such similar ways that if we know the way one behaves we can immediately predict how the other will behave.

In both cases

$$\begin{array}{ccccc} \text{force} & = & \text{field strength} & \times & \left(\begin{array}{c}\text{mass}\\ \text{or}\\ \text{charge}\end{array}\right) \\ \uparrow & & \uparrow & & \uparrow \\ \text{vector} & & \text{vector} & & \text{scalar} \\ \text{quantity} & & \text{quantity} & & \text{quantity} \end{array}$$

The direction of the field at any point is the direction of the force it would exert on a charge or a mass at that point.



Fig. 33.9

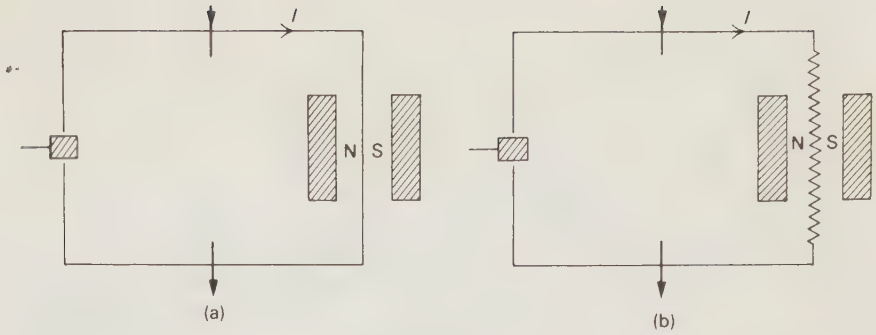


Fig. 33.10

b) Magnetic

The magnetic field strength is unlike electrostatic and gravitational field strengths in the way it is related to the magnetic force:

force	=	magnetic flux density	×	product of current and length (current element)
↑ vector quantity		↑ vector quantity		↑ vector quantity

The current element has direction as well as size, whereas mass and charge are scalar quantities.

We said, in introducing magnetic fields, that the study of these developed historically from the fields around permanent magnets. A compass needle was (and still is) used to detect them and the direction of the field at any point is said to be the direction indicated by the North pole of the compass. Figure 33.9 shows how the field is revealed in the case of a straight wire, carrying a current into the paper.

Alternatively, we may scatter iron filings over a sheet of paper around the wire. Each iron filing becomes a small magnet and points along the direction of the field.

But to measure B we measure the force on a current element, Il. This force is perpendicular to the direction of the field lines plotted by the compass. Is this just a historical accident and a consequence of the development of the field model in the nineteenth century?

We recall (Section 30.1) that the full expression for the force F_M exerted on a wire of length l, carrying a current I held in a field B is

$$F_M = B(Il) \sin \theta, \qquad (33.7)$$

where θ is the angle between the conductor and the field direction. Since the magnetic flux density B must be descriptive of the field and independent of the conductor on which it acts, B must not only be independent of the magnitude of the product Il but also of its direction. This demands that the direction of the magnetic field should be in a plane which is perpendicular to the direction of the magnetic force. We must continue to define the direction of B as being that in which a freely pivoted compass needle would point.

Problem 33.2 Two current balances of the type described in Section 30.1 and of the same overall dimensions are constructed. One has a straight, rigid test conductor and the other a rigid test conductor which is bent into zig-zag form. The test conductors are placed in turn in the same magnetic field B and the same current I is passed (Fig. 33.10). The forces are found to be equal. For the straight conductor the force is BIl. What value of l (the length of the con-

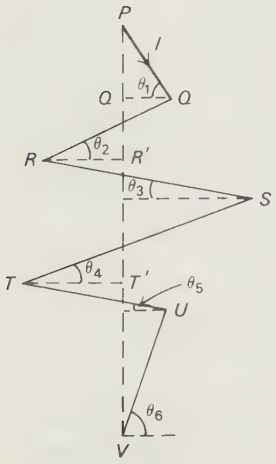


Fig. 33.11

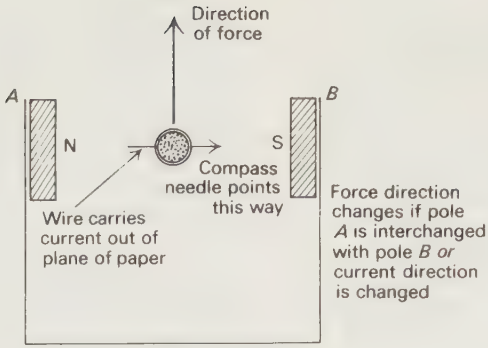
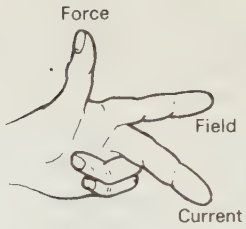


Fig. 33.12 Current field, and force. The force changes direction if pole *A* is interchanged with pole *B* or if the current direction is changed.



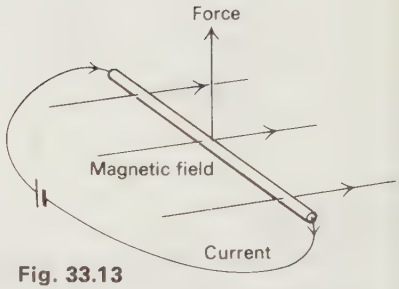


Fig. 33.13

ductor in the field) must be used when the zig-zag conductor is used?

What is the effective length of the zig-zag conductor?

Solution The zig-zag conductor may be represented as shown in Fig. 33.11. The products *Il* may be resolved into components normal to the field and parallel to the field. Use the figure to show that the components QQ', RR', SS', TT', and UU' total zero whilst the components normal to the field total *Il*.

When a current balance is placed in a magnetic field, and the current element is perpendicular to the direction of the magnetic field, the force on the balance may be up or down depending on the direction of the current in the balance and which way round we place the magnet. (See Fig. 33.12.) When compass needles were the sole means of finding field directions, the field was said to point in the direction of the force on the North-seeking end. While we retain that convention today, we can re-state it in terms of the force on a current element, as follows:

The magnetic field is perpendicular to the direction of the force on a current element and perpendicular to the direction of the current element when that force is a maximum. The magnetic field is in the direction that the first finger of the left hand would point if the thumb and second finger, extended mutually at right-angles, point in the direction of the force and the current respectively. (See Fig. 33.13.)

This is commonly known as Fleming's left hand rule and it gives us the conventional direction of the field, if the conventional direction of electric current is taken to flow from the positive to the negative terminals of a battery.

There are other rules for finding the direction of the magnetic field associated with an electric current which eliminate the need to find the direction of a force on a current element. One of the most useful is that commonly known as Maxwell's corkscrew rule, used for finding the direction of the field around a long-straight current-carrying wire. The field lines are circles concentric with the wire, and the field direction is that in which a right-handed screw would rotate if

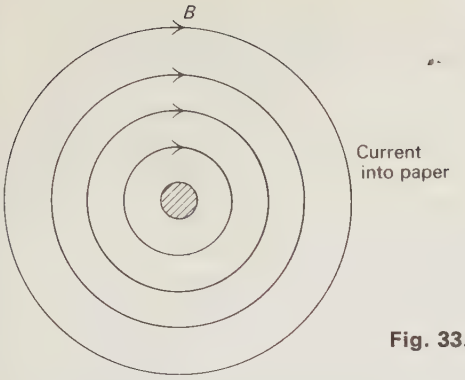


Fig. 33.14

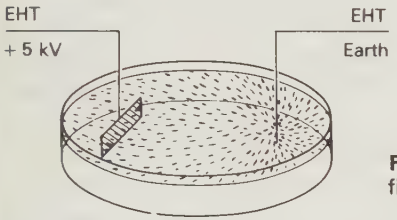


Fig. 33.15 Displaying electric fields.

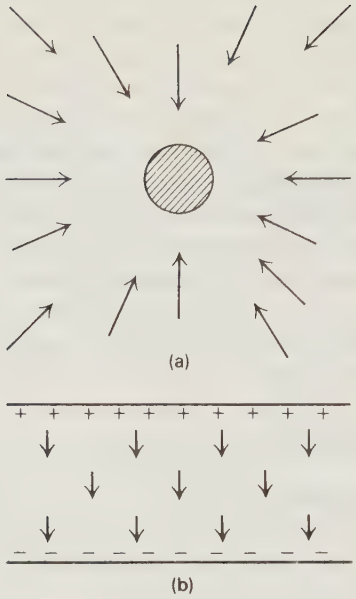


Fig. 33.16

its point is to move in the direction of the current (Fig. 33.14). You should check that this rule is identical in its conventions with Fleming's left-hand rule.

The rule can easily be adapted to find the direction of fields associated with flat coils and solenoids.

33.7 MAPPING OTHER FIELDS

The glass dish (Petri dish) (Fig. 33.15) contains a layer of castor oil and a layer of carbon tetrachloride. Between these two layers grass seed floats. The mixture is stirred with a glass rod and the plates connected to an EHT power pack capable of an output of 5 kV.

The grass seeds will align themselves with the direction of the electrostatic force at each point in the field. Field patterns for two point charges of the same or of opposite signs, for a point charge and a plane, or for two planes, are easily obtained.

Gravitational fields near to quite massive objects are so weak that it is impossible to devise a similar method for mapping them. But, near to the Earth which is the most massive object accessible to us, we may map the field by suspending a number of small masses in the space in which we are interested. The fact that the field of the Earth is distorted near to a mountain mass was indeed used by Bouguer in 1749 and again by Maskelyne in 1774 in the earliest attempts to determine the gravitational constant.

Since the gravitational force is always attractive, the field strength is always directed towards the source of the field. Electrostatic forces can be either attractive or repulsive, depending on the relative signs of the charge on the source and the charge on the test body. Conventionally, the field is said to act in the direction in which a positive test charge would move.

33.8 LINES OF FORCE

Fields may be represented pictorially by lines attached to points in space in which we might be interested. For example, the gravitational field of the Earth might be shown as in Fig. 33.16(a) whilst the electric field between two parallel sheets of charge might be shown as in Fig. 33.16(b).

However, it has long been the custom to draw such maps using continuous lines whose direction at each point is the direction of the field at that point. In Fig. 33.17 the two cases described above are drawn according to this convention. Such field lines are often called lines of force. Some care must be taken today over the phrase "lines of force." In the case of electric and gravitational fields, the forces they exert are in the same direction as the field. In the case of magnetic fields, as we shall shortly see, the field is not in the same direction as the force when it is exerted on an electric current and the phrase "line of force" can be very misleading.

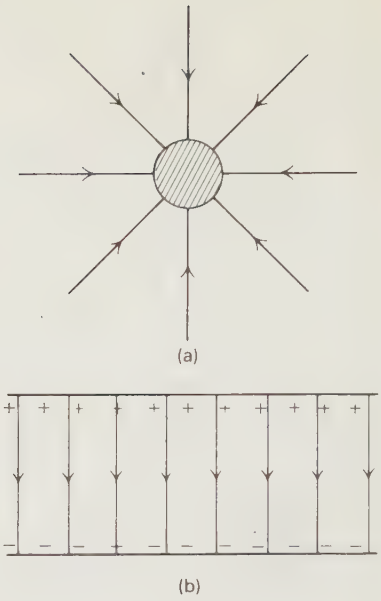


Fig. 33.17

33.9 FARADAY AND LINES OF FORCE

Michael Faraday was the first to use lines of force—continuous lines of field direction—extensively. To Faraday, these lines became more than a representation of changing field directions.

J. C. Maxwell gives a brief and easily-read account of Faraday's model for the electric and magnetic field in the same scientific paper *Action at a Distance* from which we have already quoted. What follows is a brief abstract from that account (Maxwell, 1873).

Maxwell first describes how a magnetic field pattern can be displayed by scattering iron filings in the field. He goes on

> "The mathematicians saw in this experiment nothing but a method of exhibiting at one view the direction in different places of the resultant of two forces one directed to each pole of the magnet; a somewhat complicated result of the simple law of force. "But Faraday, by a series of steps as remarkable for their geometrical definiteness as for their speculative ingenuity, imparted to his conception of these lines of force a clearness and precision far in advance of that with which the mathematicians could then invest their own formulae. "In the first place, Faraday's lines of force are not considered merely as individuals, but as forming a system, drawn in space in a definite manner so that the number of the lines which pass through an area, say of one square inch, indicates the intensity of the force acting through the area. Thus the lines of force become definite in number. The strength of a magnetic pole is measured by the number of lines which proceed from it; . . .
> ". . . By means of this new symbolism, Faraday defined with mathematical precision the whole theory of electro-magnetism, in language free from mathematical technicalities, and applicable to the most complicated as well as the simplest cases. But Faraday did not stop here. He went on from the conception of geometrical lines of force to that of physical lines of force. He observed that the motion which the magnetic or electric force tends to produce is invariably such as to shorten the lines of force and allow them to spread out laterally from each other. He thus perceived in the medium a state of stress, consisting of a tension like that of a rope, in the direction of the lines of force, combined with a pressure in all directions at right angles to them." [*Reprinted from* Proc. Royal Inst. 7.]

This was an extremely important and successful model. It was modified by Maxwell so that he was able to build up a theory of the electromagnetic field. He then showed that the model was no more than an

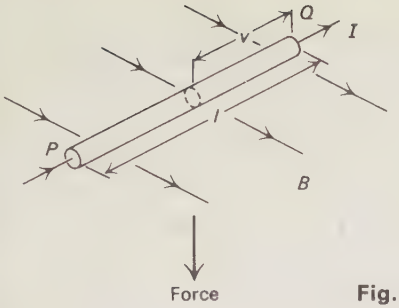


Fig. 33.18

aid to thinking and that ultimately it could be dispensed with. Faraday, however, as well as many who came after him, endowed these lines of force and the state of stress with a physical reality that they do not possess: they have no other properties than those given them in order to picture the behaviour of charges and currents in the field. Such models can eventually be a hindrance to thought rather than a help, and for this reason, we prefer today not to build such a mechanical model of field behaviour. We have even become wary of talking of "lines of force" lest we should imply that they have some reality. Such an assumption of reality was valuable to Faraday but it no longer provides us with a satisfactory model for our experience.

33.10 THE MAGNETIC FIELD AND MOVING CHARGES

Hitherto we have visualized the electric current as a flow of something to which we gave the name *electric charge*. This charge was measured in coulombs and a current of it in ampères or coulombs per second. As we shall see in Unit 8, a new model for electricity came to the fore towards the end of the nineteenth century. Campbell (1913) described the new model in these terms:

> "The fundamental hypothesis of the newer theory is that electric charges are not distributed continuously through the bodies which contain them, but are concentrated on particles the total volume of which is small compared with that of the charged body. The negative charges are concentrated on particles which have the same properties from whatever body they are derived. It is only necessary to attribute to them two properties, a charge and a mass. These particles are termed 'electrons,' a name originally invented by Johnstone Stoney to denote the indivisible unit of electric charge, the existence of which appeared to be indicated by the phenomena of electrolysis. On the other hand the particles on which positive charges are concentrated are, so far as we know, never smaller than atoms, so that the discontinuity in the distribution of positive electricity is of the same nature as that of the distribution of matter." [*Reprinted from Campbell, N. R. (1913).* Modern Electrical Theory, *Cambridge University Press.*]

According to such a model as this, the current within a wire will be made up of a flow of discrete charges. Let us now assume that this is so, leaving a consideration of the evidence until Unit 8.

Equation (33.7) gave

$$F_M = BIl,$$

as the force acting on a current-carrying wire in a magnetic field whose direction was normal to the wire. What is the force on the moving charges assumed to make up this current?

Consider the current element PQ carrying a current I. PQ is held at right angles to the magnetic field B as shown in Fig. 33.18. Suppose that the current is made up of n charges per unit length, each of size $+q$ and moving with a speed v. In unit time, all charge within a distance v of the point P on the wire will pass that point. Thus

$$I = nvq.$$

Now force

$$F_M = BIl,$$

$$\therefore F_M = B(nvq)l.$$

The total number of charges in the length l is nl. So,

$$\text{force on each charge} = F/nl,$$

$$= Bqv, \qquad (33.8)$$

if our model is correct.

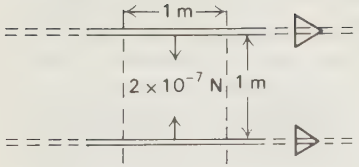


Fig. 33.19

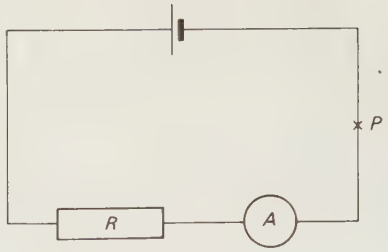


Fig. 33.20

This demands an experimental test. We must determine the force expected to act on known charges moving with known velocity. We may then predict the path which the charges will follow as they move in the magnetic field and we may compare the prediction with the path which they do in fact follow. The results of such experiments are not inconsistent with our model.

It is important to note that if we had assumed that the current consisted of a flow of electrons of charge $-e$, they would necessarily have had a velocity of $-v$ (v, in the opposite direction) in order to constitute a current in the conventional sense. So

$$F_M = B(-e)(-v)$$
$$= Bev \qquad (33.9)$$

and the force is in the same direction as before.

33.11 A UNIT FOR CURRENT

Our three field strengths are given by the following results:

$$g = F_G/m,$$
$$E = F_E/q,$$
$$B = F_M/Il.$$

No difficulty is experienced in finding g since the units of force (newtons) and mass (kilogrammes) are already determined from basic work in dynamics. To determine E, numerically, requires a unit of charge. To determine B requires a unit of current. There is no other way in which these units can be derived. We can make them what we please, but once chosen, we must stick by our decision.

But we are not entirely free. We cannot make two free choices because

electric current = rate of flow of charge.

Once we have a unit of charge we have a unit of current, or vice-versa.

For reasons of convenience (ease and accuracy of measurement), the ampere is made the subject of an arbitrary (i.e., freely chosen) definition.

We have seen already that a force acts on a current-carrying conductor in a magnetic field and that two current-carrying conductors attract or repel each other with a measurable force. It is this phenomenon which provides the international definition of the ampère (see Section 30.1). This is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section and placed one metre apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} newton per metre of length (Fig. 33.19).

This definition is not so important, for our purposes, as the realization that to measure a magnetic field we have to agree on a fundamental unit additional to the metre, the kilogramme and the second. That other unit is the ampere.

The unit of charge is derived directly from this unit of current. If, in the simple circuit shown in Fig. 33.20, 1 ampere is flowing then the charge passing the point P in each second is 1 coulomb.

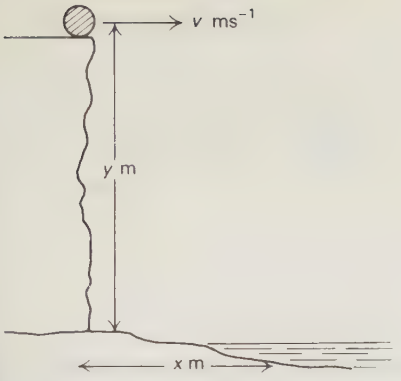


Fig. 33.21

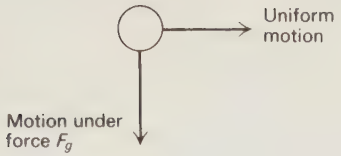


Fig. 33.22

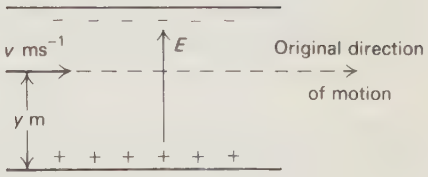


Fig. 33.23

33.12 THE BEHAVIOUR OF BODIES UNDER THE INFLUENCE OF A FIELD OF FORCE

Understanding the behaviour of masses in a gravitational field is a prerequisite for the launching of Earth satellites and planetary probes, for the solution of problems of planetary motion and for describing the motions of the stars in their galaxies. The electric motor is an application of the force acting on a current-carrying wire in a magnetic field. Mass spectrometers, particle accelerators and the Millikan experiment to determine the charge of the electron are all concerned with the behaviour of charged particles in electric and magnetic fields. This wealth of application might suggest that the remainder of our study of fields should be concerned with such matters. But in doing that we should learn little more about the force fields themselves. We should, for example, be no nearer to finding ways to calculate the sizes of these fields—and this is a vital preliminary step to the description of their action on bodies influenced by them.

The description of the resultant behaviour of bodies acted upon by fields of force is not so much a problem for field theory as one for dynamics—"given the force, find the resultant motion." Indeed, once a particular problem has been solved for one type of field, the results can often be directly applied to another. To show this is so we will look at two apparently dissimilar problems.

Problem 33.3 A ball is kicked horizontally over a cliff edge. How far away from the foot of the cliff does it fall?

Problem 33.4 An electron beam, accelerated by an electron gun, passes between two parallel plates in a cathode ray tube, between which there is a uniform electric field. How much is the beam deflected from its original direction of travel?

We will look first at Problem 33.3 and Fig. 33.21.

We will suppose the height of the cliff to be y, measuring our distances downwards from the top of the cliff. Let the distance the ball lands from the foot of the cliff be x. We will ignore any effects of air resistance.

The ball has a mass m, and so it falls under the action of a force F_G, where F_G is the force of gravity on the ball. The acceleration of the ball vertically downward is

$$\frac{F_G}{m} = g$$

from Newton II. Horizontally the ball travels at a constant velocity of v.

Why is this? Remember that, in the absence of air resistance, the only force acting is the pull of the earth downwards.

After a time t,

$$x = vt, \qquad y = \tfrac{1}{2}gt^2.$$

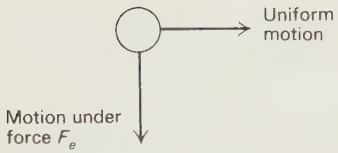


Fig. 33.24

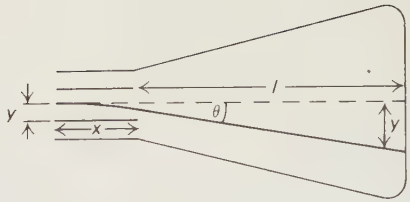


Fig. 33.25

Eliminating t between these two equations,

$$y = \frac{1}{2} g \frac{x^2}{v^2}.$$

Consequently, a knowledge of y and v will allow us to calculate x, the distance away from the cliff that the ball falls to the ground.

(Find x, if $y = 50$ m and $v = 15$ m s^{-1}. What effect do you think air resistance would have on this answer?)

The solution to this equation involves the equations of motion under constant acceleration. The ball is acted upon by the gravitational field. This force enables us to calculate the downward acceleration. Once this has been done, the field can be forgotten (see Fig. 33.22).

Now let us look at Problem 33.4 and Fig. 33.23.

We will suppose that the parallel plates are of length x and that the uniform field between them is E. y is the vertical distance measured downwards from the original direction of the beam.

Let the charge on each electron be e and the mass m. The electron comes under the action of forces in just the same pattern as did the football (Fig. 33.22).

Why is the gravitational pull of the Earth on the electron legitimately neglected in this statement?

The acceleration of the charge towards the bottom plate is Ee/m. This replaces g in Problem 33.4.

We can immediately write down the answer to this problem, by comparing it to the first. In that case

$$y = \frac{1}{2} g \frac{x^2}{v^2}.$$

So now,

$$y = \frac{1}{2} \frac{Ee}{m} \frac{x^2}{v^2}.$$

By the time the stream of charged particles have reached the edge of the parallel plates, they will have been deflected by a distance of

$$y = \frac{Ee}{2mv^2} x^2$$

from their original path.

Now the problem differs from the earlier one. The beam leaves the electric field and continues at a uniform speed in the direction it was following as it left the field until it reaches the face of the tube (see Fig. 33.25). Calculate the angle which the beam makes with the horizontal as it leaves the electric field, using the knowledge of the horizontal and vertical velocities. Calculate further the deflection of the beam when it reaches the tube face, if this is 0.25 m beyond the edge of the field.

Chapter 34
THE INVERSE SQUARE LAW

34.1 Variations in field strength
34.2 Flux and the inverse square law

34.1 VARIATIONS IN FIELD STRENGTH

Once we can measure electric, gravitational and magnetic field strengths experimentally, we can investigate how these fields vary. We would certainly expect such a field to depend on the size of the source. We might expect it to depend on its physical shape. We might expect the field strength to depend in some way on the distance from the source (or sources) of the field. The nature of the intervening medium might well have an effect.

We have observed that the gravitational field near to the surface of the earth is uniform: so is the field between charged, parallel plates. In this latter case, the strength of the field will depend on how much charge is on the plates and it may be affected by the material between the plates.

In this chapter we shall look at some of the factors which affect field strengths and see what sort of laws relate the measured field strengths to their source and surroundings. We shall look at this experimentally and we shall find one law figuring more prominently than any other—the *inverse square law*. Inverse square laws are very prominent throughout Physics. For example, the intensity of gamma radiation falls off inversely with the square of the distance from the source. Light intensities decrease in the same way. Neither of these are field strengths. First of all then, we will consider one of the reasons for the frequency with which inverse square laws occur, by considering the inverse square law for an aerosol paint spray.

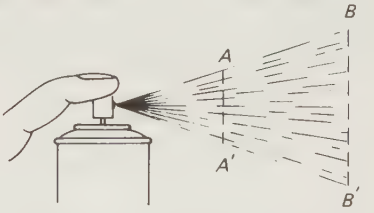


Fig. 34.1 The aerosol paint law.

Let us suppose we want to spray a patch of colour on a piece of metal held perpendicular to the path of the spray, which is fixed in one position (Fig. 34.1). We apply these restrictions because we are trying to make the problem as simple as possible. Clearly we

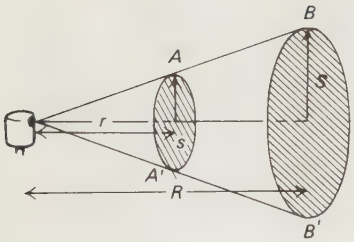


Fig. 34.2

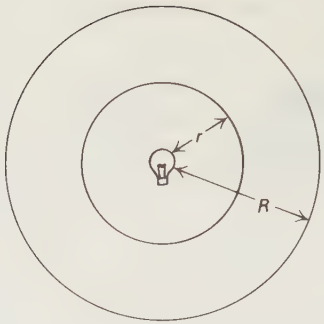


Fig. 34.3

shall cover a much bigger patch if we place the metal at BB' than if we place it at AA'. We could cover an even bigger patch if we moved the piece of metal further away still. Can we paint bigger and bigger areas satisfactorily by just moving the surface to be sprayed further away from the can?

The amount of paint per square metre will get less and less and we may not be covering the surface densely enough to give a satisfactory paint cover if the metal surface is too far away.

Why is this? Because in all cases, the same total amount of paint falls on the surface in unit time. It is just spread out more thinly as we move the painted surface away from the spray. How much more thinly? Consider Fig. 34.2. The area covered by the spray at $AA' = \pi s^2$, where s is the radius of the cone of spray at distance r. The area covered by the spray at $BB' = \pi S^2$, where S is the radius of the cone of spray at distance R. If the amount of paint P leaves the spray in unit time, the density of paint

$$\text{at } AA' = P/\pi s^2,$$

$$\text{at } BB' = P/\pi S^2.$$

But

$$s : S = r : R.$$

Hence

$$\frac{\text{density of paint at } A}{\text{density of paint at } B} = \frac{S^2}{s^2} = \frac{R^2}{r^2}.$$

The density of paint sprayed to any distance r from the source is proportional to $1/r^2$.

So the density obeys an inverse square law. This law is a consequence of the geometry of space.

A law of the same form applies to light intensities. The amount of light energy I provided by a point source per second, will under normal circumstances, be the amount passing through the surface of the sphere of radius r and also through the sphere of radius R (Fig. 34.3).

If the light is emitted uniformly in all directions the illumination (light intensity) at distance r is

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

and at distance R is

$$\frac{I}{4\pi R^2}.$$

It thus follows that the light intensity at any distance r from the source is proportional to $1/r^2$.

34.2 FLUX AND THE INVERSE SQUARE LAW

The inverse square law is a law of flow. When any quantity flows outwards from a point source uniformly in all directions at an unchanging rate then the amount passing through parallel surfaces of the same area at different distances from the source will obey an inverse square law. The amount flowing out is known as a *flux*—in the examples considered, a flux of paint and a light flux.

Under these conditions, the flux per unit area diminishes inversely as the square of the distance from

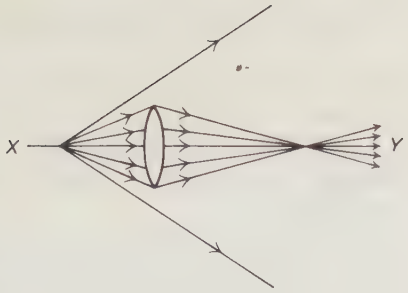


Fig. 34.4

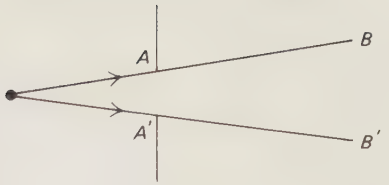


Fig. 34.5

the point source. The flux per unit area is known as the *flux density*. In the case of the paint spray, the flux density measures the degree to which the paint covers the surface to be painted; in the case of light, the flux density is a measure of the illumination of the surface on which the light falls.

Problem 34.1

a) Do you think the inverse square law will remain true in a smoke-filled room with a lamp at the centre?

b) Consider how the law will be modified when light passes through a lens as in Fig. 34.4. At what points along XY will the law remain true and how should it be expressed?

c) What happens to the light intensities if the light is passed through a narrow slit and diffraction occurs? (see Fig. 34.5). Will the inverse square law still apply in the plane BB'?

Chapter 35
THE ELECTRIC FIELD

35.1 The electrostatic force between two charged particles
35.2 A model for the electric field
35.3 A consequence of the flux model
35.4 Using the flux model

35.1 THE ELECTROSTATIC FORCE BETWEEN TWO CHARGED PARTICLES

We have referred briefly in Section 33.4 to a method of testing whether the force on a charged particle in an electric field was proportional to the charge. We asked how that charge might be measured. Here is one answer.

The charge which is to be measured is transferred by contact to a capacitor, whose capacitance is much larger than that of the conductor carrying the charge. This conductor is commonly a very light metallized sphere about a centimetre across. It has a capacitance in the order of a fraction of a picofarad. In contact with a capacitor of, say, 0.01 μF, the charge will be transferred almost completely to the capacitor from the sphere. You should estimate the percentage of the charge remaining on the sphere after this process is complete.

The problem is now reduced to one of measuring the potential differences across the capacitor. This can be done with the help of an electrometer. This is an electronic device in which the value of the current flowing in an electrometer tube is determined by the potential difference applied across two of the electrodes in the tube. This current is then amplified and its value indicated on a suitable milliammeter. This whole instrument is, in effect, a voltmeter with an extremely high impedance (about $10^{13}\ \Omega$). Such instruments can readily be calibrated and normally give, in use, a full scale deflection on the meter for an applied potential difference of 1 V. They are then capable of measuring charges in the order of 10^{-9} C and currents in the order of 10^{-11} A as well. An example may help to make this clear. Let us suppose that, having received the charge on the conductor, the capacitor (of capacitance C) has a potential difference of 1 V across it. Since $V = q/C$ the charge q conveyed to the capacitor was CV or (with a 0.01 μF capacitor) $0.01 \times 10^{-6} \times 1$ C which is 10^{-8} C. So 10^{-8} C will raise the potential difference across the capacitor to 1 V. If this charged capacitor is now connected across the input terminals of the electrometer it will provide a direct indication of charge up to this value.

To investigate the way in which the force between

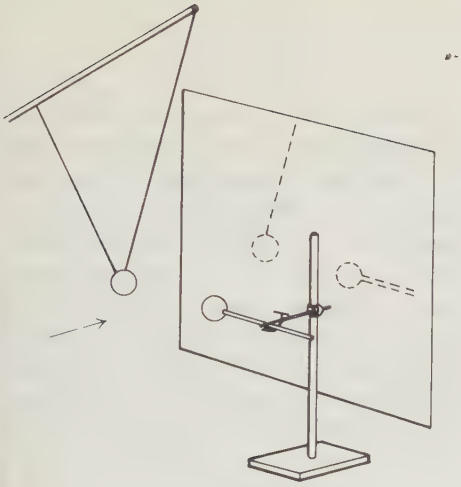


Fig. 35.1

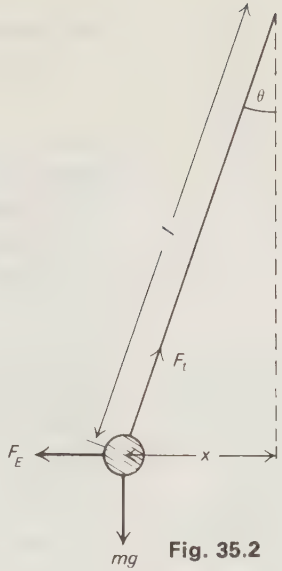


Fig. 35.2

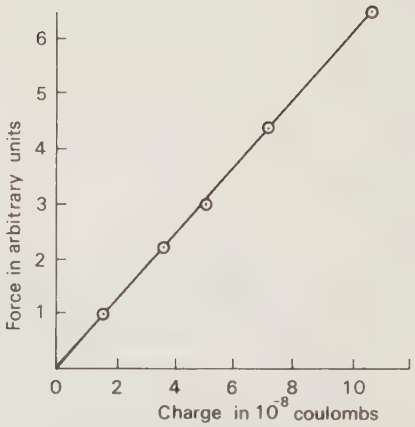


Fig. 35.3

two charges varies with those charges and with their separation we will describe two experiments. In the first, the light metallized sphere is fixed to an insulated stem and given a charge which is to remain unaltered throughout the experiment. This may be done by using either a power pack providing several kilovolts, or an electrophorus or a Van der Graaff generator. A second similar sphere, which is suspended from a well insulated nylon trapeze (see Fig. 35.1) is also charged. The first sphere, held rigidly as shown, is brought up to the suspended sphere. Repulsion occurs if the charges are of the same nature. To measure the separation of the two spheres without bringing a rule near to them it is convenient to throw a shadow of them on to a screen using light from a distant source. The angle of deflection (θ) of the nylon threads is measured. The separation of the spheres (r) is noted and the charge (q) on the suspended ball is discharged to the electrometer and so measured. The experiment is repeated with different charges on the suspended sphere but with the same separation of the spheres and with the same charge on the fixed sphere.

If F_E is the force between the two spheres, F_T the tension in the thread and m the mass of the sphere, x the displacement shown in Fig. 35.2 and l the length of the thread we have, at equilibrium

$$F_E = F_T \sin \theta,$$

and

$$mg = F_T \cos \theta.$$

Dividing

$$F_E/mg = \tan \theta,$$

so

$$F_E = mg \tan \theta,$$

$$\approx mg\, x/l,$$

provided that θ is small.

The deflection of the nylon suspension is therefore directly proportional to the force of repulsion between the two charges.

Figure 35.3 shows the relationship between the charge on the sphere and the deflection of the suspension, which, as we have seen, is a measure of the force. We conclude that

$$\text{force} \propto \text{charge},$$

and that our earlier assumption was justified.

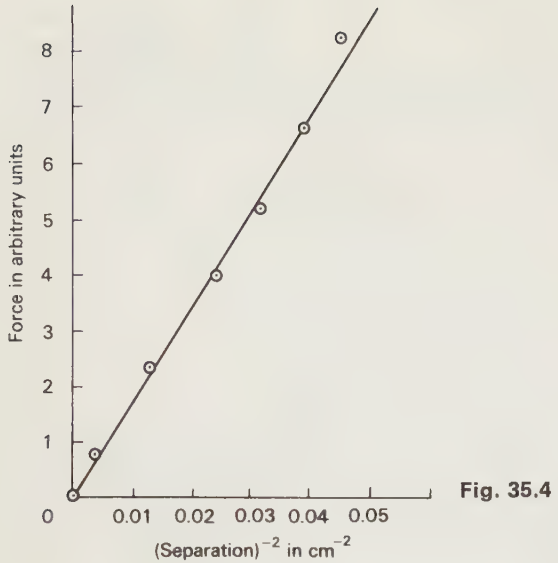


Fig. 35.4

In the second experiment, the charges on the two spheres are kept the same and the separation r is changed. Figure 35.4 shows the graph which is obtained under these circumstances when the deflection (recorded as the distance x) is plotted against

$$\frac{1}{r^2}.$$

We conclude that

$$\text{force} \propto \frac{1}{r^2}.$$

(This experiment can never be very accurate. You should suggest some of the errors likely to be encountered.)

More accurate experiments confirm that the force of one charged sphere on the other varies inversely as the square of their separation. We already know that the force is proportional to q, the charge on the suspended sphere. Further experiment can show that it is also proportional to the charge, Q, on the sphere fixed to the insulated stem.

So we find

$$F_E \propto \frac{Qq}{r^2},$$

and therefore

$$F_E = k\frac{Qq}{r^2},$$

where k is a constant of proportionality. This constant is usually written $1/4\pi\varepsilon_0$. That it appears in this form is due to an accident of history and to the logic of the spherical symmetry with which the product 4π is so closely associated. It arises in this case thus: consider the field surrounding an isolated point charge. Imagine a sphere, radius r, enclosing the charge and centred upon it (Fig. 35.5). If $F \propto Qq/r^2$ then the field is the same at every point on the sphere and is everywhere directed normally to its surface. Such a field displays spherical symmetry. We have then

$$F_E = \frac{1}{4\pi\varepsilon_0}\frac{Qq}{r^2} \qquad (35.1)$$

and for the field intensity (or force per unit charge)

$$E = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}. \qquad (35.2)$$

We have already met the constant ε_0 (Section 28.4). It is the permittivity of free space. For media other than a vacuum, the constant is written $\varepsilon_r\varepsilon_0$, where ε_r is the relative permittivity of the medium. Since the relative permittivity of air differs but little from unity, we may use ε_0 when the charges are separated by air rather than a vacuum.

The value of the permittivity constant ε_0 is

$$8.854 \times 10^{-12} \quad \text{or} \quad \frac{1}{36\pi} \times 10^{-9}$$

and its units are often quoted as $\text{J}^{-1}\ \text{C}^2\ \text{m}^{-1}$ or as F m^{-1}.

You should check these units against those which you would expect from the equation

$$E = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}.$$

35.2 A MODEL FOR THE ELECTRIC FIELD

Our experiments confirm that the electric field intensity obeys an inverse square law and, as we have

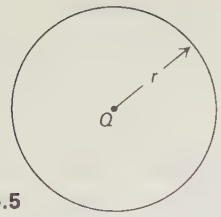


Fig. 35.5

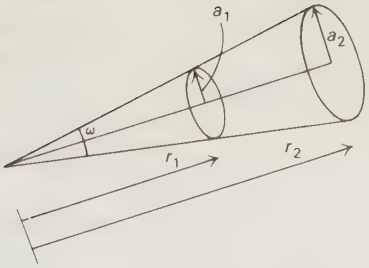


Fig. 35.6

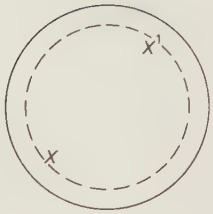


Fig. 35.7

seen (Section 34.2) this represents the typical behaviour of a flux density. We will suppose then, that E, the field intensity, does behave like a flux density.

We have no need to enquire into the reality of this analogy between the two phenomena but we can use it. It provides us with a useful model for discussion.

What would be the total flux from Q in this case? This total electric flux, ψ, must be the flux density, E, at the sphere (Fig. 35.5) multiplied by the area of the surface of the sphere.

$$\psi = E \times 4\pi r^2$$

$$= \frac{Q}{4\pi\varepsilon_0 r^2} \times 4\pi r^2$$

$$= \frac{Q}{\varepsilon_0}. \qquad (35.3)$$

We were able to calculate ψ very easily because we chose a particularly simple and symmetrical case. But the shape of the surface surrounding Q should not affect the electric flux associated with Q, if our model is a good one. To return to the flux law of light, the placing of a book relative to a lamp will certainly control how much light falls on its pages, but it does not affect how much light leaves the lamp. Similarly if we spray a surface with paint by holding the paint spray close to it we shall get a dense layer of paint. If we hold it a long way off we shall get a thin layer of paint—but the amount of paint coming from the spray per second is the same in both cases.

So we can generalize our model and assume that any charge Q is a source of electric flux, Q/ε_0. The electric field at any point is then the density of the electric flux at that point. This is a simplified statement of Gauss's law. We will look at some consequences and applications of this model shortly. But before that, let us again consider the steps we have taken in setting up this model.

We have used the fact that flux densities, like paint films and illumination, obey inverse square laws in certain symmetrical cases. Why they do so is simply a property of the geometry of space (see Fig. 35.6).

For the solid angle, ω,

$$a_1/a_2 = r_1/r_2,$$

$$\text{area } A_1 = \pi a_1^2, \qquad \text{area } A_2 = \pi a_2^2.$$

Thus

$$A_1/A_2 = a_1^2/a_2^2,$$

$$= r_1^2/r_2^2.$$

The areas are in proportion to the square of their distances from 0.

The validity of our model of the electric field depends upon the fact that experimentally the electric field E obeys the inverse square law whenever a flux density model suggests that it will do so.

Thus ultimately our model depends upon the trust we can place in the power of 2 in the expression

$$E \propto Q/r^2.$$

We assume in our model that the power is not roughly 2, or sometimes 2, but always exactly 2. Can we justify this?

35.3 A CONSEQUENCE OF THE FLUX MODEL

Figure 35.7 represents a solid conducting sphere. Let us suppose that it is given a charge q. Now imagine

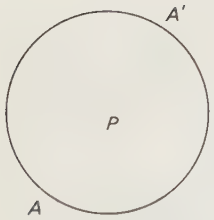


Fig. 35.8

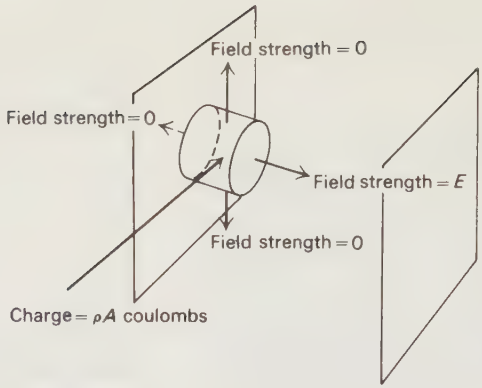


Fig. 35.9

that a second sphere, whose surface is represented by XX' is drawn inside the surface of the first sphere. Since the material, which is a good conductor of electricity, fills the whole of sphere XX', electric charges can move freely about it. If there is any electric field within this imaginary sphere, the electric charges will be acted upon by a force and so they will move. Equilibrium will only be established when no net force acts on the charges.

But if there is no net force acting on the charges within XX', there can be no electric field within XX'. Using our flux model, the electric flux through the surface XX' is equal to

$$E \times \text{area of surface} = 0, \text{ if } E = 0.$$

But we have just seen that the electric flux $= Q/\varepsilon_0$. If the electric flux through XX' is zero then,

$$Q/\varepsilon_0 = 0,$$

in other words

$$Q = 0.$$

Hence we come to the surprising conclusion that the total net charge within XX' is zero.

Now we could have drawn XX' as close to the surface AA' of the sphere as we liked; the result would have been just the same. But the conducting sphere is certainly charged—as tests outside the sphere show. So where is that charge located? Our reasoning suggests that it must all be on the surface of the sphere.

That this result is a consequence of the flux model and hence of an inverse square law may be difficult to see. The converse argument—that a surface layer of charge will lead to zero field in the interior of a sphere is more convincing: suppose AA' is a thin spherical shell carrying charge Q (Fig. 35.8). The charge can be divided up into small charges each of size Δq. The field $E_{\Delta q}$ at P due to Δq is $\Delta q/4\pi\varepsilon_0 r^2$.

It is possible mathematically, but beyond the scope of the present book, to show that the total effect of all the charges Δq into which we have subdivided Q gives rise to a net field E at P which is zero. This is only true for an inverse square law of force.

Arguments such as these show that for charged, closed metallic conductors of any shape, the total charge should reside on the surface, if the inverse square law of force is true.

In 1936, Plimpton and Lawson tested this fact very carefully. They tried to find whether any charge was left on the inner of two spheres after it had been charged and then connected to the outer one. As a result they estimated that the difference between the power to which r is raised and 2 is no greater than 2×10^{-9}.

35.4 USING THE FLUX MODEL

Provided the electric field from a given source is known to have a certain symmetry (and the field lines experiment will help to establish this), the flux model

can be used to predict the way the field strength is related to charge and distance. We will look at just one example—the field between parallel plates, which we already have some evidence for believing to be uniform. The flux law can be applied to any surface, so we will choose a surface for which it is particularly easy to calculate the total electric flux.

Our chosen surface (Fig. 35.9) is that of a cylinder of cross-sectional area A, one end of which is on the surface of one plate. If the charge on unit area of the plates is ρ (this is called the charge density) then the total charge within the chosen cylinder is ρA.

$$\text{The total flux through the cylindrical surface} = \frac{Q}{\varepsilon_0}$$

$$= \frac{\rho A}{\varepsilon_0}.$$

Let E represent a flux density through the end of the cylinder, which is situated between the plates. Experiment confirms that there is no field directed through the sides of this cylinder or through the other end. So the total electric flux through the cylinder is given by the product of the flux density E and the area of the end A.

Hence

$$EA = \frac{\rho A}{\varepsilon_0}$$

and

$$E = \frac{\rho}{\varepsilon_0}. \qquad (35.4)$$

As we shall see later, this result is confirmed by experiment.

How is the operation of a Van der Graaff generator accounted for in terms of the flux model? Examine how such a generator is constructed. The charge collects on the two almost closed conducting shells which are to be found at the base and the top of the pillar which carries the belt. The charge is carried up this pillar by the belt. How is the belt enabled to discharge at the top? Why doesn't the charge which is already there prevent the discharge of the belt?

Chapter 36
THE GRAVITATIONAL FIELD

36.1 The force of gravity
36.2 Weightlessness
36.3 Newton's test of the inverse square law
36.4 The measurement of the gravitational constant (G)
36.5 Comparison of electric and gravitational fields

36.1 THE FORCE OF GRAVITY

We have already (Unit 1) referred to Newton's theory of gravitation. It suggests that the force between two masses M and m separated by a distance, r, is given by

$$F_G \propto \frac{Mm}{r^2}$$

and that

$$F_G = G\frac{Mm}{r^2}, \qquad (36.1)$$

where G is a constant. The distance r is taken to be very much larger than the size of either of the two masses. We do not then have to specify whether r is measured from centre to centre or between the closest parts of the masses. This was certainly an acceptable approximation when applied by Newton to the planets.

It is very difficult, if not impossible, to test the dependance of this force on mass and on distance in the laboratory for the gravitational forces between "ordinary" bits of matter are far too small. However, as we shall see, it is possible to measure the value of the gravitational constant G of Eq. (36.1).

Newton was able to derive this inverse square law for gravitation from Kepler's laws of planetary motion. These summaries of the observations which had been made of the motion of the visible planets by, primarily, Tycho Brahé towards the end of the sixteenth century were arrived at by Kepler after years of study. The laws may be summarized thus.

1. The orbit of each planet is an ellipse with the sun at one focus.
2. The arm drawn from the sun to a planet sweeps out equal areas in equal periods of time. See Fig. 36.1.
3. The squares of the periods of revolution are proportional to the cubes of the average distances of the planets from the sun.

Modern observations on the solar system, on the moons of the planet Jupiter and on Earth satellites confirm the success of the three laws as summaries (but no more than summaries) of the data. See, for example, Table 36.1.

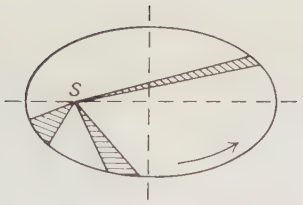


Fig. 36.1 Elliptical paths of planets round the Sun S. Eccentricity much exaggerated.

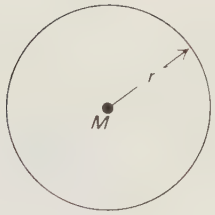


Fig. 36.2

Table 36.1 Solar system data and a test of Kepler's third law

Planet	Mean radius of orbit (R) in metres	Period of orbital motion (T) in seconds	R^3/T^2
Mercury	5.79×10^{10}	7.60×10^{6}	3.36×10^{18}
Venus	1.08×10^{11}	1.94×10^{7}	3.35×10^{18}
Earth	1.49×10^{11}	3.16×10^{7}	3.30×10^{18}
Mars	2.28×10^{11}	5.94×10^{7}	3.36×10^{18}
Jupiter	7.78×10^{11}	3.74×10^{8}	3.36×10^{18}
Saturn	1.43×10^{12}	9.30×10^{8}	3.37×10^{18}

Kepler's three laws can only be accounted for if the law of force which applies to the sun and the planets is an inverse square law. The vital case is, in fact, that described in Law 1. Only an inverse square law of force can lead to planetary orbits which are elliptical with the sun at one focus. The second law is, in fact, valid for any centrally directed force.

The story of the relationship between Newton's inverse square law and Kepler's laws and the many consequences of this law of force is a long and interesting one. (See for example Koestler, 1959.)

Let us accept the truth of this inverse square law for gravitational forces.

$$F_G = G\frac{Mm}{r^2}.$$

By analogy with the case of electrostatic force (Eq. 35.1) we see that the strength of the gravitational field may be written

$$g = G\frac{M}{r^2}. \qquad (36.2)$$

G is called the universal gravitational constant and has a value of 6.67×10^{-11} N m^2 kg^{-2}.

The comparison with electric fields can be extended by applying the flux model to gravitation. Because g obeys an inverse square law, it behaves like a flux. If the radius of the mass M is small compared with r, the field through the sphere of radius r is the same all the way round.

$$\begin{aligned}\text{The total gravitational flux} &= g4\pi r^2,\\ &= G(M/r^2) \times 4\pi r^2,\\ &= 4\pi GM. \qquad (36.3)\end{aligned}$$

36.2 WEIGHTLESSNESS

One remarkable consequence of the gravitation force law is the fact that the period of rotation of any mass, m, orbiting about the other attracting mass M at a distance, r, from it is independent of the size of m.

Consider a mass m in orbit with speed v and distance r from M.

Acceleration of the mass, m, towards $M = v^2/r$ (see Unit 2). This is provided by the force F_G exerted by M upon m, and so

$$F_G/m = v^2/r \qquad \text{(Newton II)}.$$

Since $F_G \propto m$, it follows that all masses moving around M in an orbit of radius r will have the same acceleration towards M and will all therefore have the same speed, v.

This leads to the phenomenon referred to as *weightlessness*. An astronaut in a space capsule in orbit around the Earth experiences no reaction to his weight in the capsule, since the pull of gravity on him is just sufficient to maintain him at the same orbiting speed as the capsule, at the capsule's distance from the Earth. (See Section 8.7.)

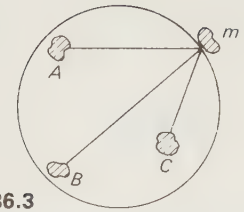


Fig. 36.3

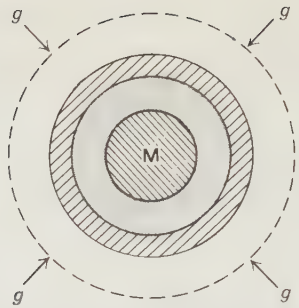


Fig. 36.4

36.3 NEWTON'S TEST OF THE INVERSE SQUARE LAW

Newton's own test of his law of gravitation was a computation of the moon's period of rotation about the Earth, on the assumption that it moved under the influence of the Earth's gravitational field.

The pull of gravity at the Earth's surface gives rise to an acceleration for all masses of 9.8 m s^{-2}. If the inverse square law applies, can we deduce what the acceleration towards the Earth of a mass as far away as the moon should be?

While we shall assume throughout that the moon's orbit is circular, the problem is not a simple one because the Earth and the moon are not point masses. Let us consider in detail the force of attraction between a mass m on the surface of the Earth and the Earth itself. This mass m is attracted by A, B, and C and all the other particles which make up the Earth (Fig. 36.3). These are all at different distances from m. We need to know the effective distance of the rest of the Earth from the mass m for the purpose of comparing this force with the force exerted by the Earth on a mass at the distance of the Moon's orbit. We can apply the flux model to this case.

Provided the mass M (Fig. 36.4) has spherical symmetry, i.e., any variations of density are a function of the radius only, the field strength at the surface of the dotted sphere just outside the surface is the same at all points. Let it be g. Then by the flux law (Eq. 36.3)

$$4\pi r^2 g = 4\pi GM$$

$$g = G\frac{M}{r^2}.$$

This field strength g is the same as it would be if all the mass M was concentrated at the centre, and we may therefore specify the Earth's radius (6.4×10^6 m) as the appropriate separation between the mass m and the Earth. So the force between m and the Earth when on the Earth's surface is

$$F_G = G\frac{Mm}{R^2},$$

where R is the Earth's radius.

The force on the same mass as far away as the Moon is

$$F_G' = G\frac{Mm}{r^2},$$

where r is the radius of the Moon's orbit (assumed to be circular).

Hence

$$\frac{\text{acceleration of } m \text{ at the Earth's surface}}{\text{acceleration of } m \text{ at Moon's orbit}} = \frac{1/R^2}{1/r^2} = \frac{r^2}{R^2}$$

$$\therefore \text{ acceleration of } m \text{ at Moon's orbit} = \frac{R^2}{r^2} \times g.$$

Now $R/r \approx 1/60$, and the acceleration of m at the Earth's surface is 9.8 m s^{-2}, so

$$\text{required acceleration} = 9.8/3600 \text{ m s}^{-2},$$
$$= 0.00272 \text{ m s}^{-2}.$$

Now centrally directed acceleration

$$= v^2_{\text{Moon}}/r,$$
$$\therefore v_{\text{Moon}} = 1.01 \times 10^3 \text{ m s}^{-1}.$$

Thus, Moon's period of rotation = 27.7 days.

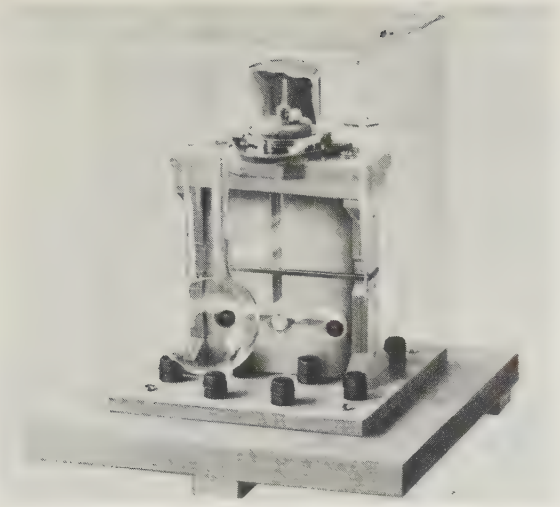


Fig. 36.5 Simple apparatus for the determination of the gravitational constant. In use, the two flasks are filled with mercury. (Courtesy of Philip Harris, Ltd.)

This is as near to the Moon's true period as such an approximate calculation might be expected to get. (It is not the time between successive full moons—which is longer. Why?)

36.4 THE MEASUREMENT OF THE GRAVITATIONAL CONSTANT (G)

The gravitational constant G cannot be determined from the force law between sun and planets. To do this, we should need to know their mass. The application of this very law is the only means of determining that. And for this we need to know G.

The measurement of the force of gravity between two known terrestrial masses was first achieved by Cavendish in 1798. Subsequently, his apparatus was improved by Boys, who was able to use the combined strength and low torsion of quartz fibres to reduce the scale of the apparatus.

Figure 36.5 shows a photograph of a modern, simple version of Boys' apparatus.

Two small masses are attached to a bar suspended from a quartz fibre with a very low torsional restoring force to twist. The displacement of these masses by gravitational attraction to the large masses seen outside the instrument can be measured. The experiment is lengthy and a speeded-up film of a set of measurements has been made the subject of a film loop (Miller). The details of this experiment are referred to in the list of references at the end of this unit.

36.5 COMPARISON OF ELECTRIC AND GRAVITATIONAL FIELDS

It is commonly said that gravitational fields and gravitational forces are much weaker than electric ones. This may seem like a comparison of French and English books: apart from the fact that they are written in different languages, what else is there to say?

And yet there is a sense in which the comparison has meaning. Consider two elementary charged particles, say two protons. They have both charge and mass. We may certainly compare the gravitational force of attraction between them with the electric force of repulsion.

Let m_p and e be the mass and charge of a proton. When they are separated by a distance r,

$$\text{the gravitational attraction} = G\frac{m_p m_p}{r^2}$$

$$\text{and the electrical repulsion} = \frac{ee}{4\pi\varepsilon_0 r^2}.$$

The ratio

$$\frac{\text{electrical repulsion}}{\text{gravitational attraction}} = \frac{e^2/4\pi\varepsilon_0}{Gm_p{}^2}.$$

Now

$$e = 1.67 \times 10^{-19}\ \text{C};$$

$$\frac{1}{4\pi\varepsilon_0} = 9 \times 10^9\ \text{J}^{-1}\,\text{C}^2\,\text{m}^{-1},$$

$$m_p = 1.67 \times 10^{-27}\ \text{kg},$$

$$G = 6.67 \times 10^{-11}\ \text{N m}^2\,\text{kg}^{-2}$$

so ratio

$$\frac{\text{electrical repulsion}}{\text{gravitational attraction}} = \frac{(1.6 \times 10^{-19})^2 \times 9 \times 10_9}{6.67 \times 10^{-11} \times (1.67 \times 10^{-27})^2}$$

$$= 1.22 \times 10^{36}.$$

We *must* agree that gravitational forces are weaker than electrical ones.

Problem 36.1 Predict Kepler's third law given Newton's law of gravitation.

Solution If a planet of mass m moves in orbit round a star of mass M, the central force of attraction must provide the central force necessary to retain that planet in its orbit. If we consider that the centre of the system coincides with the centre of the star (and if M is very large by comparison with m this is justified) then

$$G\frac{Mm}{R^2} = m\omega^2 R,$$

where R is the radius of orbit and ω is the angular frequency. If the periodic time is T,

$$T = 2\pi/\omega.$$

Substituting for ω we have

$$G\frac{Mm}{R^2} = m \cdot \frac{4\pi^2}{T^2} \cdot R$$

and

$$\frac{R^3}{T^2} = G\frac{M}{4\pi^2}.$$

Problem 36.2 From the information given in this section show that the mass of the Earth is about 6×10^{24} kg and that its average density is 5.5×10^3 kg m^{-3}.

Chapter 37
THE MAGNETIC FIELD

37.1 Measuring magnetic fields
37.2 Exploration of magnetic fields
37.3 The field at the centre of a flat coil
37.4 The magnetic field of a current-carrying solenoid
37.5 The field outside a long solenoid
37.6 The flux model and magnetic fields
37.7 The value of μ_0 (the magnetic force constant)

37.1 MEASURING MAGNETIC FIELDS

We have seen in Unit 6 that the simple current balance is a satisfactory instrument for defining the meaning of magnetic field strength (magnetic flux density) but it is not sufficiently sensitive to find out how that field strength varies from point to point in a field. Before we examine how we might do this, it is worth considering how we might produce a uniform magnetic field. Such fields occur within long solenoids. We may demonstrate this with a simple current balance of a suitable shape.

The circuit shown in Fig. 37.1 enables us to place the test conductor of the balance at various points within the solenoid. In a typical experiment with the solenoid current maintained at 3.5 amperes, the counter-balancing force which had to be applied to the balance through which 3 amperes were flowing varied as shown in Table 37.1.

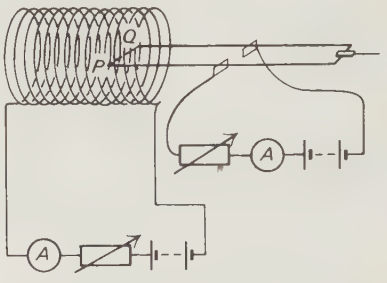


Fig. 37.1 Exploring the magnetic field within a solenoid with a current balance.

This counter-balancing force was equal in magnitude to the force exerted on the conductor PQ and provides, therefore, a measure of the magnetic field. The graph of force—and therefore of the field B—against position of the conductor within the solenoid has the form shown in Fig. 37.2.

Evidently, so long as we limit ourselves to the field in the central section of the solenoid, that field is a uniform one.

Moreover, if now we change the current in the solenoid whilst keeping that in the current balance fixed, we find the results displayed in Table 37.2.

Table 37.1

Distance of test conductor PQ from centre of solenoid (cm)	18	17	15	13	11	9	7	5	3	1
Restoring force (arbitrary units)	0.13	0.38	0.50	0.85	1.10	1.30	1.45	1.45	1.45	1.45

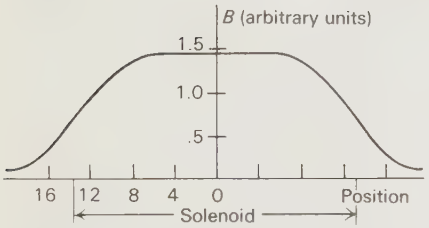


Fig. 37.2

Table 37.2

Solenoid current (A)	0	0.5	1.0	1.5	2.0	2.5	3.0
Restoring force (arbitrary units)	0	0.24	0.50	0.77	1.00	1.37	1.60

Such results suggest that the magnetic flux density within the solenoid is directly proportional to the current flowing.

In Unit 6 we saw that a change in the current (I_p) flowing in a coil which was placed within a second coil would induce an e.m.f. in this second coil. This e.m.f. depended on the rate of change of magnetic flux through it. Our coil surrounded the primary coil so that we were sure that all the flux within the primary coil passed through it.

This enables us to design a far more convenient technique for the exploration of magnetic fields. Consider a small search coil (Fig. 37.3) mounted so that it may be inserted into a solenoid with the axes of the two coils parallel to each other. If now an alternating current $I_0 \sin 2\pi ft$ is passed through the solenoid, an alternating magnetic field will be set up, and we may write

$$B = B_0 \sin 2\pi ft.$$

An alternating e.m.f. will be developed in the search coil and since this will be proportional to the rate of change of flux, we may write

$$\mathscr{E} = \frac{d}{dt}(kB_0 \sin 2\pi ft),$$

where k depends on the dimensions of the search coil. (We shall see later that $k = NA$ where N is the number of turns and A is the cross-sectional area.) Hence

$$\mathscr{E} = kB_0 2\pi f \cos 2\pi ft.$$

This alternating voltage has a peak value which is directly proportional to the peak value of the magnetic flux density developed in the solenoid. It is convenient to display it and to measure it on a cathode ray oscilloscope.

We now have at our disposal a very convenient technique for exploring the magnetic field and we have a device whose sensitivity may be increased either by increasing the number of turns or the applied frequency—or both together. The search coil is usually kept small so that the field may be sampled over a small area.

37.2 EXPLORATION OF MAGNETIC FIELDS

How far do laws similar to those which apply to electric and gravitational fields apply in this case? Experimental investigations into this problem are complicated by the vector nature of the product Il. It is difficult to examine the interaction of two current

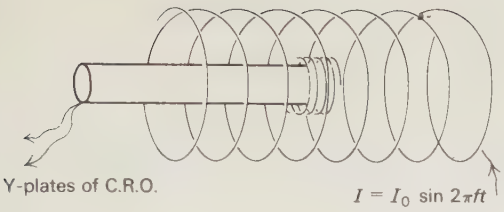


Fig. 37.3 Exploring the field within a solenoid with a search coil.

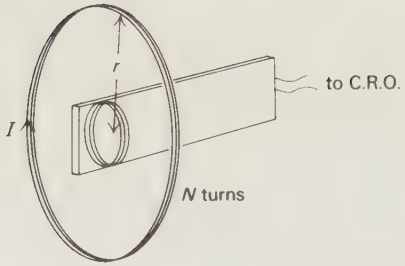


Fig. 37.4 Exploring the field of a flat coil.

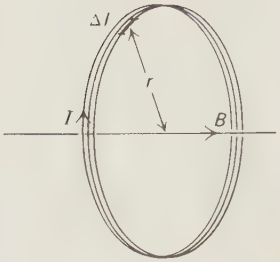


Fig. 37.5

elements with any degree of accuracy. Major difficulties arise from

a) the fields created by wires leading to the current elements,
b) the necessarily finite length of the current elements,
c) the very small forces which exist between the current elements.

However, the law describing the forces between current elements (the Biot-Savart law) can be derived from a knowledge of the fields generated by larger shapes and longer lengths of wire. We will see how this may be done in one simple case.

37.3 THE FIELD AT THE CENTRE OF A FLAT COIL

A large coil has N turns of wire, each of radius r. The co-axial search coil is mounted as shown (Fig. 37.4) and connected to the cathode ray oscilloscope. Investigation of the field at the centre of the coil, shows that

$$\begin{aligned} B &\propto I, \text{ with } N \text{ and } r \text{ constant,} \\ &\propto N, \text{ with } I \text{ and } r \text{ constant,} \\ &\propto 1/r, \text{ with } I \text{ and } N \text{ constant.} \end{aligned}$$

Thus for a coil of N turns and radius r which is carrying a current I

$$B_{\text{centre}} \propto \frac{NI}{r}$$

$$= \frac{\mu_0}{2}\frac{NI}{r}. \qquad (37.1)$$

The constant of proportionality is, in this case, written $\mu_0/2$. As with the electric force constant ε_0, the magnetic force constant μ_0 is usually to be found in association with some multiple of π. The factor (4π) is associated with the spherical symmetry of the space around a point source; (2π) with the cylindrical case and (zero π) with cases such as the one just treated.

Let us now break down the full length of wire in the coil $(= 2\pi rN)$ into small segments, each of length Δl. Each carries current I, and each segment is arranged relative to the centre like any other segment (Fig. 37.5). Each segment is perpendicular to r and will make a small contribution ΔB to the field at the centre. Our task is to find the size of that contribution.

In the whole coil of total length $2\pi rN$, there will be $2\pi rN/\Delta l$ contributions, each of ΔB. So

$$B = \frac{2\pi rN}{\Delta l}\Delta B.$$

Substituting for B from Eq. 37.1, we obtain

$$\frac{\mu_0 NI}{2r} = \frac{2\pi rN}{\Delta l}\Delta B$$

$$\therefore \Delta B = \frac{\mu_0 NI\,\Delta l}{2r2\pi rN}$$

$$= \frac{\mu_0}{4\pi}\frac{I\,\Delta l}{r^2}.$$

This is an inverse square law for the current element $I\Delta l$. The equation can only be applied if $r \gg \Delta l$. In the limit this becomes

$$dB = \frac{\mu_0}{4\pi}\cdot\frac{I\,dl}{r^2}.$$

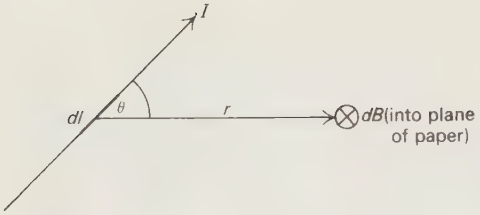


Fig. 37.6

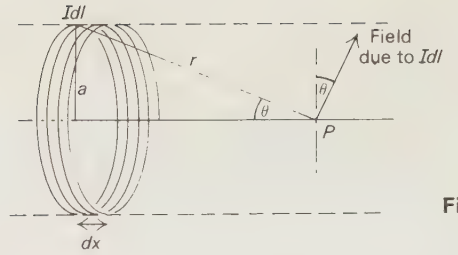


Fig. 37.7

The law is not complete in this form because we have chosen r so that its direction is perpendicular to the direction of Δl. If, as in Fig. 37.6, r makes an angle θ with Δl which is different from 90° we must write

$$dB = \frac{\mu_0}{4\pi} \cdot \frac{I\, dl}{r^2} \sin\theta. \qquad (37.2)$$

This is known as the *Biot-Savart law* and it is much more complex in form than the forms of inverse square law which apply to the cases of the electric and gravitational fields. Nevertheless it suggests that the flux model may still be used and this will be valuable when we come to consider induced e.m.f.'s.

37.4 THE MAGNETIC FIELD OF A CURRENT-CARRYING SOLENOID

A solenoid is a coil which is long compared with any measurement made across a cross-section. Where this ratio is particularly large (say more than 5 to 1) we usually refer to the coil as a long solenoid.

We have already started an exploration of the field of such a solenoid using a current balance. We have found that in regions inside the solenoid remote from either end, the field is uniform.

A current balance is too large an instrument for exploring the field in any detail, but a search coil may be used to explore the field of a similar solenoid carrying an alternating current. Such an investigation confirms that the field is indeed uniform over the cross-section of the coil as well as along its length in regions close to the centre of the coil.

What factors affect the size of this field? We already know that

$$B \propto I.$$

We might expect that it would also depend on the cross-sectional area of the coil and on the number of turns of wire. However, experiment quickly shows that B is independent of A, the cross-sectional area, but is proportional to N/l, the number of turns per unit length. Hence

$$B \propto \frac{N}{l} I.$$

What is the constant of proportionality? We have shown that for a small current element Idl, a distance r from the point at which the field dB is measured

$$dB = \frac{\mu_0 I dl}{4\pi r^2}.$$

We can use this to find the field at the centre of a long solenoid.

First we divide up the solenoid into a sequence of flat coils of length dx as in Fig. 37.7. The field at P due to a small current element Idl at Q is

$$\frac{\mu_0 I dl}{4\pi r^2}$$

in the direction shown. The field along the axis of the solenoid is thus

$$\frac{\mu_0 I dl}{4\pi r^2} \sin\theta.$$

For a complete loop

$$\sum dl = 2\pi a,$$

where a is the radius of the (circular) solenoid. Consequently the field at P due to one complete turn of wire is

$$\frac{\mu_0 I \cdot 2\pi a}{4\pi r^2} \sin\theta.$$

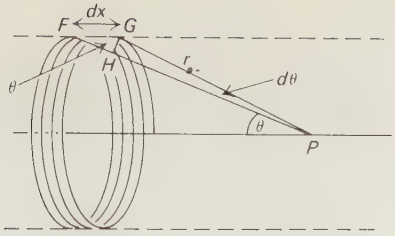


Fig. 37.8

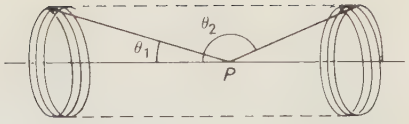


Fig. 37.9

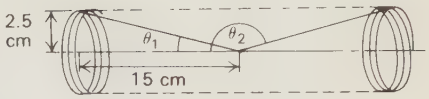


Fig. 37.10

The components of the field perpendicular to the axis will cancel each other out.

The coil length is dx, so it contains $(N/l)dx$ turns. The field at P due to the flat coil is

$$dB_P = \frac{\mu_0 I \cdot 2\pi a}{4\pi r^2} \frac{N}{l} dx \sin\theta. \qquad (37.3)$$

To find B_P we need to add up the field contributions of all the flat coils of which the solenoid is made. To do this we simplify dB_P as follows.

The small length dx subtends a small angle $d\theta$ at P (see Fig. 37.8). In the triangle FGH, right-angled at H,

$$FG = dx \quad GH = rd\theta \quad \text{and} \quad GH/FG = \sin\theta.$$

So

$$dx = rd\theta/\sin\theta.$$

But

$$a/r = \sin\theta.$$

Combining

$$dx = r^2 d\theta/a.$$

Substituting for dx in the formula for dB_P (Eq. 37.3) and simplifying gives

$$dB_P = \frac{\mu_0(N/l)I}{2} \sin\theta \cdot d\theta.$$

To find the contributions of all the flat coils we must integrate this expression:

$$B_P = \int_{\theta_1}^{\theta_2} \frac{\mu_0 NI}{2l} \sin\theta \, d\theta.$$

The angles θ_1 and θ_2 between which we perform the integration are the angles subtended at P by the ends of the solenoid (see Fig. 37.9).

If we assume that the solenoid is very long and that P is very distant from both ends, $\theta_1 \approx 0$ and $\theta_2 \approx \pi$. So

$$\begin{aligned} B_P &= \int_0^{\pi} \frac{\mu_0 NI}{2l} \cdot \sin\theta \, d\theta \\ &= \frac{\mu_0 NI}{2l} [-\cos\theta]_0^{\pi} \\ &= \frac{\mu_0 NI}{2l} \cdot 2 \\ &= \frac{\mu_0 NI}{l}. \end{aligned} \qquad (37.4)$$

This expression is only precisely valid for an infinitely long solenoid, but it is a good approximation for the field at the centre of a long but finite solenoid, with N/l turns per unit length. To see how good an approximation it is, consider a solenoid which is 5 cm wide and 30 cm long (Fig. 37.10). In this case we see from the figure that

$$\begin{aligned} B_P &= \int_{\theta_1}^{\theta_2} \frac{\mu_0 NI}{2l} \sin\theta \cdot d\theta \\ &= \frac{\mu_0 NI}{2l} [-\cos\theta]_{\theta_1}^{\theta_2} \\ &= \frac{\mu_0 NI}{2l} \cdot (\cos\theta_1 - \cos\theta_2). \end{aligned}$$

When P is at the centre of the solenoid

$$\cos\theta_1 = -\cos\theta_2$$

and

$$B_P = \frac{\mu_0 NI}{2l} \cdot 2\cos\theta_1.$$

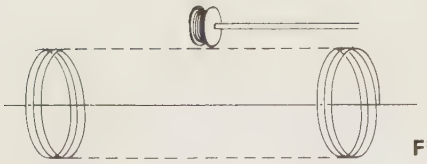


Fig. 37.11

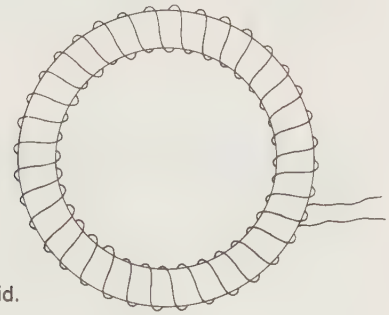


Fig. 37.12 A toroid.

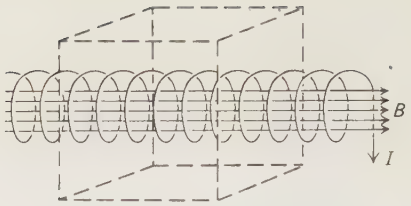


Fig. 37.13

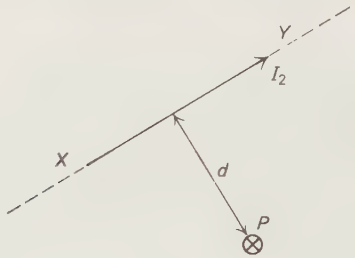


Fig. 37.14 B_P into plane of paper

From Fig. 37.10

$$\cos\theta_1 = \frac{15}{\sqrt{(15^2 + 2.5^2)}} = 0.985$$

so

$$B_P = 0.985\,\mu_0 \frac{N}{l} \cdot I.$$

This is only $1\frac{1}{2}\%$ lower than the field would have been had the solenoid been infinitely long.

37.5 THE FIELD OUTSIDE A LONG SOLENOID

A solenoid such as the one just described can typically carry a current of 5 amps a.c. With such a current flowing, a search coil placed inside the solenoid was found to have a p.d. of 6.5 V across it using a CRO.

With the same search coil placed outside the solenoid (Fig. 37.11) the p.d. was only 0.16 V.

Thus the field immediately outside the solenoid was only 2.5% of the field inside. Farther away, the field was even smaller. To a good approximation, the field outside the solenoid can be called zero. In fact this approximation turns out to be the same as calling the field inside the solenoid $\mu_0 NI$. Both approximations become exact statements for infinitely long solenoids, and for toroids, which are solenoids bent into a circle so that there are no free ends (Fig. 37.12).

37.6 THE FLUX MODEL AND MAGNETIC FIELDS

Consider the solenoid of Fig. 37.13. It is assumed to be infinitely long. We may place a closed surface around some part of it as shown. How much magnetic flux leaves the closed surface? As much as enters it. This is true no matter how the surfaces are drawn and we may say that the net flux passing through the closed surface is always zero.

37.7 THE VALUE OF μ_0 (THE MAGNETIC FORCE CONSTANT)

The value of the magnetic force constant μ_0 is directly determined by our definition of the ampere. To see how this arises we will quote without proof the result of applying the Biot-Savart law to an infinitely long straight wire (see Fig. 37.14) in which a current I_1 is flowing. In such a case

$$B_P = \frac{\mu_0}{2\pi} \cdot \frac{I_1}{d}$$

as we shall see in Section 41.3. If a second wire runs parallel to the first at a distance d and carries a current I_2 the force per unit length (from Eq. 33.7) is given by

$$F = BIl$$

and, in this case

$$F = B_P I_2 l.$$

Substituting for B_P

$$F = \frac{\mu_0}{2\pi} \cdot \frac{I_1 I_2}{d}.$$

If now I_1 and I_2 are each 1 A and d is 1 m, then, by definition (Section 33.11) this force is 2×10^{-7} N. Consequently

$$2 \times 10^{-7} = \frac{\mu_0}{2\pi}$$

and

$$\mu_0 = 4\pi \times 10^{-7} \text{ N A}^{-2}.$$

The unit in which μ_0 is quoted is expressed in many forms. You should check that the N A^{-2} used above is the same as the $\text{J}^{-1}\,\text{C}^2\,\text{m}^{-1}$ recommended by the International Union of Pure and Applied Physics and as the henry per metre (H m^{-1}) quoted in many texts.

It is customary to call this magnetic force constant the *permeability of free space*.

Chapter 38
POTENTIAL ENERGY IN FIELDS OF FORCE

38.1 Introduction
38.2 Gravitational potentials in the field around the Earth
38.3 Potential changes and potential gradients
38.4 Potential at a point in the Earth's gravitational field—the general case
38.5 Self-energy
38.6 A potential well
38.7 The special case of circular motion

38.1 INTRODUCTION

The concept of field strength was introduced so that we could describe a measure of the force that a field might exert on a mass, a charge or a current element placed at any point within it. This description was independent of any particular mass or charge or current element. It was a description of a general property of the field.

Any object which is under the influence of a force (i.e., existing in a force field) possesses potential energy.

For example, in discussing the case of the gravitational potential energy of a mass at rest at some height h above an arbitrary zero (for example, the floor) we have seen that this energy is expressed by the product mgh.

Just as we defined gravitational field strength at a point as the force experienced by unit mass placed at the point in question, we may now define a *gravitational potential* at a point as the potential energy possessed by unit mass placed at that point. Taking the surface of the Earth as an arbitrary zero (Fig. 38.1), the gravitational potential (V_G) at a height h is gh.

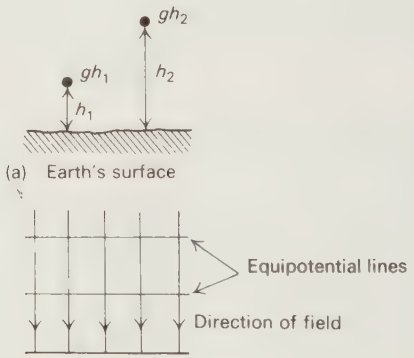


Fig. 38.1

In this example, points with the same gravitational potential can be connected by a series of lines parallel to the Earth's surface. These are called *equipotential lines*. In three dimensions, points with the same potential all lie in surfaces, called *equipotential surfaces*.

What is the shape of a gravitational equipotential surface close to the Earth's surface?

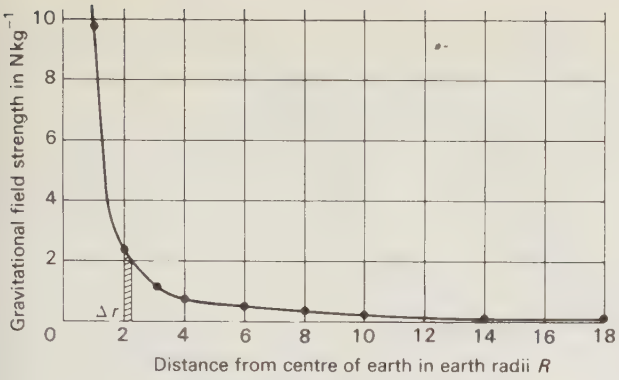


Fig. 38.2 The gravitational field strength in the vicinity of the earth.

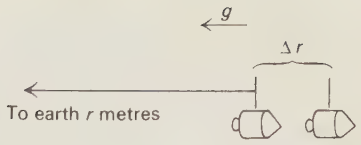


Fig. 38.3

38.2 GRAVITATIONAL POTENTIALS IN THE FIELD AROUND THE EARTH

The calculations of potential in a uniform field are easy enough. But practical problems may not be so simple.

Consider a space craft of mass m. How much energy must be given to it in order to get it clear of the pull of the Earth? If the Earth's gravitational field strength remained constant at all distances from the Earth's centre, the question would have no answer. But the field is not uniform; it diminishes according to an inverse square law. This is shown in Fig. 38.2 in which the gravitational field strength g is plotted against distance from the Earth measured in Earth radii.

When the spacecraft is 9 Earth radii ($9R$) from the Earth's surface (about 60 000 km—a sixth of the way to the Moon) the gravitational force has fallen to about 1% of its value at the surface. The pull of the Earth has fallen to about 0.1 N for each kilogramme of spacecraft. Compared with the amount of energy already utilized in getting this far, very little more is required to increase the distance further.

At a distance of $19R$ (one third of the way to the moon) the pull of the Earth is only 0.25% of its value at the surface.

Ultimately, the spacecraft can get so far away that the gravitational pull due to the Earth is smaller than any stated amount. We may then say that the spacecraft is infinitely remote from the Earth and that no further energy has to be provided to free it from the gravitational pull of the Earth.

The question is, how much energy has to be supplied to get the spacecraft into such a position?

Consider the spacecraft at a distance r from the Earth's centre (Fig. 38.3). At this distance the field intensity g is given by

$$g = G\frac{M}{r^2}.$$

In order to move the craft a further distance Δr along the radius, additional energy has to be transformed to gravitational potential energy. The energy gained per unit mass, $\Delta E_P = g\,\Delta r$.

We can calculate this with the help of the graph (Fig. 38.2) of field strength against distance. At the distance of $2R$ indicated $g = 2\ \text{N kg}^{-1}$. If we take $\Delta r = 0.2R$, ΔE_P is given by the area under the graph (shaded in the figure) between values of r of $2R$ and $2.2R$.

To extend the calculation out to 20 Earth radii would involve us in 100 such computations. A less accurate result may be obtained by taking a larger value for Δr. We will take $\Delta r = R$, a large increment indeed.

Table 38.1 shows the first four summations for $g\,\Delta r$. You should complete it as far as $r = 19.5R$ and plot the graph.

The graph of the variation of the potential energy per unit mass of the spacecraft (which is the

Table 38.1

Average value of r	Value of g at r in N kg^{-1}	Δr in m	$g\,\Delta r$ in joules	Total energy given to craft to reach r in joules
$1.5R$	4.3	R	$4.8R$*	$4.8R$
$2.5R$	1.6	R	$1.7R$*	$6.5R$
$3.5R$	0.8	R	$0.8R$	$7.3R$
$4.5R$	0.5	R	$0.5R$	

* From a careful estimate of the areas under the curve.

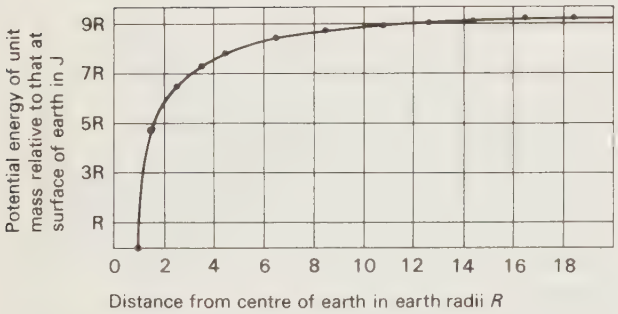


Fig. 38.4 The potential of a unit mass in the vicinity of the earth.

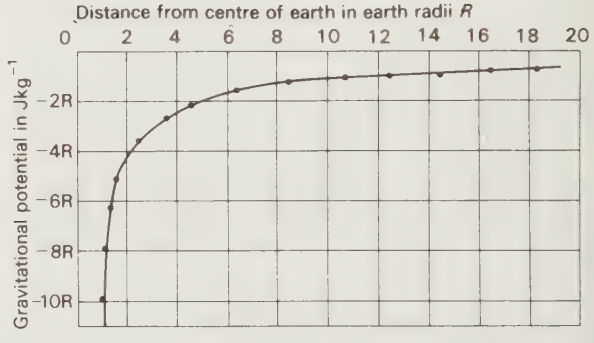


Fig. 38.5 Gravitational potential in the vicinity of the earth.

last column of figures) plotted against its distance, r, from the Earth is shown in Fig. 38.4.

To get the spacecraft 19 Earth radii away we have to provide a total of $9.3R$ joules for each kilogramme of mass. For a load of, say, 5000 kilogrammes (about the mass of an Apollo command module), and taking R as 6.4×10^6 m, the total energy provided must be

$$5000 \times 9.3 \times 6.4 \times 10^6 \text{ J} = 2.98 \times 10^{11} \text{ J}.$$

You should calculate the velocity which the craft would acquire were it now to fall back to the Earth's surface.

The spacecraft requires still more energy to get completely free of the Earth's pull. The potential energy-distance graph does not seem to be reaching a limit yet. In fact it approaches the limiting energy very slowly. If we go on with our calculations, it turns out that the limit is $9.8R$ joules—but we should have to make calculations to one hundred Earth radii from the Earth to get within 1% of this value.

This graph of potential energy per unit mass against distance (Fig. 38.4) is also a representation of the way in which the potential at points in the Earth's field changes with distance. The potential is measured relative to the Earth's surface taken as an arbitrary zero. It would be more rational to transfer the zero to some point far out in space where we could agree that no additional energy is required to move the mass further away from the Earth. We shall, in future, refer gravitational field potentials to this position.

To change from the former system to this one, we need to subtract $9.8R$ joules from the numbers we have calculated. This gives the graph of Fig. 38.5. The potential of each point is, of course, negative and this is characteristic of potentials when the forces are attractive in nature. Forces of repulsion give rise to positive potentials.

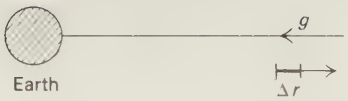


Fig. 38.6

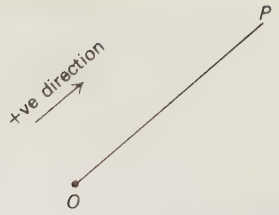


Fig. 38.7

38.3 POTENTIAL CHANGES AND POTENTIAL GRADIENTS

a) Sign convention

A body which is moving in a gravitational field of force may lose or gain potential energy. Whenever the body is displaced in the direction of the force which is exerted on it in the field, it will lose potential energy, changing this energy to some other form.

If the body, assumed to have unit mass, moves through a distance Δr in the direction of the force F due to the field, then the resultant change in the potential energy is given by

$$\Delta E_p = - F \Delta r,$$

and this defines the change in potential, written ΔV.

The change in potential is negative (representing a loss) when the displacement Δr is in the same direction as the force.

If the displacement is in the opposite direction to that in which the force acts, then

$$\Delta V = - F(- \Delta r) = F \Delta r.$$

The change in potential is now positive (representing a gain).

The expression

$$\Delta V = - F \Delta r, \qquad (38.2)$$

is a general result, applying to all fields. F is the appropriate field strength. To get the correct sign for the change in potential when a body moves in a field of force we must pay due regard to the convention which gives rise to this formula.

As an example let us consider a unit mass which is moved away from the centre of the Earth. Its potential energy and its potential both increase. Or, looked at from the point of view of a zero which is placed at infinite distance, the potential energy and the potential both become less negative. We may apply our formula (38.2).

The change in potential, ΔV_G, in moving through a positive displacement Δr (away from the Earth, in the direction of increasing r) Fig. 38.6, is given by

$$\Delta V_G = - F \Delta r,$$
$$= - (- g) \Delta r.$$

In this last step g is negative because it is in the direction of decreasing r. And so

$$\Delta V_G = g \Delta r.$$

This is positive and represents an increase in ΔV_G as we expected.

If, now, the mass moves towards the Earth, Δr is negative

$$\Delta V_G = - (- g)(- \Delta r),$$
$$= - g \Delta r.$$

This is negative and represents a decrease in ΔV_G as we expected.

Examples

a) Suppose a field intensity g is in the direction PO in Fig. 38.7, where the direction OP is regarded as the positive direction. How does the potential change as we move away from O?

$$\Delta V_G = - (- g)(+ \Delta r),$$
$$= + g \Delta r.$$

So the potential increases.

b) Suppose now that the field intensity is in the direction OP in Fig. 38.7. How does the potential change as we move from P towards O?

$$\Delta V_G = -(+g)(-\Delta r),$$
$$= +g\,\Delta r.$$

So the potential increases.

c) Energy is required to move a body from P towards O. What is the direction of the field?

In this case the change in potential is to be positive whilst Δr is negative. So the product $-(g)(-\Delta r)$ must be positive. It follows that g must be positive, and so in the direction OP.

b) Potential gradient

We have, in the general case of a gravitational field,

$$\Delta V_G = -g\,\Delta r.$$

So

$$g = -\frac{\Delta V_G}{\Delta r}.$$

In the limit as

$$\Delta r \to \text{zero}$$

$$g = -\frac{dV_G}{dr}. \qquad (38.3)$$

The quotient dV_G/dr is known as the *potential gradient* at the point chosen. We see that the field intensity at a point is equal to the negative potential gradient at that point. Or, conversely, we may define potential V_G in terms of the equation thus:

$$dV_G = -g\,dr.$$

You will have noticed that we have deliberately chosen simple cases, always measuring the displacement and the field intensity in the same line. It is not necessary to do this; the same results are obtained but, in place of g, we must use the component g_r in the direction of r.

Precisely the same argument may be applied to other cases of force fields and the relevant potentials and potential gradients arrived at in exactly the same way.

38.4 POTENTIAL AT A POINT IN THE EARTH'S GRAVITATIONAL FIELD—THE GENERAL CASE

As a unit mass moves through a distance Δr away from the Earth the change in potential is given by

$$\Delta V_G = -g\,\Delta r.$$

At a distance r

$$g = -\frac{GM}{r^2}.$$

So

$$\Delta V_G = \frac{GM}{r^2}\Delta r.$$

In the limit

$$\text{as } \Delta r \to 0$$

$$dV_G = \frac{GM}{r^2}dr.$$

The total energy change in moving from $r = r_1$ to $r = r_\infty$ is

$$\int_{V_G}^{0} dV_G = \int_{r_1}^{\infty} \frac{GM}{r^2}\cdot dr$$

$$[V_G]_{V_G}^{0} = GM\left[-\frac{1}{r}\right]_{r_1}^{\infty}$$

$$0 - V_G = -\left(0 - \frac{GM}{r_1}\right)$$

$$\therefore V_G = -\frac{GM}{r_1}.$$

If the mass moves from a point distant r from the centre of the Earth to infiinity

$$V_G = -\frac{GM}{r}. \qquad (38.4)$$

Since, when $r =$ the radius of the Earth R

$$\frac{GM}{R^2} = 9.8 \text{ N kg}^{-1}$$

$$GM = 9.8R^2$$

and the gravitational potential at a distance r from

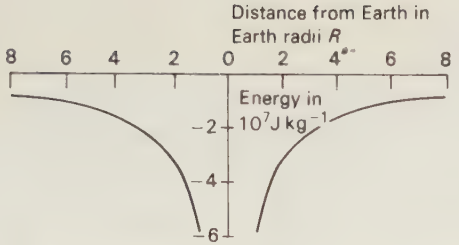


Fig. 38.8 The gravitational potential well of the earth.

the centre of the Earth may be written

$$V_G = -\frac{9.8R^2}{r} \qquad (38.5)$$

which gives V_G in J kg^{-1} when R and r are in metres.

Problem 38.1 You should plot a graph of the potential V_G against r, writing r in terms of whole number multiples of the radius of the Earth R. How does this graph compare with the one obtained arithmetically? Account for any differences between them.

At the Earth's surface, the gravitational potential is $9.8R$ joules per kilogramme. This is the energy which must be given to each kilogramme of the spacecraft in order to free it from the pull of the Earth.

Using the fact that the radius of the Earth is 6.4×10^6 m, calculate the gravitational potential energy of the spacecraft as it rests on the Earth's surface. Suppose the spacecraft were to return to the Earth from an infinite distance where the gravitational potential energy due to the Earth's field was zero. Assuming that all this energy was converted to kinetic energy, how fast will it be travelling when it reaches the Earth's atmosphere? What sensible assumption will you make about the thickness of the atmosphere?

38.5 SELF-ENERGY

The realization that a mass on the surface of the Earth has a certain (negative) potential energy relative to points at a great distance from the Earth leads us to see that each of the individual masses which go to make up the Earth has a negative gravitational energy relative to the rest of the mass of the Earth. This self-energy is very high for a mass like the Earth. It is the energy that would be needed to break up the Earth and distribute all the pieces to distant parts of the universe. Calculations about the gravitational potential of masses in the universe due to all the other masses figure prominently in theories of the evolution of the universe. (See also Problem 38.3 at the end of this chapter.)

38.6 A POTENTIAL WELL

A spacecraft, almost stationary a long way from Earth, is attracted towards it. It falls towards Earth, its kinetic energy increasing as its potential energy decreases, just like a pail of water will fall down a well, if the rope to which it is attached, breaks. This analogy has led to the variation of potential in an attractive force field being referred to as a *potential well*. Figure 38.8 is a graph of the Earth's potential well.

A body in a potential well will be acted on by an attractive force towards the centre of the well. We have so far only considered what will happen to a spacecraft held stationary a long way from the Earth. Under these conditions, when the restraining force is removed, the spacecraft falls towards the Earth with increasing speed. Is this the only possible motion for a mass in a gravitational potential well?

Here is an experiment you can do, to find out. Set up a large funnel (at least 30 cm in diameter at the mouth) so that its axis is vertical as shown in Fig. 38.9. The funnel forms a potential well. Hold a ball-bearing at the edge and then release it. It runs down the funnel (and out through the central hole, if it is small enough). This is something like the spacecraft hurtling in towards the Earth. Now hold the ball-bearing some way inside the lip and give it a small impulse in a direction making an angle with the line to the centre (Fig. 38.9b).

The ball no longer drops down to the centre but moves on an ellipse. Sometimes the ball is close to the centre; sometimes it is far away.

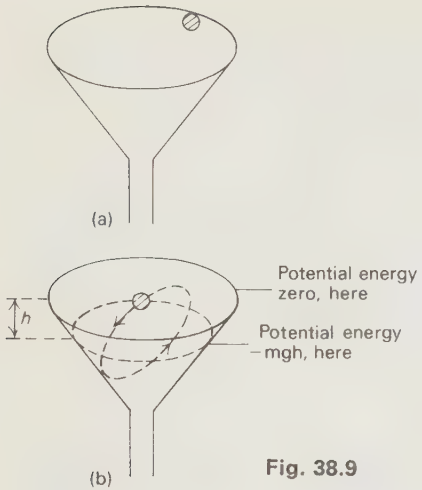


Fig. 38.9

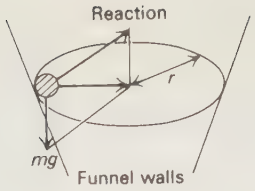


Fig. 38.10

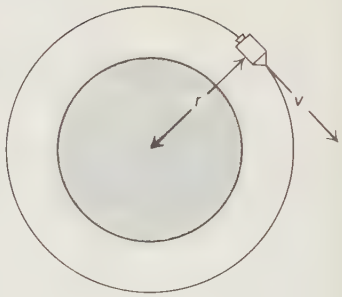


Fig. 38.11

When the ball was given its initial impulse, it had a total energy E.

$$E = E_p + \tfrac{1}{2}mv^2.$$

E_p is its potential energy. This is negative if we reckon potential energies at the lip to be zero.

E must remain constant throughout the motion. (Why?)

Thus the lower the ball the more negative is E_p and so the larger is $\frac{1}{2}mv^2$. Do your observations confirm this?

What can you say about the size of E_p compared with the size of $\frac{1}{2}mv^2$ if the ball is to stay in the funnel?

38.7 THE SPECIAL CASE OF CIRCULAR MOTION

In one particular case, this ellipse becomes a circle. Can you set the ball-bearing into a circular orbit in the funnel? The forces on the ball have a resultant towards the centre of the funnel, in the plane of the orbit as shown in Fig. 38.10. Ideally the ball maintains a constant distance from the attracting centre. What in fact happens to the ball? Why does it behave like this?

In the special case of motion in a circle, we can work out the value for the total energy E for a spacecraft or satellite in the Earth's gravitational field. If the craft is moving with velocity v in a circle radius r as shown in Fig. 38.11,

$$E = E_p + \tfrac{1}{2}mv^2$$
$$= -G\frac{Mm}{r} + \frac{1}{2}mv^2.$$

From Newton's laws of motion applied to motion in a circle (see Unit 2)

$$\frac{mv^2}{r} = G\frac{Mm}{r^2}.$$

Hence

$$mv^2 = G\frac{Mm}{r}.$$

Substituting this value for mv^2 in the expression for E, we get:

$$E = -G\frac{Mm}{r} + \tfrac{1}{2}G\frac{Mm}{r}$$
$$= -G\frac{Mm}{2r}. \qquad (38.6)$$

First, notice that E is negative. This is another good reason for reckoning potentials as zero at infinitely great distance from the source of attraction. It leads

to the simple rule that for all closed orbits in a gravitational field, the total energy of the orbiting body must always be negative.

Secondly, the bigger the radius of the orbit, the bigger (less negative) is E. For a mass m on the Earth's surface, from Eq. (38.4),

$$E' = -G\frac{Mm}{R},$$

where R is the radius of the Earth and E' is the total energy.

To put a satellite into a circular orbit of radius r, it must be given additional energy ΔE.

$$\begin{aligned}\Delta E &= E - E' \\ &= -G\frac{Mm}{2r} - \left(-G\frac{Mm}{R}\right) \qquad \text{(from Eq. 38.6)} \\ &= GMm\left(\frac{1}{R} - \frac{1}{2r}\right). \qquad (38.7)\end{aligned}$$

The bigger r, the bigger must this additional energy be. For an orbit close to the Earth's surface:

$$\begin{aligned}\Delta E &= GMm\left(\frac{1}{R} - \frac{1}{2R}\right) \\ &= \frac{GMm}{2R}.\end{aligned}$$

ΔE can take any value from $GMm/2R$ to GMm/R, resulting in all possible circular orbits of radius between R and ∞.

To place a satellite in a circular orbit requires great skill, since the satellite must be travelling in just the right direction (perpendicular to the Earth's field) at just the right speed. Usually, such precision is unattainable. A small departure from the necessary conditions leads to an elliptical orbit instead.

Problem 38.2 If this additional energy is all given in the form of kinetic energy ($\frac{1}{2}mv^2$) how fast must the craft be travelling to go into a circular orbit just grazing the Earth's surface (ignoring, of course, viscous forces)?

Problem 38.3 This is a step-wise calculation leading to an important expression for the loss in gravitational potential energy when a massive body is put together from a large number of discrete bits collected together from points very distant from each other. We shall apply the resulting expression to the Earth, but it also plays an important part in theories of the nature of the Universe.

a) Write down an expression for the gravitational potential energy of unit mass on the surface of a sphere of radius r and mass M.

b) What is the potential energy lost by a mass m in coming from a long distance away to the surface of the sphere?

c) Write down an expression for M in terms of the radius r and the density ρ of the sphere and substitute this value for M in the expression derived in (b) above.

$$\left(-\frac{4}{3}G\pi r^2\rho m\right).$$

d) Suppose that the mass m is deposited on the surface of the sphere to form a shell of thickness Δr (which is much smaller than r) and density ρ the same as that of the larger sphere. Write down an expression for m in terms of r, Δr and ρ.

e) Substitute this expression in (c) above.

$$\left(-\frac{16\pi^2}{3}G\rho^2 r^4\,\Delta r\right).$$

f) If this expression is integrated from $r = 0$ to $r = R$ we can find the total potential energy lost by the matter in making up a sphere of radius R if it accumulates together from places originally far away from each other. Do the integration (it is a very easy one).

$$\left(E_{\text{total}} = -\frac{16\pi^2}{15}G^2\rho^2 R^5\right).$$

g) Since M, the total mass of the sphere is $4\pi R^3\,\rho/3$, find an expression for E_{total} in terms of G, M, and R.

$$\left(E_{\text{total}} = -\tfrac{9}{15}G\frac{M^2}{R}\right).$$

h) Let us suppose that the Earth was formed from a similar accumulation of particles. Calculate the total gravitational potential energy lost by such a process.

($M_{\text{Earth}} = 6 \times 10^{24}$ kg, $R_{\text{Earth}} = 6.4 \times 10^{6}$ m),

(-2.3×10^{32} J).

i) Suppose that the mean specific heat capacity of the Earth of 0.5×10^{3} J kg^{-1} K^{-1}. Calculate the mean temperature rise that might be expected in the formation of the Earth by this process. Comment on your result.

(7.5×10^{4} K).

Chapter 39
POTENTIAL IN THE ELECTRIC FIELD

39.1 Definitions and measures
39.2 A further look at field potentials
39.3 Energy stored in an electric field
39.4 Electric potential due to a dipole

39.1 DEFINITIONS AND MEASURES

All the ideas about potential which we have developed for gravitational fields may be used equally with electric fields.

We defined a gravitational potential V_G in terms of the equation

$$dV_G = -\,g\,dr.$$

Let us define an electric potential in terms of

$$dV = -\,E\,dr,$$

where E is the field strength.

From the equation it is clear that the units of the electric potential V must be (N/C) m or $\mathrm{N\,C^{-1}\,m}$. Since the newton metre is the joule, the unit of electric potential is the joule per coulomb ($\mathrm{J\,C^{-1}}$). We have already met this under the name of the *volt*.

In the earlier work with electric currents, the volt has been used to describe the energy associated with an electric current and, particularly, with the battery supplying the current. We have used voltmeters to measure potential differences in circuits. It is the ready availability of such instruments as these which makes the concept of electrical potential so very useful. Since the potential between any point and some chosen reference point can be measured directly, the field strength E can be deduced immediately by the application of the relationship $dV = -E\,dr$ or

$$E = -\frac{dV}{dr}. \qquad (39.1)$$

We may define the electric field strength at a point as the negative of the potential gradient at that point.

Again we observe that E is, in the general case, the component of the potential in the direction of r.

Because of the experimental importance of this relationship, field strengths are often quoted in units of volt $\mathrm{metre^{-1}}$ rather than as newton $\mathrm{coulomb^{-1}}$. This is legitimate for

$$\frac{\text{volt}}{\text{metre}} = \frac{\text{joule}}{\text{coulomb} \times \text{metre}}$$

$$= \frac{\text{newton} \times \text{metre}}{\text{coulomb} \times \text{metre}} = \frac{\text{newton}}{\text{coulomb}}.$$

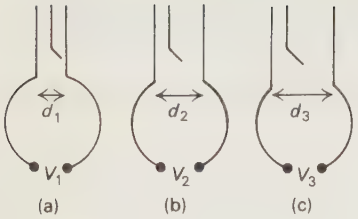


Fig. 39.1

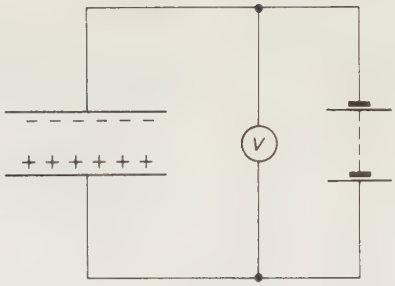


Fig. 39.2

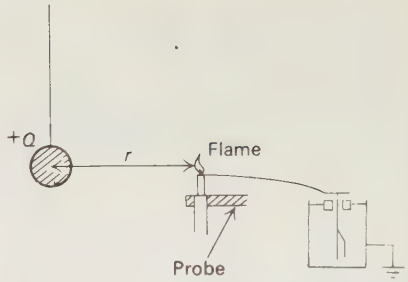


Fig. 39.3 Exploring the potential in the vicinity of a charged sphere.

The foil indicator described in Chapter 33 may be used to demonstrate the application of this approach to stating the strength of the electric field. We may use such an indicator to explore the parameters which may affect the strength of a uniform field.

Consider a pair of parallel conducting plates which are connected to a continuously variable source of electric potential up to a few kilovolts (Fig. 39.1). If the probe is charged, the deflection of the foil at a definite separation of the two plates and for a definite potential difference can be recorded (by, for example, projecting a shadow of the probe on to a screen). If now the separation of the plates is changed, it will also be found necessary to change the potential difference in order to produce the same strength of field. In a typical experiment the values given in Table 39.1 were obtained.

Evidently the ratio V/d is constant. This ratio is, of course, the potential gradient between the two plates and it provides a measure for the intensity of the field (which is uniform) between the plates

$$E = -\frac{V}{d}. \qquad (39.2)$$

Table 39.1

Separation of plates (d) in metres	0.030	0.040	0.051	0.063	0.073
Potential difference (V) in volts	2000	2700	3500	4200	5000
Ratio V/d in volts per metre	67000	68000	69000	67000	68000

Voltmeters based on galvanometers (i.e., re-calibrated milliammeters) require the continual flow of charge for their operation and may be used in measurements of field potential only when the surface potential of the conductors is maintained by some external means, as for example in Fig. 39.2.

However, there are voltmeters which depend on the accumulation of charge at two points in amounts which are proportional to the electric potential difference between these points. These include the cathode ray oscilloscope (with direct connection to the Y plates) and the gold leaf electroscope. Such instruments themselves require small amounts of charge to operate and care must be taken to ensure that they do not thus affect the field to be measured in any significant way. Provided that there is a suitable source of charge to charge an electroscope it may be used to measure the potential at a point in a field. Of course, it will be necessary to calibrate such an instrument before use. This is readily effected by applying known potential differences to it.

An experiment to investigate the variation of potential in the space near to a charged body is described by R. W. Pohl (1930).

Figure 39.3 shows a charged conducting sphere charged by contact with a power pack supplying, say,

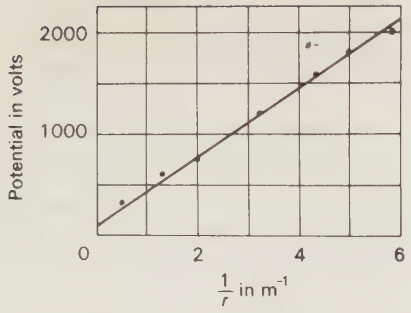


Fig. 39.4 The potential near to a charged conducting sphere plotted as a function of $1/r$.

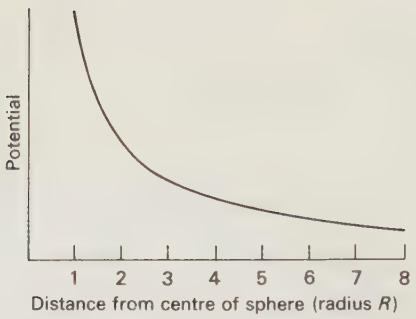


Fig. 39.5 The potential near to a charged sphere.

5 kV. The small probe is an insulated gas jet which is ignited and which then supplies a flow of equal numbers of positive and negative ions into the space. This probe is connected to the leaves of the electroscope which is held some distance away from the rest of the apparatus. The whole demonstration is performed in a space which is relatively free of such other things as pieces of furniture. It is found that the deflection of the leaf of the electroscope depends on the difference between the electric potential at P and the Earth. The Earth provides us with a practical and convenient reference point for potential for it is such a large conducting body that it is unlikely that the addition or subtraction of charge will make its surface potential measurably different from that theoretical infinitely distant point.

Experiment will show that the potential of the probe is inversely proportional to the distance from the probe to the centre of the conducting sphere. One set of results obtained with such equipment is displayed as Fig. 39.4. The line joining the points is straight but the failure of the line to pass through the origin is due to a zero error.

It is legitimate to enquire how it is that an ion source, such as the small flame, senses potential at a point. Suppose that the probe, without the flame, was brought up to the vicinity of the charged sphere from a great distance. A negative charge would appear on it and a corresponding positive charge would appear on the leaves of the electroscope since these two are connected by a conductor.

If now the flame is lit and equal numbers of positive and negative ions are produced at the probe, more negative ions will leave the probe than positive ones and the charge on the probe will be neutralized. This leaves the charge on the leaves of the electroscope. Under these conditions there is no potential difference between the probe and the electroscope, which itself indicates the potential at the probe tip.

We may return to the analogy with the gravitational case, and describe the potential at the point P in the vicinity of a charge Q by means of the equation

$$V = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r} \tag{39.4}$$

since we have already seen (Eq. 35.2) that

$$E = \frac{1}{4\pi\varepsilon_0}\frac{Q}{r^2}.$$

We notice that V is positive for positive values of Q. So energy is gained from the test charge, $+q$, as it is moved away from Q. The graph of V, the potential of a charged sphere of radius R, against r the distance from the centre has the form shown in Fig. 39.5.

If Q is negative the potential energy curve will have the same form as the gravitational potential energy curve. In this case energy will be needed to separate the two charges. This is the case with an electron and a proton forming a hydrogen atom. Assuming that, on average, the electron is 10^{-10} m away from the proton and that the charge on the

proton is 1.6×10^{-19} C, the potential at a distance of 10^{-10} m is

$$\frac{e}{4\pi\varepsilon_0 r} = \frac{1.60 \times 10^{-19} \times 9 \times 10^9}{10^{-10}} \approx 14 \text{ volts.}$$

The electron must have an amount of energy equal to eV joules or $1.60 \times 10^{-19} \times 14$ J, i.e., about 23×10^{-19} J. This is often expressed in units known as *electron volts* (eV).

In accordance with our convention that the zero of potential lies at an infinite distance from the proton, the energy of the electron must be negative. The electron at a distance of 10^{-10} m from the proton has a potential energy equal to -23×10^{-19} J (or -14 eV).

If we assume that the electron is in motion around the proton, it will also have some kinetic energy $\frac{1}{2}mv^2$. If it moves in a circle of radius r, the force towards the centre is mv^2/r, provided by the electrical force of attraction, $e^2/4\pi\varepsilon_0 r^2$.

$$\begin{aligned} mv^2/r &= e^2/4\pi\varepsilon_0 r^2, \\ \tfrac{1}{2}mv^2 &= \tfrac{1}{2}(e^2/4\pi\varepsilon_0 r), \\ &= \tfrac{1}{2} \times 23 \times 10^{-19} \text{ J}, \\ &= 11.5 \times 10^{-19} \text{ J}. \end{aligned}$$

So the electron needs an additional 11.5×10^{-19} J in order to escape from the atom (ionization). This energy can be checked experimentally. This is a very rough calculation—the answer is clearly very sensitive to the value taken for r. However, it is close enough to the measured value to give us some confidence in the validity of the calculation.

We must not, however, assume that this gives a complete picture of atomic structure. As we shall later see in this Unit, such an accelerating electron, ought to lose energy continuously in the form of electromagnetic radiation—and thus collapse into the proton.

Calculations such as this suggest that the inverse square law is valid right down to a separation between the two charges of 10^{-10} m. Indeed calculations involving nuclear energies suggest that it remains valid down to distances of at least 10^{-14} m.

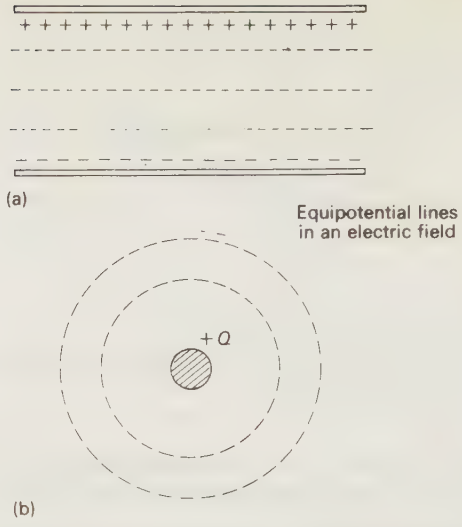


Fig. 39.6 Equipotential lines in an electric field.

On the other hand, inverse square law results have been used in astronomical distances. We may say that the inverse square laws have proved useful over a range of at least 1 to 10^{33}.

39.2 A FURTHER LOOK AT FIELD POTENTIALS

So far we have assumed implicitly that energy changes within a field take place along the direction of the field strength. What will happen if movement takes place along any path? To discuss this we shall find it convenient to use the ideas of the equipotential line and surface (see Section 38.1). Figure 39.6 shows equipotential lines in (a) a uniform field and (b) an inverse square law field.

Consider the paths shown in Fig. 39.7(a). The field between the two plates is uniform.

How much energy will be gained if a test charge is moved from A to B? We can break the path down into any number of zig-zag steps (Fig. 39.7b) which either go along the field direction or along an equipotential surface. Along the latter, no changes in energy take place. So the net change in energy will be exactly the same as though the test charge had moved in the direction of the field from one equipotential line through A to an equipotential line through B.

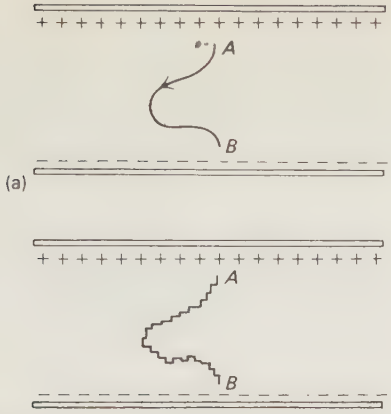


Fig. 39.7

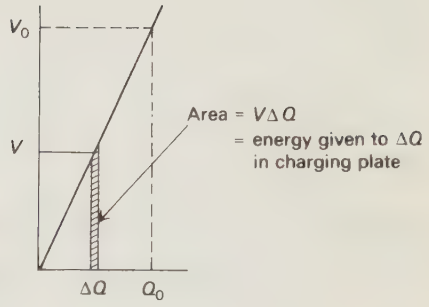


Fig. 39.8

From this, it follows that if the test charge returns from B to A by *any* path the energy lost will be identical with the energy gained in moving from A to B. Thus the potential at any point in electric or a gravitational field has a unique value relative to any chosen zero. It does not depend on the path a charge takes to reach it. Fields which behave like this are called *conservative* (see Unit 2).

We are now in a position both to describe and to measure the electric field in the space around a charged conductor by two means: (a) in terms of the charge and its distribution relative to the point under consideration (b) in terms of the potential gradient at the point in the field which is being considered.

As an example, let us consider once again the uniform field between a pair of parallel plates. Let us suppose that the area of a plate is A, that their separation is d and that one carries a charge $+Q$ whilst the other carries $-Q$. If the field strength between the plates is E we have, from Eq. (35.4)

$$E = \frac{Q/A}{\varepsilon_0} \quad \text{or} \quad E = \frac{\rho}{\varepsilon_0}.$$

On the other hand we may obtain a value for the field strength from a measurement of the potential gradient between the plates using an experimental arrangement similar to that shown in Fig. 39.2. In this case, if the measured potential difference is V and the plate separation remains d

$$E = \frac{V}{d}.$$

Equating these two values for E gives

$$\frac{\rho}{\varepsilon_0} = \frac{V}{d}$$

and hence a value for the permittivity may be found from the equation

$$\varepsilon_0 = \rho \frac{d}{V} = \frac{Q}{A} \cdot \frac{d}{V}. \tag{39.5}$$

39.3 ENERGY STORED IN AN ELECTRIC FIELD

The concept of field potential has so far been used as a device to give a measure of the energy a charge or mass would have if it were at that point. But the energy concept can be taken further. Suppose a pair of plates are separated by a fixed distance d and joined to a battery through a resistor. (The resistor is there merely to slow up the charging rate.)

A graph of charge on the plates against potential difference across them is shown in Fig. 39.8. V is proportional to Q (see Chapter 28). This is a general property of all conductors. To bring about this

charging process, the battery has to supply a total charge Q_0. When the plates are uncharged, this requires no energy expenditure. But as the plates approach the fully-charged condition the last bit of charge ΔQ has to "overcome" a potential difference of V_0. The battery has to provide ΔQ with energy $V_0\,\Delta Q$.

How much energy does the battery supply in all? It is the area under the V/Q graph (Fig. 39.8) which is $\frac{1}{2}Q_0V_0$.

What has happened to this energy? It is stored in the field E between the plates and, from Eq. (35.4)

$$E = \frac{\text{charge density}}{\varepsilon_0} = \frac{Q/A}{\varepsilon_0}.$$

Now

$$E = \frac{V_0}{d},$$

hence we may write

$$Q_0 = \varepsilon_0 EA, \qquad \text{and} \qquad V_0 = Ed.$$

So energy stored in field

$$= \tfrac{1}{2}Q_0V_0,$$
$$= \tfrac{1}{2}(\varepsilon_0 EA)\,(Ed),$$
$$= (\tfrac{1}{2}\varepsilon_0 E^2)\,Ad.$$

Now Ad is the total volume occupied by the field E. We may thus describe the quantity

$$\tfrac{1}{2}\varepsilon_0 E^2, \tag{39.6}$$

as the energy density of the electric field.

It gives the measure of the energy stored within the field—potential energy which may eventually be converted to some other form. This measure of energy density has been established for a very simple uniform field. The expression $\frac{1}{2}\varepsilon_0 E^2$ also applies to non-uniform fields. In this case, the energy density varies from point to point in the field.

39.4 ELECTRIC POTENTIAL DUE TO A DIPOLE

We are considering this example because the dipole field is very important in relation to many molecules.

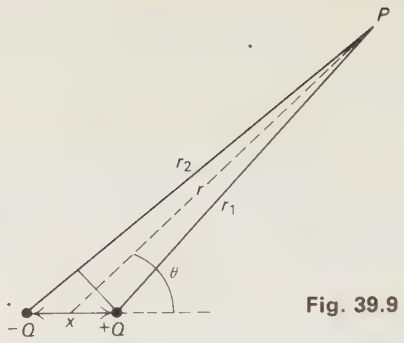


Fig. 39.9

A dipole consists of two equal, but opposite charges $\pm Q$ separated by a distance, x (Fig. 39.9). The potential at a point P, distance r_1 from $+Q$ and r_2 from $-Q$ is given by

$$V = \frac{Q}{4\pi\varepsilon_0 r_1} - \frac{Q}{4\pi\varepsilon_0 r_2}$$
$$= \frac{Q(r_2 - r_1)}{4\pi\varepsilon_0 r_1 r_2}.$$

If x is much smaller than r_1 and r_2 we may, in this case, write

$$r_1 = r_2 = r.$$

However, $(r_2 - r_1)$ is not negligible and we may write

$$r_2 - r_1 = x\cos\theta,$$

(θ being measured relative to the line from $-Q$ to $+Q$). Thus

$$V_{\text{dipole}} = \frac{Qx\cos\theta}{4\pi\varepsilon_0 r^2}. \tag{39.7}$$

Qx is called the *dipole moment*. Any potential described by a $\cos\theta/r^2$ law is called a *dipole field* and a corresponding dipole moment can be calculated for its source, regardless of the exact distribution of charges which go to make up the dipole.

Problem 39.1 A pair of metal plates 25 cm long and 25 cm wide are separated by a gap 1.4 mm thick. The lower plate is earthed and the upper one is raised to a potential of 200 V above earth. On discharging the plates to a capacitor which is connected across

an electrometer, they were found to have carried 9.4×10^{-3} C of charge. Calculate a value for the permittivity of free space assuming that the plates were in vacuo. You should first calculate the charge density and then the potential gradient. Comment on your value for ε_0.

Solution The field between the plates may be written in two ways using either Eq. (35.4) or (39.2). Equating the two values for the field intensity E we have

$$\frac{Q/A}{\varepsilon_0} = \frac{V}{d}$$

and

$$\varepsilon_0 = \frac{Q}{A} \cdot \frac{d}{V} = \frac{9.4 \times 10^{-3}}{625 \times 10^{-4}} \times \frac{1.4 \times 10^{-3}}{200}$$

$$= 10.5 \times 10^{-12} \text{ F m}^{-1}.$$

This value of ε_0 is 20% higher than the accepted value. The formulae used assume ideal conditions for isolated plates between which a uniform field extends right to the edges and there falls to zero. These conditions may be approached more closely if the plates are surrounded by a guard ring. This is raised to the same potential as the plates but is isolated from them. Readings of the charge on the plates are made for various values of separation and a graph plotted of charge against the reciprocal of the separation.

Alternatively, but less accurately, the effects may be allowed for by reading the value of the charge stored on the plates for various separations and for various areas of overlap. Similar data of a more extensive sort will be found in Problem 7.13 at the end of Unit 7.

Chapter 40
ELECTROMAGNETIC INDUCTION

40.1 Induction in a magnetic field
40.2 The electromotive force induced by a changing magnetic flux
40.3 The direction of the induced e.m.f.
40.4 The energy stored in a magnetic field
40.5 The self-inductance of a solenoid
40.6 An alternative form of Faraday's law of electr - magnetic induction

40.1 INDUCTION IN A MAGNETIC FIELD

A piece of wire is connected across the terminals of a sensitive galvanometer—one giving a deflection of a centimetre or more per μA (Fig. 40.1).

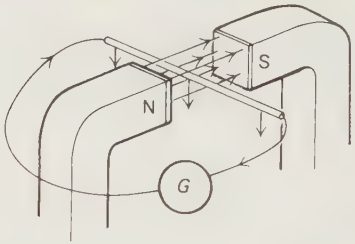


Fig. 40.1

If this wire is moved sharply up or down between the poles of a permanent magnet, the galvanometer needle is deflected momentarily. Experiment shows that the deflection is only produced when the wire moves through the field. The direction of the current depends upon the direction of the magnetic field and the direction of movement of the wire. The size of the current depends on the speed with which the wire moves through the field and the strength of the magnetic field. (See Chapter 31.)

This induced current might have been predicted from (a) the force experienced by charges moving in a magnetic field, and (b) the model of a current in a wire as a flow of free charges.

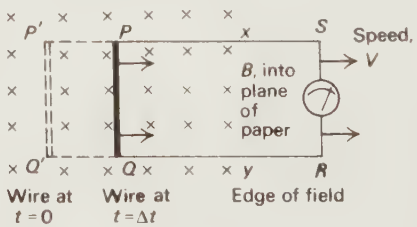


Fig. 40.2

Consider the situation shown in Fig. 40.2 in which the conductor PQ of length l carries free positive charges each of $+e$. PQ is part of a circuit $PQRS$ which moves from left to right at a speed v through a uniform magnetic field of intensity B. This field,

whose direction is perpendicular to the plane of the circuit, extends to the line XY as shown.

As each charge moves along with speed v, a force F will act on each charge in a direction perpendicular to B and also to v, i.e. along l.

$$F = +Bev. \tag{40.1}$$

By the left-hand rule, this force is directed from Q to P. This will result in a flow of positive charge in the direction Q to P. The charge will be acted on by a force F over the length of wire, l, in the magnetic field and it will consequently gain energy

$$Fl = Bevl.$$

(You should consider carefully why the force acting on charges in those parts of the circuit which are at right angles to PQ can be ignored.)

Now precisely the same effect could have been produced in the circuit in the absence of any motion through the magnetic field by inserting a source of e.m.f. into the circuit. The magnitude of this e.m.f. would have to be equal to the product Bvl to produce the same current as we now observe.

Thus the movement of the circuit through the magnetic field generates an e.m.f. in the conductor PQ of magnitude where

$$\mathscr{E} = \frac{Bevl}{e}$$

$$= Bvl.$$

Since the flow of positive charge is in the direction $PSRQ$, the end P of the wire l will be the conventional positive end, while Q is negative.

(You should repeat this calculation for the case in which the freely-moving charges are negative. Does this affect the direction of the induced electrical potential difference across PQ?)

Suppose that the circuit $PQRS$ has moved forward from $P'Q'$ to PQ, a distance Δs, in time Δt. Then

$$v = \frac{\Delta s}{\Delta t}.$$

Consequently

$$Bvl = \frac{B\,\Delta s\,l}{\Delta t}.$$

Inspection of the diagram shows that

$$B\Delta sl = B(\text{Area } P'Q'YX - \text{area } PQYX).$$

It is, in other words, the change in the product of (the field strength) and (the area through which that field passes inside the circuit $ABCD$) in the time Δt.

We have already found (Section 35.4) that this product (field strength × area) is a very useful one and we have used it to produce a model of a field based on the concept of flux. We found that this gave a good picture of the inverse square law of force. Now again, we see that we can usefully introduce the idea of a magnetic flux which is the product of the magnetic field strength and the area through which it passes. The field strength itself can thus be described as the *magnetic flux density* (or magnetic flux per unit area).

When we needed a symbol for the electric flux, we used ψ. We shall write the magnetic flux, ϕ. So $B\,\Delta sl = \Delta\phi$, the change in magnetic flux passing through the circuit.

We can now write the electrical potential difference across the ends AB as

$$\mathscr{E} = \Delta\phi/\Delta t. \tag{40.2}$$

This result is one aspect of the law known as Faraday's law of electromagnetic induction. $\mathscr{E}$ is the induced electromotive force. It is correct to describe it as an electromotive force rather than a potential difference as it describes the total energy given to the charges in the wire, in the same way that a battery's e.m.f. describes the total energy given to the charges.

Historically, the quantity B which we have used to describe the field of force surrounding a current-carrying wire was first introduced through a study of electromagnetic induction. As we have seen, the magnetic flux is a more useful concept than the magnetic field strength in this context. So magnetic flux, which in our units would be measured in tesla metre^2, has a named unit of its own, the *weber*. Since B is the magnetic flux density, another unit for B would be that of magnetic flux density—the weber per square metre (Wb m^{-2}). You will frequently find B referred to as the magnetic flux density rather than

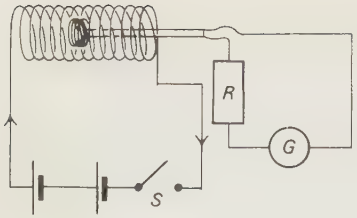


Fig. 40.3

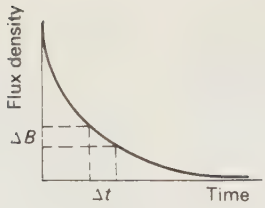


Fig. 40.4

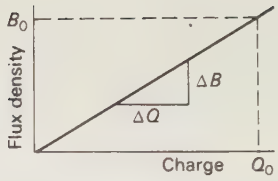


Fig. 40.5

the magnetic field strength and its units given as either

weber metre^{-2} (Wb m^{-2})

tesla (T)

or

newton ampère^{-1} metre^{-1} (N A^{-1} m^{-1}).

All these are different names for precisely the same unit.

To return to Faraday's law, we can see that in terms of instantaneous rates of change,

$$\mathscr{E} = d\phi/dt$$

and that if the circuit (Fig. 40.2) has N turns of wire, rather than just one,

$$\mathscr{E} = N\, d\phi/dt \qquad (40.3)$$

If this was all there was to the law of electromagnetic induction, there would be little that was remarkable about it. But, as we shall now see, the law has a much wider application than to wires bodily moved through magnetic fields.

40.2 THE ELECTROMOTIVE FORCE INDUCED BY A CHANGING MAGNETIC FLUX

Suppose that a coil of N turns and cross-section A is placed within a current-carrying solenoid so that its face is perpendicular to the direction of the magnetic field, as shown in Fig. 40.3. When the switch S is either opened or closed an induced current will flow in the circuit containing the small coil. This current flows only during the short period of time in which the magnetic field in the solenoid builds up or collapses. It is of interest to ask whether the e.m.f. induced can be described in the same way as the e.m.f. which was induced when the circuit of Fig. 40.2 was moved relative to the field.

To test this let us first consider an experiment in which an e.m.f. is induced in a series of coils of different areas and with different numbers of turns.

When the switch S is opened, the flux density B_0 in the coil falls rapidly to zero. Let us assume that we may write the e.m.f. induced in the small coil as

$$\mathscr{E} = N \frac{d\phi}{dt}.$$

$d\phi/dt$ will vary in an unknown way as the field collapses and so the e.m.f. $\mathscr{E}$ does not remain constant. But we may use the work already done in Section 31.3 which suggests that the form of the change may be represented by the graph of Fig. 40.4.

During the small interval of time Δt, there will be a small change in flux density ΔB. During this interval we may write

$$\mathscr{E} = N \frac{\Delta\phi}{\Delta t}$$

$$= NA \frac{\Delta B}{\Delta t}.$$

So

$$\mathscr{E}\, \Delta t = NA\, \Delta B.$$

This e.m.f. induced in the coil causes a current I to flow round the circuit which includes the coil. If this circuit has a resistance R,

$$\mathscr{E} = RI$$

$$\mathscr{E}\, \Delta t = RI\, \Delta t$$

$$= R\, \Delta Q,$$

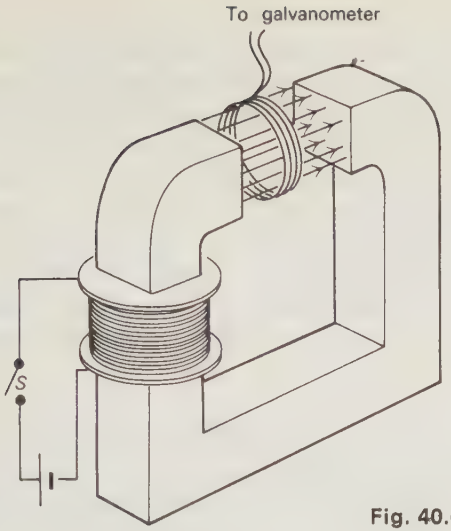


Fig. 40.6

where ΔQ is the small amount of charge flowing round the circuit (which includes the coil), as the field through the coil changes by ΔB. Consequently,

$$R\,\Delta Q = NA\,\Delta B$$

or

$$\frac{\Delta B}{\Delta Q} = \frac{R}{NA}.$$

A graph of the flux density against charge will be a straight line through the origin with slope R/NA (Fig. 40.5). As the field drops from B_0 to 0, the total charge Q induced to flow in the circuit will be:

$$Q_0 = (NA/R)B_0. \qquad (40.4)$$

We have already seen (Unit 6) that a ballistic galvanometer can be used to measure charge, provided the charge flows through the coil of the instrument in a time which is much shorter than its natural period of swing.

Any galvanometer of period about 2 seconds will, in fact, do to test this relationship. Q_0 can be shown to be proportional to N and A. B_0 can be measured using a current balance, and Eq. (40.4) can be checked.

This justifies, as a consequence, our original assumption that $\mathscr{E} = N\,d\phi/dt$ correctly describes the magnitude of the induced e.m.f., whether a conductor physically cuts through magnetic flux, or whether the flux in the circuit changes for some other reason.

The coil considered above was held at right angles to the direction of the magnetic field in the solenoid and we assumed that the area of the cross-section of the coil was the effective area in which the flux linkage occurred. In consequence we found that the induced e.m.f. was a function of the coil area. This needs some care in interpretation, however. Consider yet another example—that of a coil placed between the poles of an electromagnet, as in Fig. 40.6.

With the plane of the coil at right angles to the field direction as before, the e.m.f. induced when S is closed, will be found to be a function of the coil area, but only so long as this area is no bigger than that occupied by the field. For larger coil areas the induced e.m.f. becomes independent of the coil area. This is what we might have anticipated had we carefully interpreted these experiments in terms of our original statement of Faraday's law in relation to the bodily movement of a circuit through a field. There we saw that the induced e.m.f. depended upon the change in magnetic flux through the circuit and not on the product of the change in flux density (B) and the area occupied by the circuit. It is precisely the same in this case. The quantity which determines the magnitude of the induced e.m.f. is the rate of change of magnetic flux. This, in turn, is determined by the magnitude of the field and the area through which it passes.

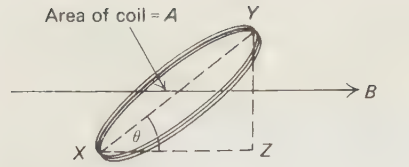
The same care must be taken when the coil makes an angle θ (other than a right angle) with the field direction as in Fig. 40.7(a). The area of the coil through which the field passes in the figure is proportional to the length of the side YZ of the triangle XYZ. The cross-sectional area (A) of the coil is proportional to the side XY of that triangle.

Since

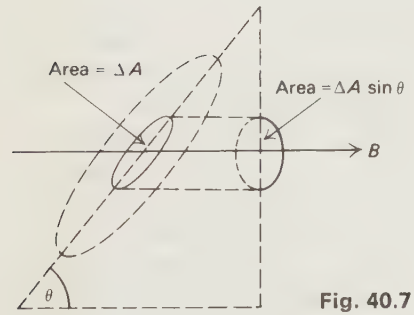
$$YZ = XY \sin\theta$$

the flux through the coil is $B(A \sin\theta)$.

If the field intensity B varies from point to point in the plane of the area through which it passes, the calculation of the flux is more difficult. It is necessary



(a)



(b)

Fig. 40.7

to consider the area of the coil to be broken up into a number of small elements of area (ΔA) through each of which the field may be regarded as reasonably constant (see Fig. 40.7b).

The flux through the element ΔA is $B(\Delta A \sin \theta)$ and the total flux through the coil is

$$\sum_{\substack{\text{over all} \\ A}} B(\Delta A \sin \theta)$$

or, in integral form,

$$\phi = \int_0^A B \sin \theta \, dA. \qquad (40.5)$$

40.3 THE DIRECTION OF THE INDUCED E.M.F.

Faraday's law is often written

$$\mathscr{E} = -N \frac{d\phi}{dt}. \qquad (40.6)$$

The $-$ve sign is put there to remind us about a law which helps to find the direction in which the induced e.m.f. acts. This is Lenz's law which states that the induced e.m.f. is always in such a direction that it opposes the changes producing it.

This is a consequence of the law of conservation of energy. The idea of opposition is often called to mind by putting a $-$ve sign in Faraday's law, but we shall not use it at this stage as we have no convention for the positive direction of changing flux.

If it is necessary to work out the actual direction of induced currents, as in explaining the action of a dynamo, Lenz's law can be very conveniently used by remembering that the change in the magnetic flux due to the induced current must tend to oppose the change in the magnetic flux causing the induced current.

40.4 THE ENERGY STORED IN A MAGNETIC FIELD

We have already seen that energy must be transferred to set up an electric field and that the field itself can be considered to have a store of potential energy (Section 39.3).

In a similar way, we may now see that energy has to be supplied in the setting up of a magnetic field. Let us consider the setting up of a magnetic field in a long solenoid with length l and N turns (Fig. 40.8). We have already seen in Section 31.3 how the current through the solenoid increases with time, according to Fig. 40.9. To understand why the current rises in this way, we found it helpful to separate out the inductive effects from the purely resistive effects. The explanation is repeated again here but now attention is concentrated on the changing magnetic flux within the solenoid rather than on the currents and voltages within the circuit. First we redraw the circuit separating the magnetic field effects from the purely resistive effects (Fig. 40.10). In Fig. 40.10 R represents the pure resistance of the coil. When the switch is closed, the rapidly increasing current sets up a changing and increasing magnetic flux through the solenoid.

This changing flux induces a potential difference across the very coil which is carrying the current. This effect is referred to as self-induction and the law of conservation of energy tells us that this new voltage will be in the opposite direction to that of the battery. Let us suppose that the battery provides an external p.d. of V volts. At any instant when the current has some value I, part of that p.d. will appear across

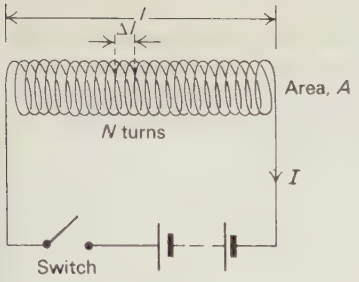


Fig. 40.8

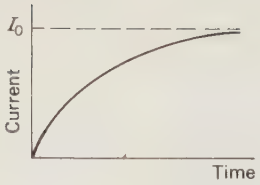


Fig. 40.9

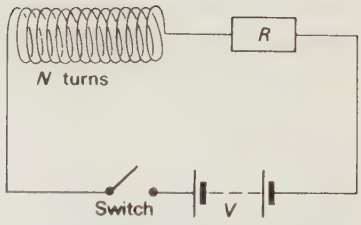


Fig. 40.10

R—an amount IR. The remainder will be overcoming the self-induced voltage in the coil caused by the increasing magnetic flux ϕ. At any instant

$$V = N\frac{d\phi}{dt} + RI,$$

Since $\phi \propto I$, we can write $\phi = kI$, where k is some constant for the particular solenoid.

So

$$V = Nk\frac{dI}{dt} + RI. \qquad (40.7)$$

At first I is very small, consequently, RI is small and dI/dt is large. The current increases rapidly. As I rises, RI increases. So dI/dt gets smaller and smaller until, when $I = I_0$, $V = RI_0$ and $dI/dt = 0$.

In a short period of time Δt, during which the current remains approximately steady at I, $I\,\Delta t$ units of charge pass round the circuit. The battery provides $VI\,\Delta t$ units of energy to this charge. Since

$$V = Nk\,dI/dt + RI,$$

we can write

$$VI\,\Delta t = (Nk\,dI/dt)I\,\Delta t + RI^2\,\Delta t.$$

The final term represents the energy converted into heat in the circuit. The first is the energy supplied by the battery to the charge to overcome the self-induced p.d. in the solenoid. What happens to this energy? The only other change taking place in the circuit is that the magnetic field in the solenoid is becoming stronger. By the law of conservation of energy, this energy must be stored in the increasingly strong field. We will now find a relationship between the field strength and the stored energy.

The field within this long solenoid of length l, cross-sectional area A and N turns depends upon the current I which is flowing through it. As we have seen, the field intensity B is given by Eq. (37.4)

$$B = \mu_0\frac{N}{l}I$$

for regions well away from the ends of the solenoid. N/l is the number of turns per unit length. Since the field outside this long solenoid is essentially zero the entire store of energy must be within the coil.

Consider a small region of the field which is contained within a length Δl near to the centre of the solenoid (Fig. 40.8). It will contain $(N\,\Delta l)/l$ turns.

If the current is changing and has, at the moment considered, a value I, the e.m.f. induced in this short length of coil will be

$$\mathscr{E} = \frac{N}{l}\Delta l\frac{\Delta\phi}{\Delta t}.$$

Since $\phi = BA$, this may be written

$$\mathscr{E} = \frac{N}{l}\Delta lA\frac{\Delta B}{\Delta t}. \qquad (40.8)$$

In a short interval of time Δt:

increase in stored energy $= \mathscr{E}I\,\Delta t$ (Eq. 25.2).

Substituting from Eq. (40.8) for $\mathscr{E}$ and from Eq. (37.4) for I we have:

$$\text{increase in energy stored} = \left(\frac{N}{l}\,\Delta l A\,\frac{\Delta B}{\Delta t}\right)\left(\frac{B}{\mu_0}\,\frac{l}{N}\right)\Delta t$$

$$= \frac{A\,\Delta l}{\mu_0}\,B\,\Delta B.$$

Now $A\,\Delta l$ is the volume occupied so the change in energy per unit volume (energy density) is $(1/\mu_0)B\,\Delta B$.

In the limit as Δt tends to zero we have

$$\text{energy density} = \int_0^{B_0} \frac{1}{\mu_0}\,B\,dB$$

$$= \frac{1}{2}\cdot\frac{B_0^2}{\mu_0}. \qquad (40.9)$$

This may be compared with the corresponding energy density in an electric field of $\frac{1}{2}\varepsilon_0 E_0^2$.

Problem 40.1 Calculate an approximate value for the total energy stored in the magnetic field of a 500 turn coil 25 cm long and of mean diameter 5 cm when carrying a current of 5 amperes. Remember that Eq. (40.9) gives the energy density (i.e. energy per unit volume). (You should find that the total store of energy is about 3×10^{-2} J.)

40.5 THE SELF-INDUCTANCE OF A SOLENOID

In Unit 6, the coefficient of self-inductance L of a coil was introduced through the equation

$$\text{induced e.m.f.} = -L\,\frac{dI}{dt}. \qquad (40.10)$$

In evaluating the energy stored in the magnetic field of a solenoid we used the expression (*cf.* Eq. 40.6).

$$\text{Induced e.m.f.} = N\,\frac{d\phi}{dt}$$

$$= Nk\,\frac{dI}{dt}.$$

Evidently

$$L = Nk.$$

We can calculate an approximate value for this self-inductance. For the greater part of the length of a long solenoid of cross-sectional area A, length l, and N turns

$$B = \mu_0\,\frac{N}{l}\,I.$$

Since $\phi = BA$ we have

$$\phi = \mu_0\,\frac{AN}{l}\,I$$

and

$$\frac{d\phi}{dt} = \mu_0\,\frac{AN}{l}\,\frac{dI}{dt}.$$

Hence the induced e.m.f.

$$V = \mu_0(AN^2/l)\,dI/dt.$$

Comparing this with Eq. (40.10) we see that

$$L = \mu_0\,\frac{AN^2}{l}. \qquad (40.11)$$

Later in this Unit we shall wish to use this expression to write down an order of magnitude value for the inductance of a coil of a single turn.

Problem 40.2 Calculate an approximate value for the self-inductance of the coil referred to in the previous problem. How would you check your value of about 10^{-3} H experimentally?

40.6 AN ALTERNATIVE FORM OF FARADAY'S LAW OF ELECTROMAGNETIC INDUCTION

We have already seen that the electric charges within a conductor which is moving through a magnetic field are subject to a force which is described by the equation $F = Bev$. An induced e.m.f. is established across the moving wire of value

$$\mathscr{E} = \frac{d\phi}{dt}$$

in consequence.

However, there is another aspect of electromagnetic induction for which this particular description

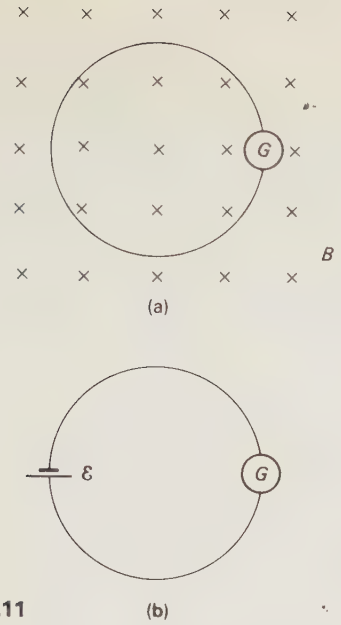


Fig. 40.11

is inadequate. An e.m.f. is induced not only when a conductor is in motion through a field but also when that field changes in strength.

Consider the following circuit (Fig. 40.11a). The circuit does not move but we arrange for the magnetic field B to decrease steadily with time. The induced e.m.f. is still described by the equation

$$\mathscr{E} = \frac{d\phi}{dt}.$$

But we can no longer invoke $F = Bev$ to describe the force which the charges within the wire are experiencing.

Let us look at the situation from the point of view of those charges. Suppose the changing magnetic field were to be replaced by a battery of e.m.f. $\mathscr{E}$ as in Fig. 40.13(b). The charges behave in precisely the same way as they did when under the influence of the changing magnetic field. But we must ascribe a different reason to this movement. We now say that an electric field E is set up in the wire and that this field drives the charges round the circuit. If we were to rely on our observations of the behaviour of charges alone, we could not distinguish between their behaviour under the influence of a changing magnetic field, a battery or even a charged capacitor. The simplest way to treat these observations is to say that there is no difference in the forces driving the charges round the circuit. If we do this, we are bound to assume that a *changing magnetic field has set up an electric field.*

We have seen that for fields in general, the force per unit quantity (i.e. field strength) is equal to the negative potential gradient at the point concerned (see Eq. 38.4). In this case, the high degree of symmetry suggests that the field intensity will be the same at all points in the loop. This, indeed, is why we chose a circular loop. It follows that, if the length of the loop of wire is l, the magnitude of the field strength within it is given by

$$\text{field strength} = E = \frac{\mathscr{E}}{l}$$

so

$$El = \mathscr{E}$$

therefore

$$El = \frac{d\phi}{dt}. \qquad (40.12)$$

The direction of the field E depends on both the direction of the magnetic flux and whether this is increasing or decreasing. Each case can be worked out from first principles with the help of Lenz's law but we shall give a rule for this shortly.

There is a great deal of evidence in support of this view of Faraday's laws of electromagnetic induction. One important type of particle accelerator called the Betatron depends for its operation on the existence of this field quite independently of any conducting wire.

Charged particles (such as protons) are injected into the hollow circular tube shown in Fig. 40.12 and are accelerated round a circle by the electric field created by a changing magnetic field. An additional magnetic field provides the necessary central force. For a fuller description of the way the Betatron and other charged particle accelerators work see Wilson and Littauer (1962).

Returning to the circuit of Fig. 40.13, the presence of the electric field does not depend on the existence

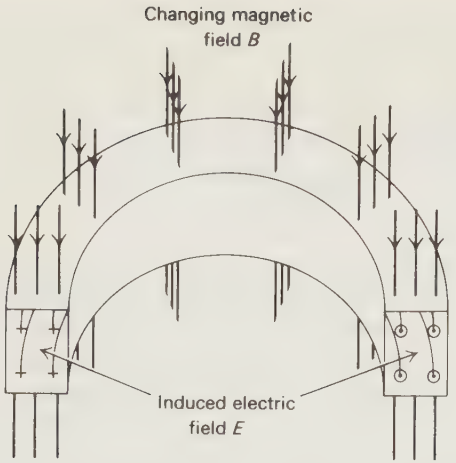


Fig. 40.12 The principle of the betatron accelerator.

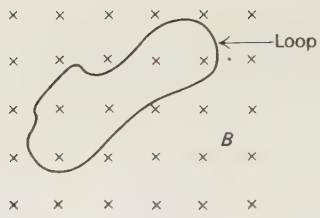


Fig. 40.13



Fig. 40.14

of the wire. The field exists wherever there is a changing magnetic flux. The wire merely provides charges upon which the field may act.

One or two further things must be said about this form of Faraday's law.

First of all, induced e.m.f.s are always measured in complete circuits. The induced e.m.f. is equal to the rate of change of magnetic flux within the closed loop of the circuit.

Thus the product (El) must be taken round a closed loop.

Of course, E may vary in size as we pass round any arbitrary closed loop in the changing magnetic field and we must allow for this. Nor need the loop necessarily be a circle. It could have some such shape as is shown in Fig. 40.13.

Although there are mathematical ways of taking these factors into account we shall not concern ourselves with anything but the simplest applications. We will content ourselves by writing

$$(El)_{\text{round loop}} = \frac{d\phi}{dt} = \frac{d}{dt}(BA)_{\text{within loop}}. \qquad (40.13)$$

This induced electric field exists round a loop of any size placed in the magnetic field. The magnitude of the electric field depends on the area of the loop as we can see in Fig. 40.14. Consider the left-hand circuit of radius r and apply Eq. (40.13).

The electric field E which is established does not have a unique value at a particular point in space—it depends on the circuit in which it is operative. Neither does it have a unique direction. Consider the two loops of wire shown in Fig. 40.14. The electric field at P in the solid line loop is diametrically opposed to the electric field at P in the dotted line loop. We have to agree that this induced electric field is a very different thing from the original electric field which we explored in connection with static electric charges.

These difficult ideas can only be fully expressed in mathematical language. The development of this model of the magnetic and electric fields was due to J. C. Maxwell and, although it will not be possible to deal with it fully here, we may hope to gain some appreciation of the wonderful utility of this triumph of nineteenth century physics.

Chapter 41
MAXWELL'S MODEL OF THE ELECTROMAGNETIC FIELD

41.1 Introduction
41.2 Maxwell's equations
41.3 Loops of magnetic flux
41.4 A search for a fourth equation of the electromagnetic field
41.5 The directions of the induced electric and magnetic fields
41.6 Electromagnetic oscillations and electromagnetic waves
41.7 The special theory of relativity

41.1 INTRODUCTION

We have already seen how Faraday with his lines of force set up a model to describe the way in which a force generated by the presence of a distant charge or magnet could act on other charges or magnets some distance away.

This essentially non-mathematical theory had a mixed reception. The lines, or tubes, invented by Faraday had no reality other than the extent to which they could be used to describe observed events.

Nevertheless the region of space surrounding masses, magnets, and charges does seem to possess special properties which do not appear to be dependent on the presence of other masses, magnets, and charges for the forces to act on.

This becomes particularly clear as soon as we appreciate that a link exists between electric and magnetic fields. A changing magnetic field gives rise to closed loops of electric flux. Attention is diverted from the currents and charges to the space surrounding them.

Maxwell, strongly influenced by Faraday's model of a field of force, attempted to describe the properties of this field in mathematical language and so to dispense with the elastic tubes of Faraday's theory.

Maxwell himself used a number of mechanical models to set up his equations but ultimately, as Einstein wrote:

> "He showed that the whole of what was then known about light and electromagnetic phenomena was expressed in his well-known double system of differential equations in which the electric and the magnetic fields appear as the dependent variables. Maxwell did, indeed, try to explain or justify these equations by intellectual constructions. But he made use of several such constructions at the same time and took none of them really seriously, so that the equations alone appeared as the essential thing and the strength of the fields as the ultimate entities, not to be reduced to anything else." [*From Newman, J. R.*, James Clark Maxwell. *Copyright © 1955 by Scientific American Inc. All rights reserved.*]

How can we hope to give a picture of this most

spectacular of mathematical theories? Many scientists would say that without the mathematical background we should not attempt it. But this would be to deny a place, in a book which has emphasized the function of models and theories, to one of the greatest and most productive theories of the modern era. Furthermore, it can help us to see that theories and models in physics do not have to be concrete pictures of the unimaginable in terms of billiard balls and water waves.

So we will start this last chapter of this Unit with another quotation, this time from Newman himself:

> "What are the Maxwellian equations? I cannot hope to give an easy answer to this question, but at the cost of deliberate oversimplification I must try summarily to explain them, for they are at the heart of the theory."

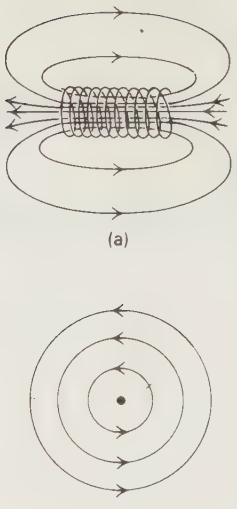


Fig. 41.1

41.2 MAXWELL'S EQUATIONS

The first two of Maxwell's equations are mathematical statements of an already productive model—the flux model. We have already seen that the inverse square law of force allows us to interpret both the electric field and the magnetic field strength as a flux density. Faraday's lines of force were a pictorial representation of this fact.

Maxwell's first equation states in mathematical terms that we may use this model for the electric field strength and that consequently the net electric flux emerging from a small volume of space is equal to

$$\frac{\text{total charge in the volume}}{\varepsilon_0}.$$

His second equation states that this flux model is also viable for magnetic fields but that since magnetic flux always forms closed loops (think of the iron-filing pattern round a current-carrying wire), the total (net) flux through any volume in space is always zero.

These first two equations have a symmetry about them which we have come to expect of electric and magnetic fields.

The third law incorporates Faraday's law of electromagnetic induction, but in the electric field form which we have just developed (Eq. 40.13).

A changing magnetic field sets up closed loops of electric flux.

$$(El)_{\text{round a loop}} = \frac{d}{dt}(\phi)_{\text{through the loop}} \qquad (41.1)$$

is the nearest we can get to its mathematical statement.

So much was observed fact at the time Maxwell set up his equations. But, he felt, there ought to be a fourth equation, symmetrical with the third, as the second was with the first.

Was there, in other words, any sense in talking about closed loops of magnetic flux being set up by a changing electric field? Or, in our simple terms, was

$$(Bl)_{\text{round a loop}} = \frac{d}{dt}(\psi)_{\text{through the loop}}? \qquad (41.2)$$

Let us try to explore the possibility experimentally.

41.3 LOOPS OF MAGNETIC FLUX

There is certainly sense in this idea—magnetic flux always forms closed loops (see Fig. 41.1). Consider the magnetic field of strength B, within the inner section of the solenoid (Fig. 41.2) which, in order to make our equations apply exactly, we assume to be

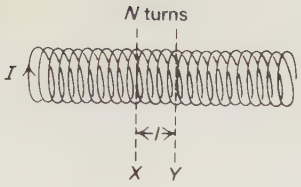


Fig. 41.2

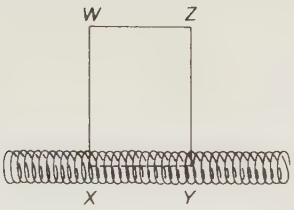


Fig. 41.3

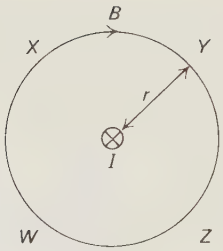


Fig. 41.4

of infinite length. This inner section XY extends over a length l and has N turns. From Eq. (37.4) we have

$$B = \mu_0 \frac{N}{l} I$$

and so

$$Bl = \mu_0 NI.$$

If now the solenoid is stretched so that the section XY now extends over a length $2l$, the new field will be given by

$$B' = \mu_0 \frac{N}{2l} I$$

and so

$$B'2l = \mu_0 NI.$$

With the decrease in the number of turns per unit length from N/l to $N/2l$, the field intensity decreases and

$$B' = \tfrac{1}{2}B.$$

However, the product ($B \times$ length over which B acts) has remained unchanged and we have

$$(Bl_{\text{over which } B \text{ acts}}) = \mu_0 NI.$$

We may write this

$$(Bl_{\text{over which } B \text{ acts}}) = \mu_0 I \times \text{no. of times } I \text{ loops } B \tag{41.3}$$

for N, the number of turns in the chosen length of solenoid, is a measure of the number of times the current I loops around the field. (We may compare this with Eq. (40.6) which may be written

$$\text{(Induced e.m.f.)} = \frac{d\phi}{dt} \times \text{number of times the circuit loops } B.)$$

In Eq. (41.3), the product B was not taken round a closed loop but we may do this by considering Fig. 41.3.

Around the closed loop $WXYZ$, the field intensity along the lines YZ, ZW, and WX is zero for this solenoid is infinitely long. Only in the section XY can we detect a magnetic field B. So, around this closed loop we have, from Eq. (41.3)

$$B\ XY + 0 + 0 + 0 = (\mu_0 NI)XY$$

or

$$(Bl)_{\text{round a loop}} = (\mu_0 I)_{\text{through the loop}} \tag{41.4}$$

where we are to understand by the term

$$(\mu_0 I)_{\text{through the loop}}$$

that we have multiplied the measured value of the current I by the number of times this current is looped around the closed loop.

Let us apply this new equation to a case which we have not yet considered, that of a long straight current-carrying wire. In this case the closed loop $WXYZ$ has a total length of $2\pi r$ and the current I passes once only through this loop (Fig. 41.4). We have

$$(B2\pi r) = \mu_0 I$$

whence

$$B = \frac{\mu_0 I}{2\pi r}. \tag{41.5}$$

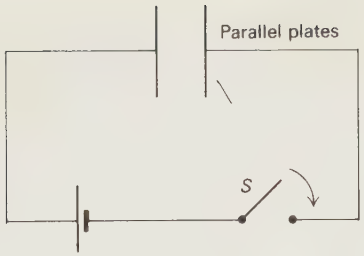


Fig. 41.5

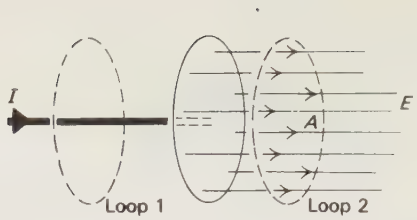


Fig. 41.6

This expression for the magnetic flux density B near to a long straight wire carrying a current can be derived by the application of the Biot-Savart law, but it has been obtained by a procedure which is much simpler to operate. This new law which is stated in Eq. (41.4) is known as *Ampère's law*.

41.4 A SEARCH FOR A FOURTH EQUATION OF THE ELECTROMAGNETIC FIELD

Consider the circuit shown in Fig. 41.5. When the switch S is closed, a current flows in the circuit until the plates are fully charged. During this time a magnetic field surrounds the wire. It was Maxwell who first asked whether it is reasonable to suppose that this magnetic field stopped short at the plates. If it did not, but surrounded the region between the plates, what was its cause, for no charges flowed here?

The magnetic field is present only during the charging or discharging processes. When the plates are fully charged, the magnetic field is absent.

In the first case a current flows; in the second case it does not. What of the region between the plates? In the first case the electric field is increasing or decreasing; in the second it is static.

So if a magnetic field surrounds the region between the plates, it is presumably related to this changing field. Electric flux, ψ, through area A, equals Q/ε_0 (Flux law, Eq. 36.8). In time Δt, $I\,\Delta t$ coulombs flow on to the plate increasing the flux by an amount

$$\Delta\psi = \frac{I\,\Delta t}{\varepsilon_0}.$$

Hence

$$I = \varepsilon_0 \frac{\Delta\psi}{\Delta t}.$$

Around the wire, the magnetic field is given by Ampère's law (Fig. 41.6)

$$(Bl)_{\text{round a loop}} = (\mu_0 I_{\text{through the loop}}).$$

If the same field exists around the changing electric field between the two plates:

$$(Bl)_{\text{round a loop}} = \mu_0 \left(\varepsilon_0 \frac{\Delta\psi}{\Delta t}\right)_{\text{through the loop}}$$

or, writing this in terms of instantaneous rates of change

$$(Bl)_{\text{round a loop}} = \mu_0\varepsilon_0 \frac{d}{dt}(\psi)_{\text{through the loop}}. \qquad (41.6)$$

Is this the fourth law for which we are looking? Maxwell assumed so, but our derivation of it—very close to that of Maxwell's in principle—is something of an inspired guess.

41.5 THE DIRECTIONS OF THE INDUCED ELECTRIC AND MAGNETIC FIELDS

In using these loop laws the electric and magnetic flux loops have definite directions in which the fields go round the loops. We will work them out.

a) Direction of E in $(E.l)_{\text{round loop}} = \dfrac{d}{dt}(\phi)_{\text{through loop}}$

If the magnetic field, B, through a conducting loop is increasing in the direction shown in Fig. 41.7.

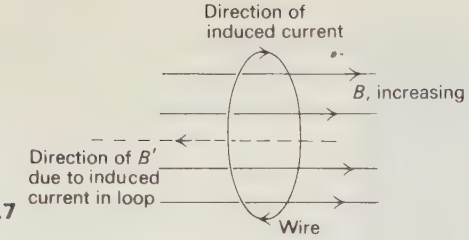


Fig. 41.7

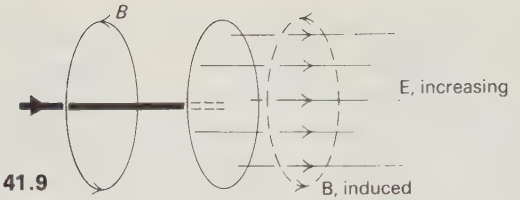


Fig. 41.9

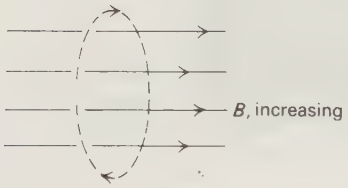


Fig. 41.8

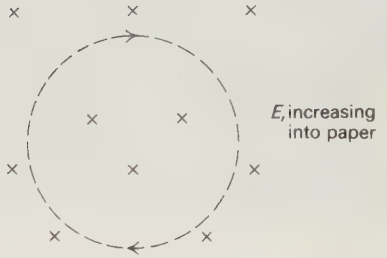


Fig. 41.10

Lenz's law tells us that the induced current must be in such a direction that its own magnetic field B' opposes the change. B' is thus oppositely directed to B and the direction of the induced current in the loop follows from the right-handed screw rule. This is also the direction of the electric field in the loop (Fig. 41.8) and thus the direction of the electric field if the conducting loop is removed.

b) Direction of B in $(Bl)_{\text{round loop}} = \mu_0\varepsilon_0 \dfrac{d}{dt}(\psi)_{\text{through loop}}$

This can be worked out from the known direction of the magnetic field round a current-carrying wire which leads to the plate of a capacitor as in Fig. 41.9. The direction of the current gives the direction of the electric field which is increasing between the plates. The accompanying magnetic field direction is obtained from the right-handed screw rule. You will notice that E_{induced} in Fig. 41.8 is oppositely directed from B_{induced} in Fig. 41.9 for the same direction of increasing flux.

By convention, if the electric flux is increasing into the paper as in Fig. 41.10, the positive way round the loop is clockwise. Thus we write,

$$(Bl)_{\text{round loop}} = \mu_0\varepsilon_0 \frac{d}{dt}(\psi)_{\text{through loop}} \qquad (41.7)$$

but

$$(El)_{\text{round loop}} = -\frac{d}{dt}(\phi)_{\text{through loop}}. \qquad (41.8)$$

So we summarize:

A simple guide to Maxwell's equations of the electromagnetic field in free space.

1. Net electric flux through a small volume
   $= (\text{charge in volume})/e_0$
   $= 0,$
   if there are no charges.
2. Net magnetic flux through a small volume $= 0$.
3. $(El)_{\text{round loop}} = -\dfrac{d}{dt}(\phi)_{\text{through loop}}$.
4. $(Bl)_{\text{round loop}} = \mu_0\varepsilon_0 \dfrac{d}{dt}(\psi)_{\text{through loop}}$.

But what do we gain from such a theory? We have not put it to a test, and the fourth equation is

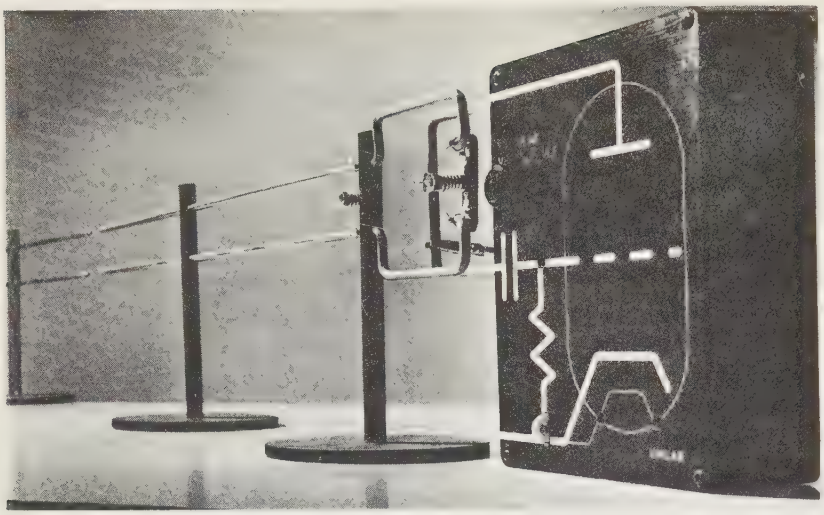


Fig. 41.11 A v.h.f. oscillator. (Courtesy of the Unilab division of Rainbow Radio, Blackburn.)

based on an inspired guess. What new predictions can we make if these four equations are a good model of the electromagnetic field? Einstein and Infeld (1961) wrote in *Evolution of Physics*:

> "In Maxwell's theory, there are no material actors. The mathematical equations of this theory express the laws governing the electromagnetic field. They do not, as in Newton's laws, connect two widely separated events; they do not connect the happenings *here* with the conditions *there*. The field *here* and *now* depends on the field in the *immediate neighbourhood* at a time *just past*. The equations allow us to predict what will happen a little further in space and a little later in time, if we know what happens here and now. They allow us to increase our knowledge of the field by small steps. We can deduce what happens here from that which happened far away by the summation of these very small steps. In Newton's theory, on the contrary, only big steps connecting distance events are permissible. The experiments of Oersted and Faraday can be regained from Maxwell's theory, but only by the summation of small steps each of which is governed by Maxwell's equations.
>
> A more thorough mathematical study of Maxwell's equations shows that new and really unexpected conclusions can be drawn and the whole theory submitted to a test at a much higher level, because the theoretical consequences are now of a quantitative character and are revealed by a whole chain of logical arguments.
>
> Let us imagine an idealized experiment. A small sphere with an electric charge is forced, by some external influence, to oscillate rapidly and in a rhythmical way, like a pendulum. With the knowledge we already have of the changes of the field, how shall we describe everything that is going on here, in the field language?
>
> The oscillation of the charge produces a changing electric field. This is always accompanied by a changing magnetic field. If a wire forming a closed circuit is placed in the vicinity, then again the changing magnetic field will be accompanied by an electric current in the circuit. All this is merely a repetition of known facts, but a study of Maxwell's equations gives a much deeper insight into the problem of the oscillating electric charge. By mathematical deduction from Maxwell's equations we can detect the character of the field surrounding an oscillating charge, its structure near and far from the source, and its change with time. The outcome of such deduction is the *electromagnetic wave*."

This was the important prediction of Maxwell's theory which no other model of electric and magnetic behaviour had yet suggested. So we will see if, in a simplified way, we can make the prediction for ourselves.

41.6 ELECTROMAGNETIC OSCILLATIONS AND ELECTROMAGNETIC WAVES

We have already seen (Chapter 32) that an oscillating electric current can be established by connecting a charged capacitor to an inductor.

If the capacitor has a value of C and the inductor a value of L, then the frequency f of the oscillations is given by

$$f = \frac{1}{2\pi\sqrt{LC}}.$$

Normally these oscillations die away very quickly indeed, but they may be maintained by suitably amplifying a proportion of the oscillating voltage and feeding it back into the oscillatory circuit to make up for the losses.

The frequency of these oscillations can be made very high indeed, by suitably reducing the sizes of the capacitance and the inductance.

The circuit in the photograph (Fig. 41.11) consists of an inductance of one turn and the capacitance which exists between the anode and grid of the thermionic vacuum tube (valve).

Let us find the order of magnitude of the frequency at which such a circuit might oscillate. We have seen (Eq. 28.6) that, for a parallel plate capacitor

$$C = \frac{Q}{V} = \varepsilon_0 \frac{A}{d}.$$

Inside the tube, the anode and the grid consist of two concentric cylinders about 3 cm high and 1.5 cm in diameter with a separation of about 3 mm. If we regard them as together constituting a cylindrical plate capacitor with a plate area of about 14×10^{-4} m^2 and a plate separation of 3×10^{-3} m the capacitance should be of the order of 4×10^{-12} F, since ε_0 is $1/36\pi \times 10^{-9}$ J^{-1} C^2 m^{-1}. You should check this computation.

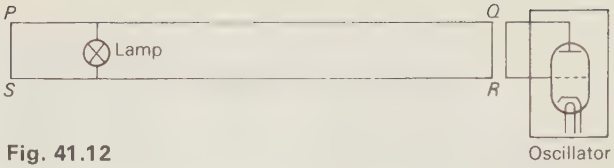


Fig. 41.12

To find the inductance L let us assume that the loop connecting the anode and the grid has the inductance of a one turn long solenoid. This is a drastic approximation (why?) but is justified in that we are only concerned to obtain an order of magnitude for the frequency. Applying Eq. (4.11) we have

$$L = \mu_0 N^2 \frac{A}{l}.$$

The cross-sectional area of this one turn solenoid is about $(0.1)^2$ m^2 and l is about 0.5×10^{-2} m. Remembering that μ_0 is $4\pi \times 10^{-7}$ N A^{-2} you will find that the self inductance L is about 2.5×10^{-6} H.

Substituting these values for C and L in the equation $f = 1/2\pi\sqrt{LC}$ you will find that the frequency is about 50 MHz. In terms of order of magnitude, this circuit should produce electrical oscillations with a frequency in the order of 10^8 Hz. In practice the measured frequency will be found within the range 1 to 2×10^8 Hz.

We can regard this oscillator as a generator of very high frequency a.c. Let us couple it inductively to the circuit $PQRS$ which consists of two long parallel wires about 10 cm apart, along which a conductor, which is fitted with a small lamp, can slide (see Fig. 41.12).

We discover some very surprising things if now we move the lamp conductor along the conductors. It lights in some places and not in others. In fact it alternates between lit and unlit as it moves from PS to QR, with a distance of about 1.2 metres between the points at which it is most brightly lit.

If the connection is broken between P and S, the lamp still lights at some places and not at others, but now at different spots—still separated by 1.2 m.

It is difficult to avoid comparing this variation in the brightness of a lamp with the alternations of

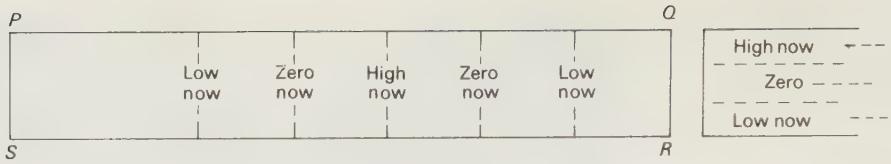


Fig. 41.13

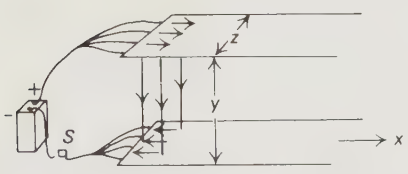


Fig. 41.14

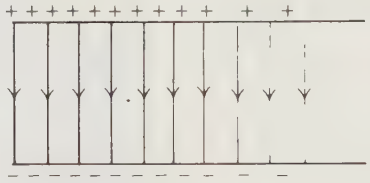


Fig. 41.15

intensity which take place along a standing wave as we pass from node to antinode.

Is this a reasonable explanation of the observed effects? Let us suppose that the speed with which changes in currents and p.d.'s can be transmitted is strictly limited. If the frequency of alternations at QR is very high, then the information "high p.d. now" may have only got part way down the line before the p.d. across QR starts to send out the information "zero p.d. now." After a short while we have the situation shown in Fig. 41.13.

A potential wave passes down the circuit. At the far end we have the same sort of boundary conditions which exist for standing waves on a slinky spring. In this case, the p.d. remains zero at all times because it is short-circuited. So a potential wave is reflected in opposite phase to the oncoming wave, so that every "high now" is cancelled by a "low now" at the far end. As this reflected wave returns it will set up a standing potential wave, antinodes where "high now" corresponds to "high now" and "low now" corresponds with "low now."

(What will you have at the reflecting end—an antinode or a node?)

Such a standing wave of current and p.d. is necessarily accompanied by a similar pattern of magnetic and electric fields between the wires. Where the current is high, there will be a high magnetic field. Where the p.d. is high there will be a high electric field. Thus we may be able to seek a reason for the finite speed of propagation of the measured currents and p.d.'s by seeing if we can find any reason why changes in an electromagnetic field are transmitted through the field at a strictly limited speed. Why indeed should such information take any time to travel at all?

In order to investigate this problem quantitatively let us imagine an experiment—a thought experiment.

A battery is connected by a number of wires to two wide plates as shown in Fig. 41.14. The multiplicity of leads ensures that the whole edge of each plate receives information from the battery at the same instant. Switch S is closed. The plates start to charge up, the top plate positively, the bottom plate negatively, and an electric field is set up between them. The information that the battery has been connected across the plates takes time (we assume) to travel down the plates—let us suppose that it travels at a speed v. After a short time, the field between the plates may be represented as in Fig. 41.15.

As the field spreads out along the space between the plates, charge must be supplied by the battery continuously. So there is a steady flow of charge along the plates up to the field edge. This will give an electric current outwards along the top plate and inwards along the bottom—setting up a magnetic field between the plates as well.

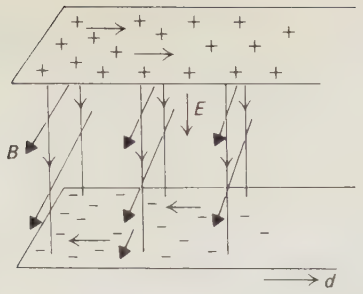


Fig. 41.16

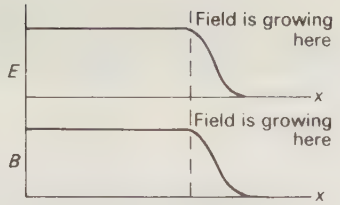


Fig. 41.17

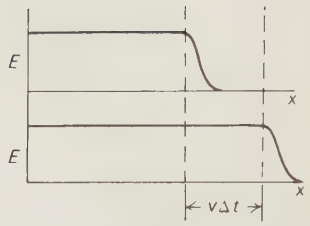


Fig. 41.18

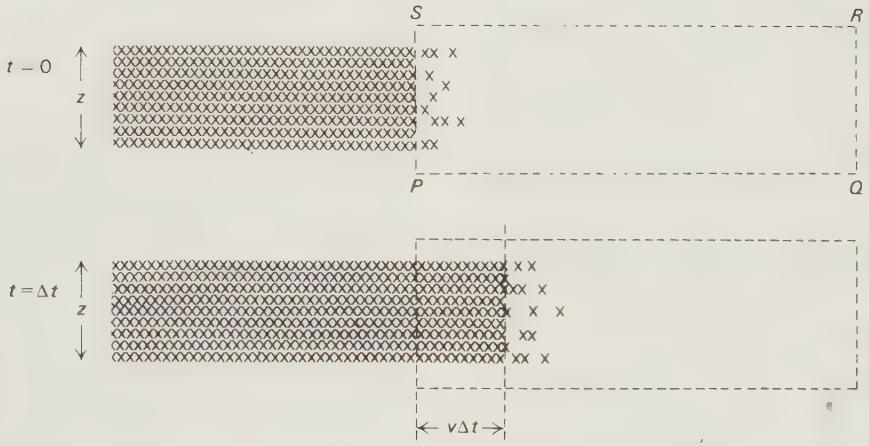


Fig. 41.19

If the right-hand screw rule is applied to any strip of current along each plate, it will be seen (Fig. 41.16) that the direction of this magnetic field between the plates is perpendicular to E and coming out from the paper. If the plates carrying the current are very wide compared with their separation, the magnetic field is almost entirely confined to the region between the plates and can be shown to be uniform to a very good approximation—indeed accurately so if the plates are infinitely wide.

Graphs of the intensities of the electric field and the magnetic field after a time, t, will appear as in Fig. 41.17. So we have an edge of electric and magnetic field moving down between the plates.

We have established all of this by considering the flow of charges on the plate. But so far there is no reason why this field edge should not move outwards as quickly as it likes.

Let us look again at this field edge moving forward, but this time, we will ignore the charges moving on the plate and use instead other laws about the behaviour of changing fields to see what ought to happen.

Figure 41.18 shows two graphs of the electric field intensity, E between the plates, separated in time by Δt. We will continue to assume the field edge moves forward at speed v.

At a time Δt later, the field edge will have moved a distance $v\,\Delta t$. Let us draw a map of the electric field as though we are looking along the direction of the field (see Fig. 41.19). The crosses represent the ends of arrows pointing into the paper. Where the

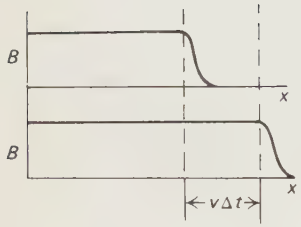


Fig. 41.20

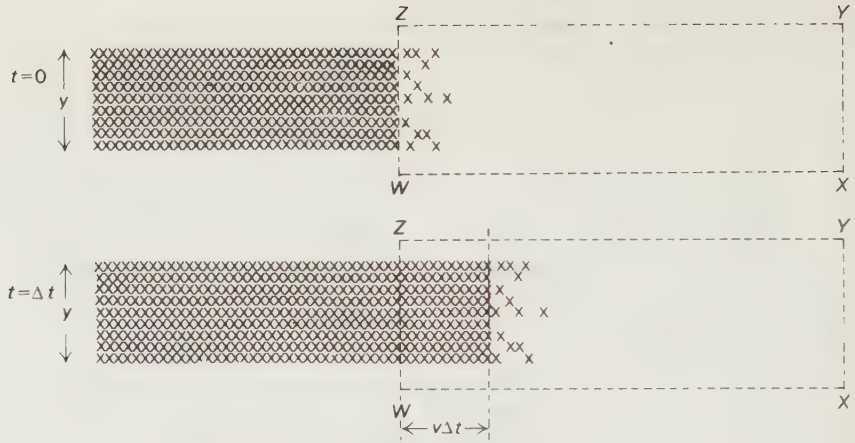


Fig. 41.21

field gets weaker in the edge we will draw fewer crosses.

Within the loop $PQRS$, an electric flux which is directed into the plane of the paper has increased by an amount

$$\Delta\psi = E\,(vz\,\Delta t).$$

Hence by the loop law (Eq. 41.7) we can write

$$(Bl)_{\text{round } PQRS} = \mu_0\varepsilon_0 \left(\frac{\Delta\psi}{\Delta t}\right)_{\text{through } PQRS}$$

$$= \mu_0\varepsilon_0 Evz.$$

Now we shall make the assumption that the field induced by the changing electric flux is the same as the field left behind the moving edge. By considering the currents flowing in the plate, we have seen that this is a uniform field perpendicular to E confined to the region between the plates.

Hence $(Bl)_{\text{round } PQRS} = Bz$, since $B = O$ everywhere except along the length z of PS.

Hence

$$Bz = \mu_0\varepsilon_0 Ezv,$$

$$B = \mu_0\varepsilon_0 Ev. \qquad (41.9)$$

So we can see how a travelling electric field can set up alongside it a magnetic field, travelling forward at the same rate. The direction of B according to the law of induction will be along SP. You should check that this is the direction we would expect B to be in, in terms of the currents in the top and bottom plate.

We can apply the same treatment to the edge of magnetic field moving along between the plates. The graphs for the B-field are shown in Fig. 41.20.

Looking along the B-field (in the direction S to P), the corresponding field maps are given in Fig. 41.21. Within the loop $WXYZ$ a magnetic flux has increased into the paper by an amount

$$B\,(yv\,\Delta t).$$

Hence by the loop law of Eq. (41.8), we can write

$$(El)_{\text{round } WXYZ} = -\frac{\Delta}{\Delta t}(\phi)_{\text{through } WXYZ},$$

$$= -Byv.$$

Using the same argument as before (that the induced E field in the edge should correspond to the field left behind it)

$$(El)_{\text{round } WXYZ} = Ey,$$

since $E = 0$ everywhere except along the length y of WZ.

Hence

$$Ey = -Byv,$$

$$E = -Bv. \qquad (41.10)$$

We can thus see that, conversely, a travelling magnetic field can set up an electric field travelling alongside it. The direction of E according to the law of induction will be along ZW. You should check that this is the direction we would expect E to be in, in terms of the charges accumulating on the plates.

We have assumed throughout that the induced electric field E is equal to the electric field we see set up by the accumulated charges. Similarly, we assume that the induced magnetic field B is equal to the magnetic field set up by the currents in the plates.

Hence we have written

$$B = \mu_0 \varepsilon_0 E v,$$

and

$$E = Bv,$$

(ignoring $-$ *ve* signs which simply gave directions).

Eliminating E from these two equations:

$$B = \mu_0 \varepsilon_0 (Bv)\, v.$$

Hence

$$1 = \mu_0 \varepsilon_0 v^2,$$

or

$$v = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}. \qquad (41.11)$$

How large is this speed?

$$\mu_0 = 4\pi \times 10^{-7} \text{ henry metre}^{-1},$$

$$\varepsilon_0 = \frac{1}{36\pi} \times 10^{-9} \text{ farad metre}^{-1}.$$

Hence,

$$\mu_0 e_0 = \tfrac{1}{9} \times 10^{-16} \text{ metre second}^{-1}.$$

So

$$v = 3 \times 10^8 \text{ metre second}^{-1}. \qquad (41.12)$$

This we recognize as the speed of light.

Before seeking some experimental support for this, we will make one further check on our analysis which, it must be admitted is founded on reasonable assumptions rather than on exact deductions. We have seen that the unit in which μ_0 (the permeability of a vacuum) is measured is the newton ampere^{-2} (see Section 37.7). We have also seen that the unit which is ε_0 (the permittivity of a vacuum) is measured is the joule^{-1} coulomb^{-2} metre^{-1}. This latter unit is equivalent to the coulomb^2 newton^{-1} metre^{-2} (see Section 35.1). Consequently the unit in which $\mu_0 \varepsilon_0$ is quoted are:

$$\frac{\text{newton}}{\text{ampere}^2} \quad \frac{\text{coulomb}^2}{\text{newton metre}^2}.$$

Since the coulomb represents the quantity of electricity passing round a circuit when 1 ampere flows for one second, we can write

$$\text{coulomb}^2 = \text{ampere}^2 \text{ second}^2.$$

Thus, the units of $\mu_0 \varepsilon_0$ are second^2 metre^{-2}, and the units of $1/\sqrt{(\mu_0 \varepsilon_0)}$ are metre second^{-1}, the units of speed, or velocity. This gives us some check on the result we have derived.

We can check the size of this speed experimentally. It is possible to send pulses down a long co-axial cable and to observe the return of the reflected pulses on a cathode ray oscilloscope (see Fig. 41.22).

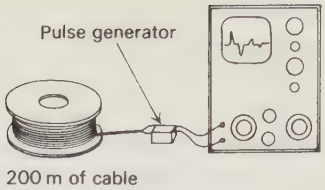


Fig. 41.22

The pulses are sent out at a frequency of about 200 kHz and the time base of the oscilloscope is set to the high speed of 1 cm per microsecond. The separation of the pulses on the oscilloscope tube corresponds to a speed of 2×10^8 metre second^{-1}. This is of the same order of magnitude as $1/\sqrt{\mu_0 \varepsilon_0}$. (The difference can be entirely accounted for by the material insulating one wire from the other, which has a relative permittivity greater than 1.)

We have tried to demonstrate why information of electrical changes takes time to pass down a pair of metallic conductors (a *wave guide*). We could have

done this using nothing more than the currents and voltages along the plates and their own self-capacitance and self-inductance. However, we choose to turn our attention instead to the fields within the gap between the plates and we saw that a changing magnetic field could itself set up an electric field, while conversely a changing electric field could set up a magnetic field.

The calculations were performed on the assumption that the electric and magnetic fields so set up must correspond to the fields accounted for by the charges and currents on the plates.

However, the setting up of these forward-going fields was not dependent on those charges and currents. The charges and currents merely served to maintain the steady fields behind the edge.

Within the edge itself, the magnetic field grows alongside the electric field according to the relationship

$$B = \mu_0 \varepsilon_0 E v.$$

This growing magnetic field is itself changing and can set up an electric field according to the relationship

$$E' = - Bv.$$

The directions are such that this new field E' is in the same direction as E. You can check this for yourself.

If $E' = E$, then the changing magnetic field and the changing electric field will be self-maintaining within the edge. Now $E' = E$, only if

$$v = 1/\sqrt{(\mu_0 \varepsilon_0)}.$$

You will notice that this only applies if both E and B are changing. Indeed you may have realized that the rate of change of E and B must itself be changing with time—or in other words d^2E/dt^2 and d^2B/dt^2 must be non-zero. If they were zero, dE/dt would be constant and hence dB/dt would be zero and vice versa—there would be no self-maintaining edge. This in no way invalidates our previous arguments, for within the space of the field edge, the charges in the plates are accelerating, which is sufficient to ensure that d^2E/dt^2 is non-zero.

Since a pulse of electric field can be regarded as two, "field edges" back-to-back it may be possible to transmit a pulse or a succession of pulses of electric

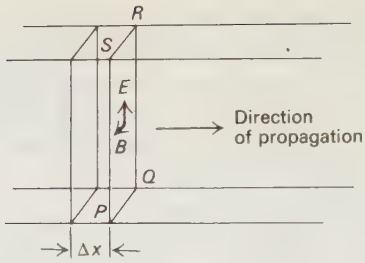


Fig. 41.23

and magnetic field through space, away from the support of any charges and currents. Such a disturbance would be an electromagnetic wave. This guess is given further credibility by the fact that there is a whole range of waves which do indeed travel at just the speed of the field edge down the wave guide. Light is an example of such a wave.

Furthermore such waves are likely to be transverse in character, since the direction of both the electric and magnetic fields are at right angles to the direction of propagation of the pulses.

It is beyond the scope of this Unit to show that Maxwell's equations are capable of realizing such a prediction, but we may guess that the mechanism of their propagation is very closely linked with the way the field edge was propagated down a wave-guide.

It is easy to see, however, that such an electromagnetic wave will carry energy with it.

We have seen (Eq. 39.6) that an electric field E has a store of energy associated with it of $\frac{1}{2}\varepsilon_0 E^2$ per unit volume. Similarly (Eq. 40.9) the magnetic field B has a store of energy of $\frac{1}{2}(B^2/\mu_0)$ per unit volume.

Let us suppose that we have a plane electromagnetic wave of cross-sectional area A travelling with velocity c as shown in Fig. 41.23.

Within a thin slice of thickness Δx of space containing this wave, the electric intensity E and the magnetic field intensity B are essentially uniform provided that Δx is small compared with the wave length. The total energy contained within such a slice will be

$$\left(\frac{1}{2}\varepsilon_0 E^2 + \frac{1}{2}\frac{1}{\mu_0}B^2\right) A \, \Delta x. \qquad (41.13)$$

Since $E = Bc$ from Eq. (41.10) and $c = 1/\sqrt{\varepsilon_0\mu_0}$ from Eq. (41.11)

$$B = E\sqrt{\varepsilon_0\mu_0}.$$

Substituting in Eq. (41.13) we have

$$\begin{aligned}\text{energy} &= \left\{\frac{1}{2}\varepsilon_0 E^2 + \frac{1}{2}\frac{1}{\mu_0}(E\sqrt{\varepsilon_0\mu_0})^2\right\} A\,\Delta x \\ &= \{\tfrac{1}{2}\varepsilon_0 E^2 + \tfrac{1}{2}\varepsilon_0 E^2\} A\,\Delta x \\ &= \varepsilon_0 E^2 A\,\Delta x\end{aligned}$$

and we can see that, to this total store of energy, the magnetic and the electric fields contribute equally.

Now this energy will all pass through the face $PQRS$ of the slice in time $\Delta x/c$. Therefore the rate of passage of energy through the face $PQRS$ is

$$\begin{aligned}&\frac{\varepsilon_0 E^2 A\,\Delta x}{\Delta x/c} \\ &= c\varepsilon_0 E^2 A.\end{aligned}$$

A rate of transfer of energy is a measure of the power and so we have the power S transmitted through unit area of a plane travelling electromagnetic wave given by

$$S = c\varepsilon_0 E^2. \tag{41.14}$$

We can represent this power by a vector whose magnitude is S and whose direction is the direction in which the power is propagated, i.e., the direction of travel of the wave. Such a vector is known as *Poynting's vector* and we have now worked out its value for the special case of a plane travelling wave.

Since both the electric field and the magnetic field fluctuate with time, the power S will also fluctuate with time. If E changes sinusoidally, we may write

$$E = E_0 \sin wt.$$

Now we have seen in Unit 6 that the average value of E^2 taken over a complete cycle will be $\frac{1}{2}E_0^2$. Consequently the mean radiated power per unit area (i.e., the intensity) is

$$\bar{S} = \tfrac{1}{2}c\varepsilon_0 E_0^2. \tag{41.15}$$

We can now find an expression for the electric field strength E_r at relatively large distances from a point source in terms of the total power radiated, since at such distances the wave will behave as though it were plane. If we assume that the total power radiated is P_0 and that this power is radiated uniformly in all directions (see Fig. 41.24) the power passing through unit area at a distance r from the source is

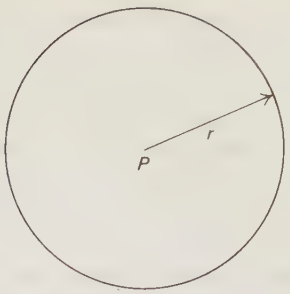


Fig. 41.24

$$\bar{S} = \frac{P_0}{4\pi r^2}.$$

Substituting from Eq. (41.15) we have

$$\tfrac{1}{2}c\varepsilon_0 E_r^2 = \frac{P_0}{4\pi r^2}$$

whence

$$E_r = \frac{1}{r}\sqrt{\frac{P_0}{2\pi c\varepsilon_0}}. \tag{41.16}$$

Problem 41.1 Calculate the electric field strength at a distance of 5 m from a sodium lamp which is radiating monochromatic light of 5 W radiated power.

If we assume that the lamp is small we can apply Eq. (41.16)

$$E_r = \frac{1}{5}\sqrt{\frac{5}{2\pi \times 3 \times 10^8 \varepsilon_0}}.$$

Now

$$\varepsilon_0^{-1} = 36\pi \times 10^9 \text{ J c}^{-2} \text{ m}^{-2}$$

and so

$$\begin{aligned}E_r &= \frac{1}{5}\sqrt{\frac{5 \times 36\pi \times 10^9}{2\pi \times 3 \times 10^8}} \\ &\approx 3.5 \text{ V m}^{-1}.\end{aligned}$$

41.7 THE SPECIAL THEORY OF RELATIVITY

There is much experimental evidence to suggest that the speed of light measured in a vacuum is independent of the motion of either the source or the observer. That it is independent of the motion of the source is characteristic of a wave motion. For example, the speed of sound is independent of the motion of its source. That the speed of light is independent of the speed of the observer or apparatus used to measure it came as a great surprise. We recall that the speed of sound relative to an observer is not independent of the motion of the observer.

As if this were not enough, evidence now exists to suggest that the speed of light in a vacuum is a limiting speed for the movement of all material objects.

The consequences of this experimental fact are far reaching. Our concepts of force and energy lead us to believe that the transference of energy can be measured in terms of forces which act during this transference and of the distance moved by their point of application. This leads us to equate the kinetic energy of a moving body with $\frac{1}{2}mv^2$.

An electron accelerated in a suitable linear accelerator can attain a speed of 3×10^8 m s^{-1} (or as close to this as makes no difference). Are we to assume that after reaching this speed that it gains no more energy? Experiment suggests not. The energy transferred to the moving electrons can be measured in terms of their charges and the potential difference through which they pass. The energy they possess as a result of the acceleration can be converted into heat by allowing the electrons to hit a metal target. As the potential difference is raised, so does the energy attained by the electrons rise—and it continues to rise long after their speed has apparently reached 3×10^8 m s^{-1}.

We have defined force as that which changes the motion of a body. Is this basic definition of force (and therefore of energy) wrong? We prefer to think not. Instead we accept other consequences. If the kinetic energy is still to be described in terms of the work done on a body, we find we must assume that the inertial mass of a body is no longer an unchanging entity. Instead, it becomes a function of the body's speed *relative to the person making the measurements.* We usually refer to this person as "the observer."

This need to describe events in relation to particular observers instead of in relation to some ill-defined *absolute* system is the central tenet of the theory of special relativity.

Galileo and Newton recognized that the laws of mechanics were independent of the uniform motion of one observer relative to another. In other words, Newton's laws of motion are equally true whether they are tested in a laboratory fixed on the Earth, in an aircraft moving at a steady speed relative to the Earth, or in a spaceship which is not accelerating relative to the Earth. This truth is often referred to as the *principle of Galilean relativity.*

Einstein recognized that the universality of the speed of light merely represented the extension of the principle of Galilean relativity to the whole of physics. We have seen that the speed of light in a vacuum is related to two fundamental quantities,

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}.$$

This is a law of electromagnetism. We saw that electromagnetic fields are self-propagating when they move at this speed relative to the observer. Electromagnetic laws as well as mechanical laws are independent of the state of motion of an observer provided he is not accelerating relative to the Earth.

It might appear that the Earth is a specially chosen reference system. In a way it is, since all our fundamental experiments have been made in laboratories fixed to it. Now we recognize that even the Earth suffers small accelerations (as it moves around the sun for instance). Scientists find it useful to visualize a frame of reference in which all our assumed laws of physics are exactly true—this is called an *inertial frame of reference.* For most purposes a frame fixed relative to the Earth can be regarded as an inertial frame, however. An observer in a frame of reference which is in uniform motion relative to an *inertial* frame will also (by Einstein's reasoning) be in an inertial frame.



Fig. 41.25 A laser clock.

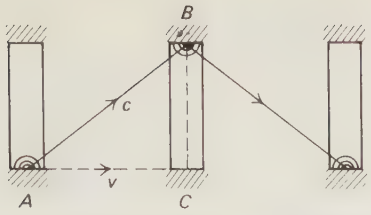


Fig. 41.26 Stationary observer's view of a moving laser clock.

We cannot explore in this book all the consequences of the principle of relativity, but we will look very simply at two of the most important.

First of all we will consider how two different observers in uniform motion relative to each other measure *time*. To do this, we shall suppose that we are one of the observers and that another is moving steadily past us (on a balloon, in a train, or a spacecraft, if you like). We will equip the observer who moves past us with a special type of clock—a *laser clock*. We choose such a clock to illustrate the consequence of the constancy of the speed of light. It can be shown, however, that what is true for *this* clock is true for *all* clocks.

In this laser clock, we imagine a pulse of light bouncing up and down between mirrors at either end which are distance l apart (see Fig. 41.25). We will suppose that every time the pulse hits end A the clock gives out a tick, or moves some clock hands. The clock will have a natural period (to the observer moving with the clock) of

$$t_0 = \frac{2l}{c},$$

where t_0 is a rather special time interval—the time interval the clock records between each pair of ticks according to the observer moving along with it. We call t_0 the *proper time*.

Now we will see how the clock appears to behave from *our* point of view. As the light pulse bounces up and down, so the clock moves steadily to the right at speed v relative to us. The light pulse seems to traverse the line AB as it moves from the bottom mirror to the top. (Fig. 41.26.)

Now $AB:AC::c:v$, since the light pulse moves along AB with speed c *according to us*.

Hence

$$\frac{AB}{BC} = \frac{c}{\sqrt{(c^2 - v^2)}}$$

and

$$AB = \frac{c}{\sqrt{(c^2 - v^2)}}\, l.$$

As the light pulse travels from A to B and returns to A, it seems to us to travel a distance $2AB$ or

$$\frac{2lc}{\sqrt{(c^2 - v^2)}}.$$

But it does this at speed c and thus takes a time

$$t = \frac{2lc}{\sqrt{(c^2 - v^2)}} \frac{1}{c}$$

$$= \frac{2l}{\sqrt{(c^2 - v^2)}}$$

Thus whereas the observer moving with the clock says it ticks with a period

$$t_0 = \frac{2l}{c}$$

we say it ticks with a period

$$t = \frac{2l}{\sqrt{(c^2 - v^2)}}$$

$$\frac{t}{t_0} = \frac{c}{\sqrt{(c^2 - v^2)}}$$

$$= \frac{1}{\sqrt{(1 - v^2/c^2)}}$$

$$\therefore t = \frac{t_0}{\sqrt{(1 - v^2/c^2)}}.$$

Let us suppose that the laser clock ticks seconds according to the observer moving with the clock. We will further suppose by way of example that the clock

and observer move at the high speed relative to us of 0.6 c (that is 1.8×10^8 m s^{-1}).

Then

$$v^2 = 0.36c^2$$

and

$$\frac{v^2}{c^2} = 0.36$$

so that

$$\left(1 - \frac{v^2}{c^2}\right) = 0.64$$

and

$$\sqrt{\left(1 - \frac{v^2}{c^2}\right)} = 0.8$$

$$\therefore t = \frac{t_0}{0.8}$$

$$= 1.25t_0.$$

Thus in one minute of *our* time, the moving clock only appears to make 60/1.25 = 48 ticks. Since each tick interval is called one second by the observer accompanying the clock, *we* say that the moving clock is running slow. It records only 48 s in the time our clock records 60.

There is a good deal of experimental evidence for this apparent running slow of a moving clock—an effect referred to as *time dilation*. Unstable particles called μ-mesons have, according to observations made when they are at rest with respect to us, a half-life of only 2×10^{-6} s. But travelling towards us from the top of the atmosphere with a speed of 0.99 c, they appear to live much longer. During the passage of 1 second of time measured by the moving μ-meson clock, we shall record the elapse of more than 7 seconds. Without this idea we cannot, in fact, account for the number of μ-mesons reaching the lower atmosphere without decaying first. It is on the basis of such arguments and experiments that it has been proposed that a space traveller, setting out from Earth and then returning will take a shorter time over the journey according to *his* clock than we record on *ours*. Consequently it is believed that the traveller will have *aged* less than his Earth-bound friends. The arguments supporting this are more complex than we can give here but many believe this outcome to be true.

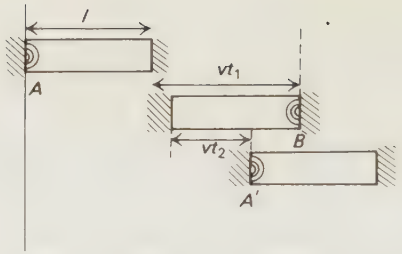


Fig. 41.27

Another consequence of the principle of relativity is to be found in the measurement of length. A body moving past is recorded as having a shorter length *in the direction of travel* (though not sideways!) than an observer moving *with* the body says it has. Again we can use the laser clock to see why this is so (see Fig. 41.27).

A pulse of light starts out from A, is reflected at B and returns to A'. According to an observer moving along with the clock this takes a time

$$t_0 = \frac{2l_0}{c}.$$

Consequently

$$l_0 = \frac{t_0 c}{2},$$

where l_0 is the length of a body according to an observer moving along with it and is called the *proper length*. To ourselves, the pulse takes a time t_1 to move from A to B (Fig. 41.28) such that

$$ct_1 = l + vt_1$$

$$t_1 = \frac{l}{(c - v)}.$$

It takes a further time t_2 to go from B back to A'

$$ct_2 = l - vt_2$$

$$t_2 = \frac{l}{(c + v)}.$$

So the time t the pulse takes to move from one end and back again is

$$t = t_1 + t_2$$

$$= \frac{l}{(c + v)} + \frac{l}{(c - v)}$$

$$= \frac{2lc}{(c^2 - v^2)}.$$

But we have seen already that

$$t = \frac{t_0}{\sqrt{(1 - v^2/c^2)}}.$$

So

$$\frac{t_0}{\sqrt{(1 - v^2/c^2)}} = \frac{2lc}{(c^2 - v^2)}.$$

Since

$$t_0 = \frac{2l_0}{c}$$

$$\frac{2l_0/c}{\sqrt{(1 - v^2/c^2)}} = \frac{2lc}{(c^2 - v^2)}$$

$$\frac{l_0}{\sqrt{(1 - v^2/c^2)}} = \frac{lc^2}{(c^2 - v^2)}$$

$$= \frac{l}{(1 - v^2/c^2)}$$

$$\therefore l_0 = \frac{l}{\sqrt{(1 - v^2/c^2)}}.$$

Since

$$\sqrt{\left(1 - \frac{v^2}{c^2}\right)} \leq 1, \qquad l_0 \geq l.$$

Thus a moving length always appears to be shorter than an observer moving *with* the length would say.

For an apparent contraction of 1% of the proper length

$$\frac{l}{l_0} = \frac{99}{100}.$$

Consequently,

$$\sqrt{\left(1 - \frac{v^2}{c^2}\right)} = \frac{99}{100}$$

$$1 - \frac{v^2}{c^2} = (0.99)^2 = 0.980$$

$$\therefore \frac{v^2}{c^2} = 0.020$$

$$v^2 = 0.02c^2$$

and

$$v = 0.14c.$$

The moving length must be travelling at a speed of 14% of the speed of light relative to the observer making the measurements—about 4×10^6 m s^{-1}. A spacecraft, escaping from the Earth's gravitational field travels at only 1.1×10^4 m s^{-1}. For all normal speeds length contraction effects are very small indeed.

Despite these effects of length contraction and time dilation, two observers moving past each other will always agree about their relative speed. If we form one set of observers, we measure the moving observer's speed by comparing the time he takes (according to *our* clocks) to move a length measured in *our* frame of reference. If the same measurements are made by the moving observer, using time in his frame of reference and lengths in his frame of reference, we shall claim that his measured lengths are too long, but his measured time intervals are too short. But as you can see by looking at the equations for time dilation and length contraction, these two effects *just cancel each other*, so both of us record the same speed of relative motion.

Why should the speed of light be a limiting speed? This is a direct consequence of our method of measuring speeds—we are compelled to make all measurements in our own frame of reference. Thus we assess a body's speed in terms of the distance it travels in our frame (as *proper length*) and the time it takes in *our frame of reference*. This latter is not a *proper time*—we know now that our clock appears to run progressively faster compared with a clock carried by the moving body, the faster that body moves. As

relative speeds approach c, the time dilation factor approaches infinity and consequently the relative speed of all bodies approaches c, the speed of light.

We started this brief introduction to the special theory of relativity by observing the surprising consequences this limiting speed c has on our ideas about inertial mass. The consequent relationship between mass and relative speed is harder to derive than the relationships we have obtained for time and length with speed, and is beyond the scope of this book. It is from the relationship between mass and relative speed that Einstein was able to show a link between the energy carried by a moving body and its apparent mass, now celebrated in the famous equation $E = mc^2$. A full discussion of the implications contained in this equation is beyond us here, but one observation might form a valuable conclusion to this brief discussion of the special theory of relativity.

From all that we have said so far, it might appear that the relationship between measured masses, lengths and time intervals and the relative speeds of observers is an artificial contrivance, designed by physicists in order to preserve a few well-known laws. We have already been able to distinguish some special lengths and time intervals called *proper length* and *proper time* which are measurements made by observers staionary relative to the length measured, or clock used. In a similar way, we are able to define a *rest mass* of an object, as the mass of the object relative to which the observer is stationary.

It was Einstein who first proposed that the relationship between energy and the *extra mass* of a body moving relative to an observer could be extended to the "rest mass" as well. In other words all mass is equivalent to a certain amount of energy—not just the apparently fictitious extra mass invented by physicists to preserve their precious laws. That this was true was violently demonstrated by the release of atomic (or, as we should say, nuclear) energy. The energy which can be released from the breakdown of a nucleus is related to the change in mass observed by the relativity equation $E = mc^2$.

It is this fact more than any other which has served both to demonstrate that the special theory of relativity is something more than an exercise in mass and energy book-keeping and to turn what may seem an esoteric exercise into something which is the "bread and butter" of a particle-physicist's life.

PROBLEMS: Unit Seven

7.1 Estimate the size of the gravitational attraction between two people standing about a metre apart.

7.2 Two light, metallized spheres, each of mass m, each carry negative charges of q. They are suspended by light nylon threads each of length a. Show that the inclination θ of each thread to the vertical is related to m and q be the expression

$$16\pi mg\varepsilon_0 a^2 \sin^3 \theta = q^2 \cos \theta.$$

In a typical experiment, θ is found to be 6°. The mass of each sphere is 1.23×10^{-4} kg and a is 0.025 m.

a) Calculate the charge carried by each sphere.
b) Estimate the total number of excess electrons carried by each sphere.

7.3 Estimate the strength of the electric field near to the surface of the nucleus of an element near the middle of the periodic table.

7.4 The electric field below a thunder cloud can be as large as 10^4 V m^{-1}. (At larger field strengths, air molecules ionize and the charge leaks away.)

Choose plausible sizes for very small and very large raindrops. What charges would be required on these extreme sizes in order to support them against the attraction of gravity?

Is it likely that charges of these magnitudes could be carried on these drops for any length of time?

If not, why not?

7.5 A current balance, initially in equilibrium, is placed between the poles of a magnet as in Fig. 33.10. When a current of 1.85 A passes through the balance, 50 cm of paper tape had to be hung over the balance to restore equilibrium. 5 m of the tape was found to have a mass of 4.23 g and 0.05 m of the current balance wire was estimated to be in the field of the magnet.

Calculate the magnetic flux density.

Without altering the current through the wire, the magnet is turned as in Fig. P7.1. What do you think will be likely to happen to the current balance? Will the end in the field of the magnet move up, indicating more force acting on it, down indicating less, or will it stay in the same place? Explain your answer.

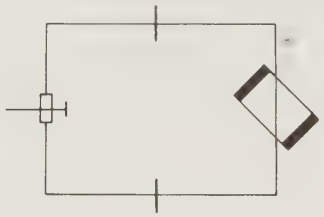


Fig. P7.1

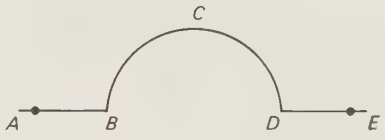


Fig. P7.2

7.6 A wire $ABCDE$ is bent into the shape shown in Fig. P7.2 where BCD is a semicircle of radius, r, and AB and D are of unit length. A current I E is passed through the wire which is placed in a uniform magnetic field B that emerges from the plane of the paper.

a) What force does the wire experience?
b) In which direction does the force act?

If the mean drift speed of the electrons in the wire is v write down an expression for the average force acting on an electron as it drifts from B to D.

7.7 It is possible in a cloud chamber or similar apparatus to observe the path followed by a high speed electron in a uniform magnetic field. (Fig. P7.3). The electron moves in the plane of the diagram.

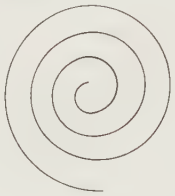


Fig. P7.3

a) What is the direction of the magnetic field?
b) What is the probable direction of movement of the electron?
c) Suggest and explain a reason why the track is a flat spiral and not a circle.

7.8 An electric field of 2.0×10^6 V m^{-1} is found to exert the same sideways deflecting force as a magnetic field of 10^{-2} Wb m^{-2} on beta particles travelling at right angles to the fields.

How fast are the beta particles going?

What further measurements would you make in order to determine e/m (ratio of charge to mass) for these particles?

Would you expect the result to be exactly the same as that found for the electron in a typical small thermionic vacuum tube?

University of Birmingham School of Education Degree of B.Ed, Third Year Examinations (1968).

7.9 A spherical rubber balloon carries a charge that is uniformly distributed over its surface. How does the electric field strength, E, vary for points

a) inside the balloon,
b) at the surface of the balloon,
c) outside the balloon,
d) as the balloon is blown up?

7.10 This is a question about the gravitational field of the Earth at points in its interior (Fig. P7.4). To calculate a value for this field strength at a point P, distance r from the centre of the Earth, proceed as follows.

a) What is the gravitational field strength at P due to the shaded mass which lies outside a sphere of radius r about the centre of the Earth?
b) What is the field strength at P due to the mass m of the Earth within the sphere of radius r? (Use Gauss's law (Chapter 35) to answer both these parts.)
c) Assuming the Earth is of uniform density write down an expression to show the gravitational field strength, g, as a function of r.
d) Plot a graph of g against r.
e) Experiment shows that the gravitational field at the bottom of deep mine shafts is usually greater than its value at the surface. Compare this result with the graph you have drawn in (d) and try to offer an explanation for the difference between this observation and your graph.

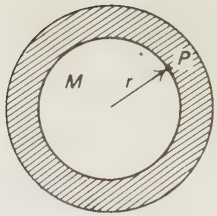


Fig. P7.4

7.11 A new energy resource could be realised by bringing about the fusion of two deuterium atoms. (Deuterium is "heavy hydrogen"—hydrogen atoms whose nucleus consists of one proton and one neutron.) In the fusion process, the deuterium would form helium and release considerable quantities of energy. However, the nuclear forces are short range and two deuterium nuclei would have to be brought to within 10^{-14} m of each other.

a) Calculate the electrical potential energy which would be gained by one deuterium nucleus if it were brought this close to another. (One way of doing this is to given the deuterium atoms kinetic energy sufficient to provide this potential energy by heating up the deuterium gas.)
b) Assuming that all molecules in a gas travel at the same speed, estimate the temperature to which the gas would have to be raised to provide sufficient kinetic energy to the molecules.
c) Explain why we might expect some products from a fusion reaction at temperatures much lower than the one you have worked out in (b).

7.12 Derive an expression for the speed that must be given to a particle near the Earth's surface in order that it

may escape from the pull of the Earth's gravitational field and hence compute its size. At what temperature is the r.m.s. velocity of hydrogen molecules equal to the escape velocity from the Earth?

The gas kinetic temperature in the upper atmosphere is of the order of 1000 K. Would you expect to find (a) much hydrogen and (b) much oxygen there? Explain how you arrive at your answer.

7.13 In a worked problem in Chapter 39 we derived a value for ε_0, but commented that errors arose from the fact that the metal plates which together made up the parallel plate capacitor were not isolated as theory demands and from the non-uniformity of the field. However, if a set of measurements of Q, the charge on the plates is taken for various values of the area of overlap, A, of the plates, these errors can be eliminated to some extent. Here is a set of such results. By plotting a suitable graph, estimate a value for ε_0.

Area in m^2	Charge in C
2.25×10^{-2}	4.0×10^{-8}
3.75×10^{-2}	5.6×10^{-8}
5.00×10^{-2}	7.8×10^{-8}
6.25×10^{-2}	9.4×10^{-8}

7.14 A football, 17.1 cm in diameter, is covered with aluminium foil. It is raised to a measured potential V and then discharged through an electrometer. The electrometer records a charge of Q coulombs. The following table gives a set of results relating V to Q:

V volts	$Q \times 10^{-9}$ coulombs
0	0
100	1.5
200	2.7
300	4.1
400	5.4
500	6.4
600	8.0

Plot a graph of V against Q.

a) What information does the graph give about the relationship between V and Q?
b) Write down a theoretical relationship between V, the potential at the surface of a charged conducting isolated sphere and Q the charge carried by the sphere.
c) Use this expression and information derived from the slope of the graph and the ball's diameter to find a value for ε_0.
d) What do you think are the likely inaccuracies in this experiment as a means of measuring ε_0?

7.15 A loosely-wound solenoid of 36 turns is 0.36 m long. It is aligned with its axis in an east–west direction. A small magnetic compass is placed inside the solenoid at its centre. When a current of 0.11 A is passed through the solenoid, the compass points in a north-westerly direction. How strong is the horizontal component of the Earth's magnetic field at this point? (You may assume that the material from which the solenoid is wound has no effect on the compass.)

In what direction would you expect the compass to point (a) outside the solenoid and (b) at one end of the solenoid when the same current passes through it?

7.16 The Earth has a permanent magnetic field whose direction over the United Kingdom makes an angle of 20° with the vertical. Its magnitude is about 5×10^{-5} Wb m^{-2}. Estimate the electric potential difference set up between Dover and Calais, two ports on opposite sides of the English channel, when the tide is running due north at 2 m s^{-1}. (Assume that Calais is 35 km due east of Dover.)

How could you attempt to check this value experimentally?

7.17 A simple electric motor consists of an armature of 10 turns wound as a rectangular coil 5 cm × 3 cm (Fig. P7.5). The armature is pivoted to rotate in the permanent magnetic field whose flux density was calculated in Problem 7.5.

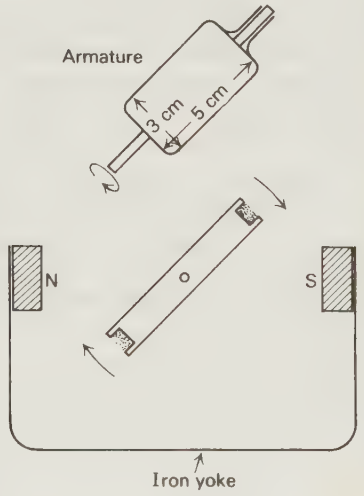


Fig. P7.5

With the armature rotating at 28 revolutions per second, the peak current through it was 2.1 A and the corresponding p.d. across it 0.64 V. Calculate the value of each of the following:

a) the maximum force acting on one side of the armature,
b) the maximum induced e.m.f. in the armature.

Devise a suitable experimental set-up for measuring the peak current passing through the armature and the corresponding p.d. across it.

7.18 A charged particle accelerator called a Betatron was briefly described in Chapter 40. In one particular machine, electrons move in an orbit of 1 m diameter at speeds very close to that of light. The steady magnetic field acting to maintain this orbit is specially shaped to maintain this orbit even though the speed and mass of the electrons may rise as their energy increases. Since, however, the electrons are injected into the machine at speeds close to that of light, their speed cannot rise much even though they gain energy in the accelerator—in fact "accelerator" is something of a misnomer; "energizer" would perhaps be better.

a) Assume the electrons travel at the speed of light. Calculate the size of the steady magnetic field needed to maintain the orbit.
b) Draw a sketch to show the direction in which the steady field acts.
c) Suppose that the magnetic flux through the plane of the orbit increases by a further 2.0 Wb in 5×10^{-3} s at a uniform rate. Calculate the value of the electric field acting at the orbit of the electrons.
d) How far will the electrons travel in this time?
e) How much will the energy of each electron increase during the time the magnetic flux is increasing?
f) What is the electric potential difference in volts through which the electrons would have to pass if they were to gain the same amount of energy?

7.19 Fig. P7.6 shows a length of coaxial cable, made of an outer sheath of conductor wrapped round a long roll of insulator, inside which is a conducting wire running along the axis of the whole cable. The capacitance of each metre of cable is 200×10^{-12} F. Some time before the instant shown, the switch S was switched to the 1.5 V battery for 10^{-9} s and then back to the position shown. At the time shown, in the region BC the inner wire carries positive charge and the outer conductor carries negative charge. The regions AB, CD are uncharged.

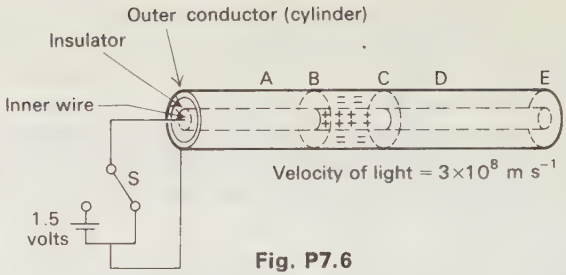


Fig. P7.6

a) Give a reason why no electricity has yet reached the distant place E.
b) Estimate roughly the length of the charged region BC (the insulator does actually reduce the velocity, but you may ignore this.)
c) Make a sketch showing the directions of the electric field around the central wire within the region BC.
d) Explain why the charge on each conductor in the region BC is about 90×10^{-12} C (again neglect any effect of the insulator).
e) What do you think will happen when the charged region reaches the end of the cable beyond B if the conductors end abruptly and are not connected to anything else or to each other?

Oxford and Cambridge Schools Examination Board Nuffield A-Level (1970).

7.20 Show that the expressions

$$S = \sqrt{\frac{\varepsilon_0}{\mu_0}} \cdot E^2 \qquad S = \frac{cB^2}{\mu_0} \qquad \text{and} \qquad S = \frac{1}{\mu_0} EB$$

are all equivalent to $S = c\varepsilon_0 E^2$ which describes the power radiated per unit area in a plane travelling wave.

Calculate the strength of the magnetic field at a distance of 5 m from a sodium lamp radiating monochromatic light equally in all directions at a power of 5 W.

What reasons could be advanced for expecting measured values of the electric and magnetic fields to be different from the calculated values?

BIBLIOGRAPHY: Unit Seven

Bennet, G. A. G. (1968). *Electricity and Modern Physics* (mks version), Arnold.
This is a detailed account which, while going beyond the scope of the present Unit, contains a careful and detailed consideration of many of the matters raised.

Campbell, N. R. (1913). *Modern Electrical Theory*. Cambridge University Press.

Einstein, A., and Infeld, L. (1961). *The Evolution of Physics*, Cambridge University Press.
This account, first published in 1938, traces the emergence of the concepts of relativity theory and quantum theory for a lay reader. In doing so, it develops the ideas involved in Maxwell's theory without going into difficult mathematics.

Feather, N. (1968). *Electricity and Matter*, Edinburgh University Press.
Chapter 2 gives a detailed account of inverse square forces in electromagnetism and the derivation of Gauss's law. Chapter 3 considers the matter of electrical units and thus develops the relationship between electricity and magnetism. Conceptually these two chapters go beyond the present Unit, but they should be read by any one wishing to get a deeper understanding of the flux model for electromagnetic fields.

Feynman, R. P., Leighton, R. B., and Sands, M. (1964). *The Feynman Lectures on Physics*, Addison-Wesley.
Volume II is concerned with electromagnetic theory. As a text it will be found to be far in advance of both this Unit and any other of the references. However, some of the discussions of fundamental principles within the text can usefully be read at this level. In particular, reference should be made to Chapters 16 and 17 dealing with induced currents and the laws of induction. These contain a careful and profound analysis of Faraday's Law of electromagnetic induction. Do not be afraid to miss out the mathematical analysis in reading it.

French, A. P. (1968). *Special Relativity*, Nelson.
Chapters 1 to 4 deal in detail with the material included in the brief introduction to special relativity in this Unit and are to be recommended.

Koestler, A. (1959). *The Sleepwalkers*, Hutchinson.
This is a classic account of the interplay of imaginative speculation and objective measurement which marked the evolution of Newton's theory of gravitation.

Maxwell, J. C. (1873). *Proc. Roy. Inst.*, **7**, 44–54.

Miller. *Measurement of G*. A film loop by the Ealing Corporation.

Newman, J. R. (1955). *James Clerk Maxwell*, Scientific American.
This is a biographical essay which traces the historical evolution of Maxwell's theory of the electromagnetic field.

Niven, W. D. (Ed.) (1965). *The Scientific Papers of James Clerk Maxwell*, Dover.
"Action at a distance," Maxwell's own account, written in 1873, of the basic ideas which still underlie the field concept, is included.

Pohl, R. W. (1930). *The Physical Principles of Electricity and Magnetism*, Blackie.
This book is important for its experimental approach. It includes an account of the flame probe. It should be pointed out that the system of units used in the text is quite different from that used in this Unit or in any other of the references.

Resnick, R., and Halliday, D. (1966). *Physics*, Wiley.
Chapters 26–39 inclusive develop the topic of electromagnetism and electromagnetic waves and give many illustrative examples and applications of the theory.

Rogers, E. M. (1960). *Physics for the Inquiring Mind*, Oxford University Press.
This is a generally valuable reference, but it is quoted here specifically for the historical development of Newton's theory of gravitation. This account is also important for the way it shows the details of physics emerging in a quest for the answer to wider, more far-reaching issues.

Tricker, R. A. R. (1966). *The Contributions of Faraday and Maxwell to Electrical Science*, Pergamon.
A useful historical survey which can be used by those who wish to obtain a fuller background to the emergence of the field concept in electromagnetism.

Wilson, R. R., and Littauer, R. (1962). *Accelerators*, Heinemann.
This small book in the Science Study Series is a generally useful reference, and contains a fuller account of the Betatron than is given in the text of this Unit.

UNIT EIGHT
Electricity and Matter

42 Electric Currents
43 The Electron
44 A Model for an Electric Current
45 Ionization and Energy Levels
46 Electricity, Matter, and Light

Chapter 42
ELECTRIC CURRENTS

42.1 Introduction
42.2 Evidence for an ionic model
42.3 Summary

42.1 INTRODUCTION

When the two terminals of an electric battery are connected by wire to electrodes separated by a liquid, such as a dilute acid or a solution of a salt, changes occur. The circuit is now surrounded by a magnetic field; the components warm up; chemical action takes place in the conducting liquid and in the battery itself. These effects are related to the electric current.

The word "current" suggests a flow—in this case, a flow of charges. As we have seen in Unit 6 measurements of the current do suggest a behaviour not unlike that which we associate with other examples of a flow. In the simple series circuit for example, the current is the same wherever it is measured. At junctions in circuits (Fig. 42.1) the current divides or combines so that as much current enters the junction as leaves it.

It is possible to build a simple model of such a current of electric charge. And when we add to such a model the concept of the volt as a measure of the energy associated with a definite quantity of electric charge (the coulomb), it becomes very tempting to elaborate this model of electrical "stuff" circulating round a circuit, transferring energy in chemical reaction, in heating, and in causing motion through the interaction of magnetic forces. It is not unlike a constant procession of bottles which start full of milk and gradually lose it as they travel from the dairy to the home and back again round the circuit; or, maybe, a fluid which flows around carrying energy as hot water in a heating system carries energy. Imagery of this sort is useful because it provides a framework within which to think about physical problems. In some cases the model serves us well for a long time as, for example, has the kinetic theory of gases. In other cases, the model is soon discarded as, for example, Maxwell's own mechanical model of the electromagnetic field.

The electrical phenomena of the simple circuit force us to ask questions about electricity which we are in no position to answer at this stage. Among them: what is it that flows in a circuit? If charge, is it positive or negative or both? How fast does it flow?

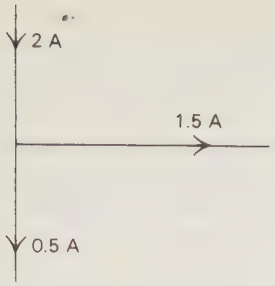


Fig. 42.1

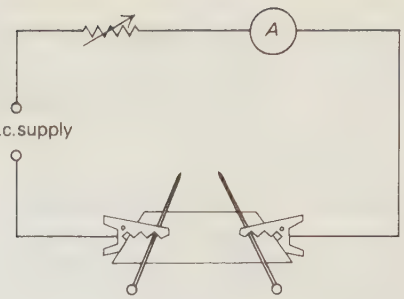


Fig. 42.2

What is electric charge—a form of matter or a property of matter? Is it particulate as is matter?

It is the purpose of this Unit to suggest tentative answers to such questions.

42.2 EVIDENCE FOR AN IONIC MODEL

One of the difficulties we experience in building a convincing model of the electric current is that we can have no direct experience of what is going on. All our evidence is indirect. But some can come from experiments in which the passage of electricity is associated with the movement of matter.

A strip of filter paper which has been moistened with ammonia solution is supported on a glass slide and connected to a d.c. power supply by means of spring clips (e.g., crocodile clips) (Fig. 42.2). A few small crystals of copper sulphate and of potassium permanganate are scattered over the surface of the paper and about 100 volts applied to the electrodes. The coloured stains which spread out from the crystals lack symmetry, the blue stains from the copper sulphate spreading towards the negative electrode and the purple stain from the potassium permanganate spreading towards the positive electrode. Switch the supply off, and the stains spread symmetrically. What do you imagine will happen if you reverse the polarity of the connections and again switch on?

The blue and the purple stains are related chemically to the atoms making up the crystals. Their spread towards the positive and negative terminals of the battery suggests that the passage of the electric current is associated with a movement of atoms from the crystals within the dampened filter paper.

It seems likely that the carriage of electric charge across the paper is, in part at least, associated with the atoms or groups of atoms which cause the coloured stains. Such electrically charged atoms or groups of atoms are called *ions*. This conclusion does not, of course, exclude other processes and it may well be that some other mechanisms are also responsible.

The chemical reactions resulting from the passage of electric charge through conducting liquids were extensively investigated by Michael Faraday, whose name is now remembered in the two laws of electrolysis. He introduced technical terms into his accounts of his researches which are still in use. Electrically conducting liquids (molten or solutions) in which chemical reactions take place are called *electrolytes*. The conductors (usually solid metal or carbon) which connect the electrolyte to the rest of the circuit are *electrodes*, that which is connected to the positive pole of the supply being the *anode* and that which is connected to the negative pole the *cathode*. Faraday is also responsible for the application of the word *ion* to the charged particles within the solution. Those carrying a positive charge or charges (e.g., H^+, Cu^{2+}, Na^+) which move towards the cathode are often called *cations*. Those with a negative charge (e.g., O^{2-}, Cl^-, SO_4^{2-}) which move towards the anode are called *anions*.

As an example of an electrolytic reaction, consider the passage of an electric current through a solution of copper sulphate. A 12 V d.c. supply is connected in series with an ammeter, a rheostat and two copper

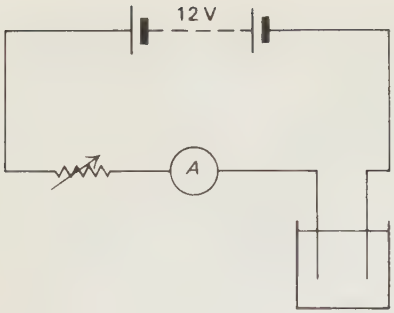


Fig. 42.3

electrodes supported in a beaker containing a strong solution of copper sulphate (see Fig. 42.3).

The copper cathode is carefully cleaned with an abrasive cloth and weighed before connection into the circuit. A current of 1 A is passed through the circuit for, say, 10 minutes. The cathode is removed, dried and re-weighed. This procedure is repeated for a range of values of both current and time.

The experimental results show that the mass (m) of copper deposited is proportional to the current (I) for constant times and to the time (t) for constant currents. So

$$m \propto I, \text{ with } t \text{ constant,}$$
$$m \propto t, \text{ with } I \text{ constant,}$$
$$\therefore\ m \propto It.$$

Now the product It is a measure of the quantity of charge passing through the electrolyte.

In general it is found that the mass of material liberated at the anode or the cathode in electrolysis is proportional to the quantity of charge passing. This statement expresses Faraday's first law of electrolysis.

It is difficult to avoid making the hypothesis that the electric charge passing through the electrolytic cell is carried by particles of matter—Faraday's ions.

The case chosen above is a simple one, for only copper appears as an end product. The experiment can readily be adapted to investigate any changes which may occur at the anode. You may wish to investigate this yourself. You could try weighing the anode before and after a charge is passed.

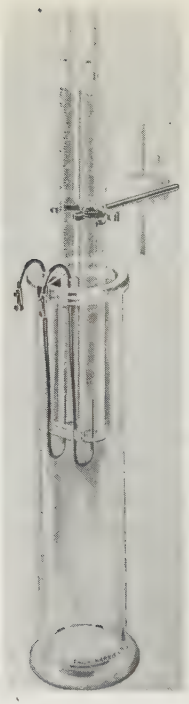


Fig. 42.4 A gas voltameter. (Courtesy of Philip Harris, Ltd.)

A more complex example involves the electrolysis of water. Now pure water is a very poor conductor of electricity whereas water containing a small quantity of an electrolyte such as sulphuric acid is a very good conductor. It seems that the acid forms ions in the water and these provide the charge carriers. But the complex reactions which occur at the anode and cathode result in the liberation of oxygen and hydrogen, the constituents of water.

Problem 42.1 Figure 42.4 shows a cell suitable for the electrolysis of acidulated water. Such cells are sometimes called *voltameters*. The electrodes are made of materials which do not react with oxygen—for example, platinum, or pure nickel.

This cell is connected into the simple circuit of the last experiment in place of the copper plating bath and a constant current (0.50 A) passed for a

known time (45 minutes). In such an experiment 174 cm^3 of hydrogen were collected in the burette placed over the cathode and 84 cm^3 of oxygen in the burette placed over the anode.

These volumes were measured over water at atmospheric pressure by levelling the liquid inside and outside the burettes. The atmospheric pressure was 102 910 N m^{-2} (771 mmHg) and the temperature 23.5° C.

Calculate the volumes of hydrogen and oxygen which would have been collected at standard temperature and pressure (making due allowance for the partial pressure of the water vapour). Standard temperature and pressure may be taken to be 273 K and 101 300 N m^{-2} (760 mmHg).

Now the density of hydrogen at s.t.p. is 0.090 kg m^{-3} while that of oxygen is 1.43 kg m^{-3}. Calculate next the masses of hydrogen and oxygen liberated in the electrolysis experiment.

What is the likely error in this estimation?

You should have found that the mass of hydrogen liberated was 14.3×10^{-6} kg and that the mass of oxygen was 109×10^{-6} kg. This gives a ratio of the mass of hydrogen to the mass of oxygen of 1.05 to 8.

This is almost identical to the ratio of the masses of hydrogen and oxygen which combine in chemical reaction and is just one half of the ratio of the atomic masses of hydrogen and oxygen (1.08:16).

Faraday and many other workers in this field made many such comparisons of the masses of atoms involved in electrolysis and found that the quantities of electricity required to liberate one gramme-atom of different elements bear a simple relationship to one another. This is a statement of *Faraday's second law of electrolysis.*

The atomic hypothesis suggests that a gramme-atom of an element will contain the same number of atoms as a gramme-atom of any other element. This number is referred to as *Avogadro's number* (N_A), (or Loschmidt's number), but to specify its existence is easier than to specify its size. Indeed its value was never known to Avogadro.

Problem 42.2 From the observations given earlier, calculate the quantity of electricity needed to liberate 1.08 g of hydrogen and 16 g of oxygen.

Accurate experiments have given a value of 96 500 C as the charge needed to liberate 1 gramme-atom of hydrogen and 193 000 C as the charge needed to liberate 1 gramme-atom of oxygen. Do your own calculations confirm these values within the limits of the experimental error?

Since 96 500 C are needed to liberate 1 gramme-atom (1.08 g) of hydrogen, it follows that each hydrogen ion (charged atom) in the acidulated water carries on the average $96\,500/N_A$ C. The evidence will not allow us to say that each ion carries exactly this charge; but there is no good reason why different hydrogen ions should carry different charges so we shall assume that the charges are identical and are each $96\,500/N_A$ C.

Building an assumption such as this into the model is an acceptable device so long as we remember that the assumption was made. In Chapter 13 when we discussed the speed of each molecule in considering a kinetic model for gases, we were careful to point out that the measured speed was only an average of a wide range of possible speeds. In that instance, that was the most reasonable assumption to make for it was implicit that the molecular velocities were changing constantly as a consequence of the many collisons which were taking place, and even had they all been the same at one moment, they would soon have become randomized.

The charge $96\,500/N_A$ C carried by each hydrogen ion is the smallest quantity of charge associated with any single ion. Let us write it e. Then, extending the hypothesis, the charge carried by each oxygen ion is $-2e$. The model can be extended to include all ions, some of which may be complex structures.

Using such an ionic picture, R. J. Clausius in 1857 advanced a model for electric conduction in an electrolyte. He suggested that the molecules of solutes spontaneously dissociated into ions within the solvent and that a continuous process of dissociation

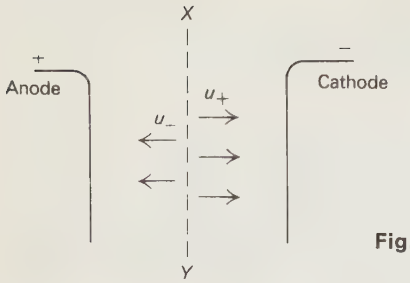


Fig. 42.5

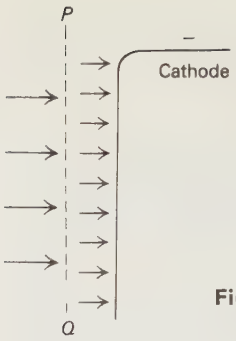


Fig. 42.6

and recombination went on. On the application of an electric field across the electrolyte, these solute ions would move and constitute an electric current. Eventually the ions would discharge at the electrodes.

Let us assume there are n positive and n negative ions per unit volume (Fig. 42.5) each carrying charge e. If these move with speeds u_+ and u_- respectively, the current through unit area of the plane XY will be

$$(nu_+e + nu_-e),$$

$$= ne(u_+ + u_-). \qquad (42.1)$$

Now, u_+ and u_- are not necessarily the same. So the part of the current carried by the cations (neu_+) will not necessarily be the same as the part of the current carried by the anions (neu_-).

At the cathode (Fig. 42.6), $ne(u_+ + u_-)$ units of charge are neutralized each second per unit area of the electrode. But, near to the cathode at the plane PQ, say, only neu_+ cations cross unit area in each second. There is, in consequence, a steady depletion of the amount of the cationic material in solution near the cathode.

We see mass m_1 of substance liberated at the cathode

$$\propto n(u_+ + u_-),$$

and mass m_2 of substance removed from the electrolyte nearby is

$$\propto n(u_+ + u_-) - nu_+$$

$$\propto nu_-.$$

So

$$\frac{m_1}{m_2} = \frac{u_-}{u_+ + u_-}. \qquad (42.2)$$

This ratio is referred to as the *transport number* for the cation and is a measure of the proportion of the current which it carries.

This analysis was first undertaken by J. W. Hittorf a very few years after Clausius had proposed his theory of electrolytic conduction. He found that concentration analyses designed to measure the cation transport number near to the anode were much less simple. We now realize that the phenomena occuring near to an electrode where ions are neutralized are extremely complex and consistent results using this simple analysis are obtained in only a few cases.

The measured current crossing unit area of the electrolyte is, as we have seen, $ne(u_- + u_+)$. At this stage, we can measure neither n nor e separately but we do know that 1 mole of solute ions carries 96 500 coulombs of charge. This quantity is known as the *Faraday constant* (F). So, if the concentration of the solute is N (measured in moles m^{-3}) and a fraction α of these are ionized at any one time

$$ne = \alpha NF,$$

and current

$$I = \alpha NF(u_+ + u_-). \qquad (42.3)$$

If the fraction α is known (and it can be determined by measurements in very dilute solution when it is legitimate to assume that the solute is

completely ionized) it is possible to combine the measurement of the transport numbers with the measure of the current flowing and so find the speeds of the ions. Such measurements show that the speed of the ions is very low. Typically, silver ions move at 5×10^{-8} m s^{-1} in a potential gradient of 1 volt per metre.

Problem 42.3 A silver voltameter has two silver electrodes 5 cm apart in a solution of silver nitrate. A potential difference of 10 volts is applied to the electrodes. How long does a single silver ion take to cross from anode to cathode?

In spite of the difficulties experienced by Hittorf and his successors in this very complex field, the basic features of the model suggested by Faraday and Clausius remain. A full analysis must take account of many factors ignored in our simple treatment. For example, what are the effects of varying the area of the electrodes, the temperature of the electrolyte, the applied potential difference, the spacing between the plates, the concentration of the electrolyte? You may wish to suggest other factors which might affect the electrolysis.

Faraday did in fact investigate such factors and convinced himself that these were secondary factors.

As N. Feather (1968) has remarked,

> "Ever since the time of Galileo the investigator of genius has known intuitively how to separate the essential from the contingent in the results of his own researches."

Subsequently, all such secondary effects have been incorporated into Clausius' model, but, at the time, they hindered acceptance of that model. Maxwell's authority strongly influenced opinion when he said in 1873

> "It is extremely improbable that when we come to understand the true nature of electrolysis we shall retain in any form the theory of molecular charges."

However, in 1881, Helmholtz was able to write

> "If we accept the hypothesis that elementary substances are composed of atoms, we cannot avoid the conclusion that electricity also, positive as well as negative, is divided into definite elementary portions which behave like atoms of electricity."

It will be recalled that we assumed that the quantity $96\,500/N_A$ coulombs (or F/N_A) appeared to be the basic quantity of charge carried by such ions as that of hydrogen. If we are to determine this quantity we need to know, by some independent means, the value of N_A, Avogadro's number. There are various ways in which this can be done. One method is based upon the diffraction of X-rays in a crystal lattice; another on the measurement of the rate of emission of alpha particles by a radioactive substance; and a third uses the kinetic theory of gases. This latter technique relies on an analysis first made by J. Perrin in 1912 and provided the first accurate determination of N_A, just 101 years after Avogadro first advanced his hypothesis. It is this last method which was described in Chapter 13. This led to a value for Avogadro's number of 6.04×10^{23} per mole.

Problem 42.4 Calculate the value of the electric charge carried by an individual hydrogen ion.

You should find that the charge carried by the hydrogen ion (which is in fact the smallest possible ionic charge) is 1.60×10^{-19} C.

42.3 SUMMARY

In the absence of direct experience of the movement of matter in metallic conductors carrying electric currents, the mechanism of current flow in an electrolyte was examined. In this case the flow was associated with a transference of matter. A model of charge transfer was developed which was compatible with other models concerned with the nature of matter, particularly the kinetic and the atomic theories.

Accepting these theories, it appears that the electric charge is intimately associated with the atoms of matter; that there is a *smallest unit* or *atom* of electricity and that its size is about 1.60×10^{-19} C.

Caution is necessary in the extension of this view to other cases but it is tempting to wonder whether the electric currents in metallic conductors represent the transport of atoms of electricity in some way. But we cannot postulate transport on ions for we have already seen (Chapter 16) that many of the properties of the solid state are described in terms of a rigid inter-locking structure of the atoms of the material.

Chapter 43
THE ELECTRON

43.1 The electronic hypothesis
43.2 Measuring the charge
43.3 Millikan's experiment
43.4 The conduction of electricity through gases at low pressure
43.5 The nature of the cathode rays
43.6 Some applications of thermionic emission

43.1 THE ELECTRONIC HYPOTHESIS

The hypothesis of discrete charges which arose from the work on electrolysis was crystallized by G. Johnstone Stoney in an address to the British Association in 1874. Stoney spoke, as we have done, of a natural unit of electric charge associated with such ions as those of hydrogen, chlorine, etc. (monovalent ions). Using such estimates of Avogadro's number as were available to him, Stoney estimated the size of this unit at about one tenth of the value accepted today. In 1891 he went further and named this unit of electric charge the *electron*.

We must recognize that Stoney used the word "electron" in a different sense from the way we use it today. For him it described the ultimate unit of electric charge. He assumed that all electrical charges were whole-number multiples of this basic unit, whether they were charges on ions or the gross charge of a rubbed ebonite rod. The testing of such an hypothesis is of fundamental importance.

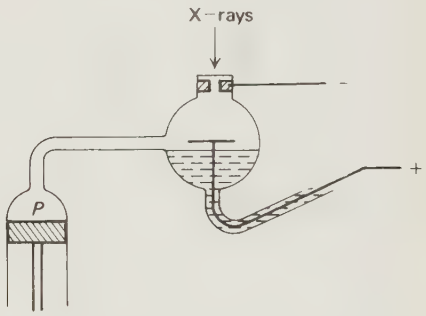


Fig. 43.1

Before we consider further evidence for this hypothesis, we ought first to make clear the difference between the testing of this hypothesis and "discovering the electron." In this latter sense, we use the word "electron" differently from the way Stoney used it. Subsequent to Stoney's invention of the term, it became clear that the movement of electric charge was associated with a specific negatively charged speck of matter. The evidence for this arose as a result of experiments carried out on the conduction of electricity through gases at low pressure. These experiments culminated in the work of J. J. Thomson.

Since this time Stoney's word for the ultimate unit of charge (electron) has been used to describe specifically the negatively charged particle which Thomson found to be associated with the passage of electricity. However, Thomson's discovery and measurement of the specific charge (e/m) of the electron did not demand that there must be a fixed elemental unit. This point is taken up again in the next section when we shall consider in more detail the evidence for the electron in the modern sense of the term.

43.2 MEASURING THE CHARGE

Many attempts were made to measure directly the small charges associated with atoms of matter. In 1897, J. S. Townsend had estimated the charge carried by the oxygen atom in electrolysis by a "falling cloud" of droplets technique which was to be so fruitful in the hands, first of H. A. Wilson and subsequently of R. A. Millikan. Thomson also used this technique in collaboration with Wilson. Figure 43.1 represents the apparatus devised by Thomson and later modified by H. A. Wilson.

X-rays were used to ionize the saturated air contained in the chamber. A sudden expansion was produced by pulling down the piston P and a cloud of water droplets condensed on the ions formed by the irradiation of the air by the X-rays. That the water droplets condensed preferentially on the ions had been shown first by Townsend (a technique subsequently developed by C. T. R. Wilson, the inventor of the expansion cloud chamber). By suitably adjusting the expansion ratio, Wilson ensured that the droplets only condensed on negatively charged ions. He then observed the rate of fall of the top of the cloud between the two horizontal brass plates, first in the absence of an electric field and then with a 2000 V battery connected across them. By observing the fall of the top of the cloud, he ensured that his measurements were made on the drops which carried the least charge.

For a water droplet which is falling freely at a constant terminal velocity the forces acting are in balance and, from Stokes' law for viscous friction

$$mg = 6\pi\eta a v_1, \tag{43.1}$$

where m is the mass of the drop, η is the viscosity of the air, a is the radius of the drop and v_1 is its terminal velocity.

If now the electric field E is applied the force (Eq. 43.1) becomes

$$mg + Eq = 6\pi\eta a v_2,$$

where q is the charge on the drop and v_2 the new terminal velocity.

Hence

$$\frac{mg}{mg + Eq} = \frac{v_1}{v_2}. \tag{43.2}$$

From Eq. (43.1)

$$a = \frac{mg}{6\pi\eta v_1}. \tag{43.3}$$

But

$$m = \tfrac{4}{3}\pi a^3 \rho g, \tag{43.4}$$

where ρ is the density of the water. Combining Eqs. (43.3) and (43.4) we can eliminate the radius a and hence find a value for m the mass of a drop. Substitution in Eq. (43.2) will then provide a value for the charge q. It was thus possible for Wilson to determine a value for the charge q. The method is clearly the precursor of Millikan's work although it contains a number of assumptions, for example, that each water droplet has the same radius and therefore mass as its neighbour in the layer at the top of the cloud. Precision was low but the experiments did lead to values of about 1.0×10^{-19} C.

Other methods led to similar results. For example, Rutherford and Geiger in 1908 gave a value of 1.55×10^{-19} C on the basis of the measurement of the charge carried by an alpha particle.

Nevertheless, it was R. A. Millikan who produced definitive confirmation of the existence of a basic unit of electric charge even though most physicists working in the field in 1909, when he started work, would have said that the case was proven. This provides a prime example of the way in which theory and experiment, at times, go hand in hand. In this case, the evidence which most strongly supports the theory was provided shortly after the theory was itself well established.

Table 43.1

Package number	Mass (in grammes)
1	25.3
2	17.5
3	6.0
4	13.0
5	12.7
6	31.0
7	105
8	42.2
9	18.6
10	12.9

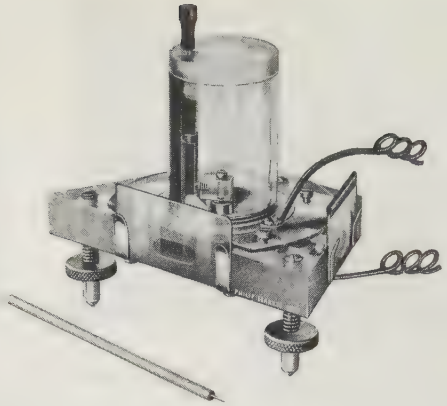


Fig. 43.2 A cell designed for use in the Millikan experiment. (Courtesy of Philip Harris, Ltd.)

43.3 MILLIKAN'S EXPERIMENT

The essence of Millikan's experiment was to charge small droplets of oil and to calculate the charge they carried by measuring their time of fall under the action of gravity alone and under the combined effect of an electric and a gravitational field. The oil droplets were produced by a spray. Many of them were charged by friction, but Millikan used independent means of both charging and changing the charge. X-rays, whose ionizing properties were well known, were used for this.

The importance of Millikan's experiment lies not so much in the fact that he was able to measure the charges carried by the oil drops but that the charges so measured were very small multiples of an assumed fundamental unit of charge. It is the deductions that can be drawn from the measured sizes of these changes, rather than the nature of the measurements themselves which are important to us.

As an analogy, let us suppose that we are presented with the masses of a number of packages. We are told that, as a working hypothesis, we may assume that each contains a different number of otherwise identical marbles. Table 43.1 gives a list of such masses (due allowance has been made for the packaging). If the hypothesis is correct then each mass should be an integral multiple of the mass of a single marble. By inspecting the table of masses we may be able to select the smallest multiple. Our task is aided if one of the packages contains only one marble.

Look at the table and try to select a lowest multiple. Then work out how many multiples of this basic unit are contained in each package.

When you have decided how many basic units (marbles) there are in each package, divide the mass of each package by the number of marbles you think it contains to find the average mass of each marble in the package.

The success of your hypothesis must be judged against the degree of consistency in your answers and the expected experimental error in the measurements.

How many marbles did you find in package 7? There were actually 17. You probably found it very difficult to decide whether there were 16, 17, or 18. Would it have made any significant difference to the average mass of each marble in this particular package, whichever figure you had chosen? With large numbers of marbles in a package it is very difficult to decide how many marbles are contained within it.

The importance of Millikan's experiment is such that there are available filmed records of modified versions of it. In one of these (*Are there electrons? a*

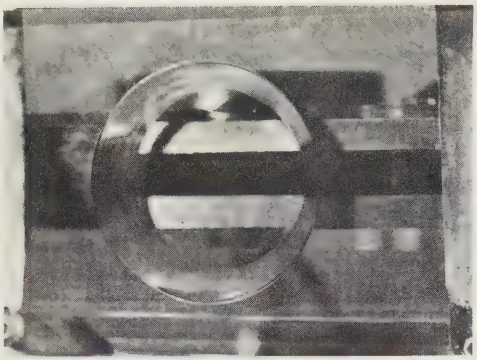


Fig. 43.3 Close-up of the plates of the Millikan cell. (Courtesy of Philip Harris, Ltd.)

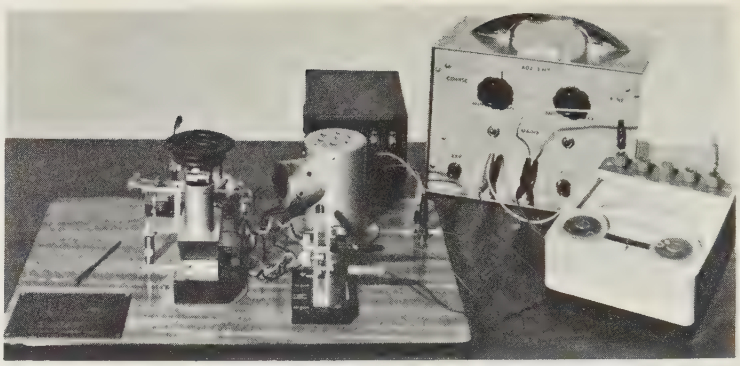


Fig. 43.4 Experimental arrangement of apparatus for the Millikan experiment. (Courtesy of Philip Harris, Ltd.)

Millikan experiment, Rank-Mullard), the oil drops are replaced by minute but extremely uniform latex spheres each of which weighs 4.4×10^{-14} N. Such spheres are suspended in water so that it is possible to flick them from a loaded brush into the Millikan cell in which the experiment is performed (Fig. 43.2).

Some of the spheres will be charged by friction in this process. Many will fall through the small hole in the upper plate into the space between the two parallel plates which make up the cell proper (Fig. 43.3). The charged spheres may be prevented from falling to the bottom of the cell by applying suitable charges to the two plates (Fig. 43.4). It is possible, by suitable manipulation of these charges, to select an individual charged sphere for careful study. Such a sphere may have a charge q and mass m.

If a uniform field E gives an upward force on the charge which just balances it against the pull of gravity, then

$$Eq = mg. \tag{43.5}$$

The field E is governed by the potential difference V across the plate and by the plate separation, d.

$$E = \frac{V}{d} \tag{43.6}$$

$$\therefore \frac{V}{d} q = mg.$$

So

$$q = \frac{mgd}{V}. \tag{43.7}$$

The potential difference across the plates is recorded when the charged sphere has been carefully balanced so that it moves neither up nor down. Next, the potential difference is switched off and the air in the space between the plates is ionized by exposing it to the radiation from a small radioactive source for a moment. Before the sphere falls to the bottom the potential difference is switched on again and the weight of the sphere is re-balanced. If the charge has been changed, this will require a new value of potential difference. Since m, g, and d are constant

$$q \propto \frac{1}{V}.$$

Readings of potential difference recorded during the filmed experiment are given in Table 43.2 and reproduced by permission of Rank Film Library.

Table 43.2

327	306
425	236
840	119
564	178
1690	59.3

Do these figures give evidence for the view that electric charge comes in integral multiples of some basic unit?

The film asks the question "Are there electrons?" You should consider the answer to this question very carefully in the light of this evidence and our previous discussion.

The observations made in the film suggest a value for the basic unit of charge of 1.7×10^{-19} C.

This compares well with the value of F/N_0 which we obtained on the basis of Perrin's value for N_0.

Subsequent experiment has only served to confirm the quantum nature of electric charge. Millikan's experiment has, over the years, become more and more a demonstration of this quantization, and less and less a method of measuring the value of the charge on the electron.

In 1925, a new method of determining the value of N_A arose from the direct measurement of X-ray wavelengths using a reflection grating by A. H. Compton. The Bragg diffraction angles of certain simple crystals can then be used to determine the spacings of the crystal planes, and these spacings in turn related to the molecular weight and density through Avogadro's number.

Early values of F/N_A, using these accurately determined values of N_A, did not agree entirely with Millikan's value for e. These discrepancies were only finally resolved when it was shown that, for bodies as small as the oil drops falling through air, the effective viscosity is different from that which was assumed by Millikan. This had led to an error in his determination of the sizes of the drops.

Recent (1969) measurements of the fundamental unit of charge give the mean value as

$$(1.6021917 \pm 0.0000070) \times 10^{-19} \text{ coulomb},$$

but to us, its existence is of more importance than an accurate knowledge of its size.

43.4 THE CONDUCTION OF ELECTRICITY THROUGH GASES AT LOW PRESSURE

Gases at atmospheric pressure are non-conductors of electricity unless the electric field to which they are subject is very high indeed. In air the field must exceed 3×10^6 V m^{-1} before a current, in the form of a violent and noisy spark, passes between the two electrodes providing the field.

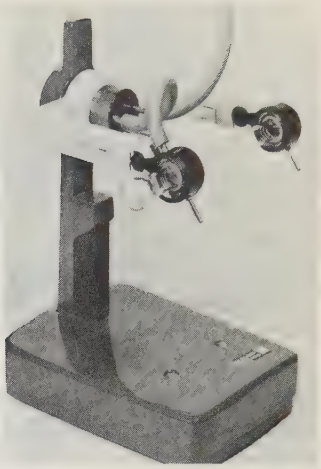


Fig. 43.5 A discharge tube. (Courtesy of Teltron Ltd.)

The lightning flash is the best example of such a discharge. Experiments with a rotating camera suggest that the structure of the flash is complex. A leading down stroke, which moves quite jerkily and slowly, ionizes the air leaving a conducting path in its wake. This path then carries the main discharge upwards at very high speed (about 40 000 km s^{-1}) in a current of from 10^4 to 10^5 A. Since the duration of this intense flash is about 10^{-4} s the quantity of electricity involved is in the order of 10 C. The potential difference between cloud and cloud or between cloud and ground may be as high as 10^8 V, so that 10^9 J may be involved in a typical discharge (Dobson, 1968).

At low pressures, however, a quieter steadier passage of electricity can take place between the two electrodes. In this case also the current is accompanied by the emission of light.

The electrodes in the tube shown in Fig. 43.5 are 15 cm apart. A potential of 2000 V between them is insufficient to cause a current to flow at atmospheric pressure.

As the pressure is reduced, a point is reached when the tube is suddenly filled with a pinkish glow. A microammeter in the circuit records a current.

As the pressure is further reduced, the pink glow retreats towards the anode, breaking up into a series of striations, leaving a dark space between the cathode and the beginning of the striations. This is the Faraday dark space. With continuing reduction in pressure, this space extends and is seen to include a further glow region close to the cathode which is itself separated from the cathode by an additional dark space. The existence and properties of this dark space were announced by Sir William Crookes in 1879. Without our modern facilities for producing high vacuua, the extent (in space) of the Crooke's dark space, as it has come to be called, was probably very small and had certainly been overlooked by a number of experimenters. While today we interpret the phenomena occurring in the discharge tube in terms of a particulate radiation emanating from the negative electrode, two independent theories were advanced to interpret the observations. It is worth looking briefly at this development in order to stress once more how difficult it is to produce unequivocal experimental results which will falsify one theory rather than another.

In British circles during the latter part of the nineteenth century a general belief grew up that the Crooke's dark space represented the mean free path of a stream of charged particles emanating from the cathode. The luminosity at the anode end of the dark space could be accounted for by the ionization of gas molecules in impact with this fast moving stream. Such a model fitted in well with observations made as early as 1858. Then Julius Plücker had seen how this region of discharge could be influenced by a magnet and that in highly evacuated tubes there was some phosphorescence of the tube wall near the cathode. This too could be influenced by a magnet.

The passage of electricity through low pressure gases excited the curiosity of almost every experimental physicist in the latter part of the nineteenth century. The theory of fast moving particles emanating from the cathode was not the only model put forward to explain the observations. The experimental work of Hertz and Lenard in Germany was interpreted in terms of what we would now call an electromagnetic radiation issuing from the cathode. British physicists would have none of this, believing that Plucker's observations could only be understood in terms of a moving charged particle. They believed the particles to be ions arising either from the cathode or from the gas in its neighbourhood. Hertz, a strong advocate of the electromagnetic wave theory, felt his model to be confirmed when he failed to deflect the radiation near the cathode with electrically charged plates placed outside the discharge tube. It was some years before J. J. Thomson was able to show that Hertz's null result was due to the neutralization of the field by the redistribution of gas ions within the tube. He was able to place the plates inside the discharge tube and to show that, at sufficiently low pressures, when the gas ions were considerably reduced in number, electrostatic deflection of the radiation could, in fact be produced (see Section 47.6).

Although, as we shall see, the particle model proved the only one capable of interpreting all the phenomena as we now know them, the alternative view may well have produced a secondary, very important success. Röntgen, in 1895, published a now famous paper on his discovery of some very penetrating radiation produced at the anode of a discharge tube on bombardment by the cathode radiation. He interpreted this new radiation as an electromagnetic radiation which he termed, for want of a better term, *X-radiation*. His original theory has never been refuted, but it is interesting to speculate that it may well have been coloured by the then current German theories of the cathode radiation. There we will leave for the moment the historical account of the discovery of cathode rays and return to our more modern experiment.

With a good vacuum pump, the pressure of the air in the tube in Fig. 43.5 can be reduced still further, until the Crooke's dark space fills the whole of the tube. A current is still registered on the meter showing that electricity still passes down the tube. A small slit in the anode allows some of the electrically charged cathode radiation through on to the chemically treated patch at the end. This glows greenly under the impact of the radiation. The glass walls of the tube emit a bluish glow and both phosphorescences are influenced by bringing a magnet up to the tube.

A small slit in the cathode allows radiation proceeding from the direction of the anode to fall on a phosphorescent screen. Here again a green patch of phosphorescence is produced. A magnet has little effect on this patch of light, however, and this alone shows it to be different in some way from the cathode radiation. We shall return to this anode radiation later.

43.5 THE NATURE OF THE CATHODE RAYS

The nature of the cathode radiation was investigated experimentally in a classic sequence of experiments by Thomson, J. B. Perrin and others. Today, we believe the tubes they used to be highly dangerous due to the emission of X-radiation resulting from the impact of high energy cathode rays on anode and tube walls. Because of this, we repeat their demonstrations using radiation produced by the thermionic effect. (It should be noted that the small discharge tube used in our initial demonstration is not dangerous as the accelerating potentials are less than 5 kV. However, the dimensions of the tube are too small to allow detailed investigation of the properties of the cathode rays. But bigger tubes need bigger accelerating potentials and this is when they may become dangerous.)

The thermionic emission of cathode rays was discovered by Elster and Geitel in the 1880s. At that time it could only be said that they found that a current would pass from a red-hot wire to another electrode in a low pressure gas, if the other electrode was positive with respect to the hot wire. Thomson was able to show that the electric current was carried by radiation with the same characteristics as the cathode rays of a discharge tube. This thermionic emission of radiation is particularly associated with tungsten.

In the tube shown in Fig. 43.6, thermionic radiation passes through a hole in the anode to form a widely divergent beam. A metallic object (conventionally in the shape of a Maltese cross) is shown to form a shadow on the phosphorescent screen which covers the end of the tube.

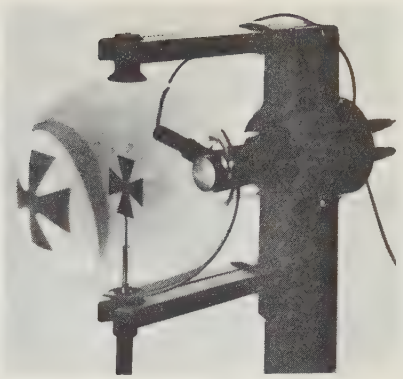


Fig. 43.6 A hot-cathode "Maltese cross" tube showing the shadow cast by the beam of electrons. (Courtesy of Teltron Ltd.)

The beam can be deflected by a magnet and shown to be consistent with a stream of negative charge. The light from the bright filament also casts a shadow but this is, of course, undisturbed by the magnetic field.

It is useful to return here to our original model of an electric current. As we noticed at the beginning of this unit, we know too little at this stage about the nature of electric current to build a satisfactory model. The evidence of the magnetic deflection of a current carrying wire shows that there is a direction of flow to be associated with the current.

If the wire moves upwards out of a magnetic field, then reversal of the battery will cause it to move downwards. Similar effects are produced by reversing the magnet. (See Fig. 33.12.) To produce a rule of thumb to describe these observations we need to describe the direction of the magnetic field and the direction of the electric current. The former is conventionally referred to as being from north pole to south. But what of the current? There is no means of telling whether positive charge flows from left to right in the diagram or negative charge from right to left. So we adopt a convention in the absence of any other evidence and say it flows from the positive terminal of the battery to the negative. In this way, Fleming arrived at his celebrated *left hand rule*.

In applying this rule to the deflection of the thermionic radiation, we find that the radiation may

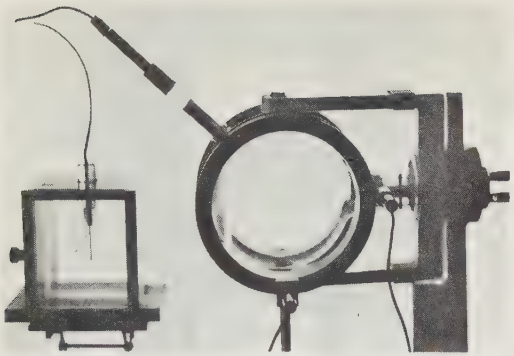


Fig. 43.7 A hot-cathode "Perrin" tube. The electron beam is deflected into the collecting cylinder by the magnetic field of the surrounding coils. (Courtesy of Teltron Ltd.)

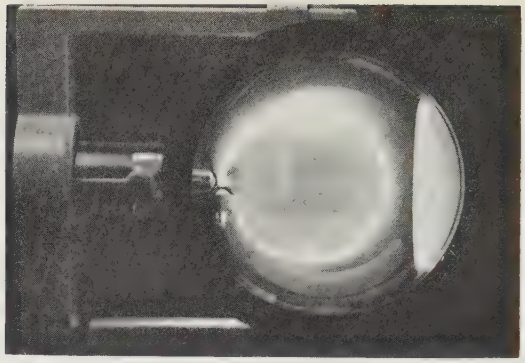


Fig. 43.8 A "fine-beam" tube. The electron beam, made visible by the re-combination of the ions it has produced, follows a circular path in the magnetic field produced by the current in the coils. (Courtesy of Teltron Ltd.)

be regarded as a conventional current flowing from the phosphorescent screen towards the anode. But the electric charges which originate in the filament must presumably be moving in the opposite direction. Thus we deduce that they are negative charges.

A more direct confirmation of this fact can be obtained using the tube shown in Fig. 43.7. This is an adaptation of the tube developed by J. B. Perrin. The thermionic radiation is deflected by a magnet into the small metal can in the arm at the top right of the diagram, which is in turn connected to a gold-leaf electroscope. The leaf is deflected by the collected charge and the electroscope is isolated from the collecting can. A test of the charge carried by the leaf shows it to be negative.

That these negative charges are carried on particles of matter seems a reasonable extension of the observations of electrolytic conduction.

Another thermionic tube (see Fig. 43.8) gives us even more information about the nature of these particles.

In this tube, the radiation emerges from a gun-like device as a thin beam whose passage through the space is made visible by the ionization of low-pressure gas in the tube. The evidence of the other experiments suggests that the beam is made up of a stream of moving electrically charged particles. Let us assume that each particle has mass m and a negative charge q. Their motion results from the application of an electric field within the gun so that the kinetic energy acquired in the field is equal to the potential energy provided. If the potential difference applied is V, the potential energy provided is qV and the kinetic energy acquired is $\frac{1}{2}mv^2$ where v is the velocity with which the particles leave the gun. So

$$qV = \tfrac{1}{2}mv^2. \qquad (43.8)$$

The resultant beam of moving charges is deflected by a magnetic field B produced by the pair of coils alongside the tube. Since the field B is at right angles to the direction of motion of the particles, the beam is deflected into a circle of radius r. The deflecting force (F) bends the path of particles of mass m and speed v into a circle of radius r.

$$F = \frac{mv^2}{r}. \qquad (43.9)$$

This force is provided by the interaction of the magnetic field and the moving charges and is given by:

$$F = Bqv. \qquad (43.10)$$

Combining Eqs. (43.9) and (43.10) we get

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

$$Bq = \frac{mv}{r}. \tag{43.11}$$

It follows from Eq. (43.8) that

$$v = \sqrt{\frac{2qV}{m}}. \tag{43.12}$$

Substituting in Eq. (43.11) we get

$$Bq = \frac{m}{r}\sqrt{\frac{2qV}{m}}$$

and simplifying we obtain

$$\frac{q}{m} = \frac{2V}{B^2r^2}. \tag{43.13}$$

Since V, B, and r can all be measured, we can calculate a value for q/m.

Problem 43.1 In a typical experiment, B was found to be 6×10^{-4} tesla, and V was 100 V. The consequent value of r was 5.7 cm. Calculate the value of q/m.

The experiment gives us the ratio of the charge carried by a particle to its mass. This is to be expected in experiments involving the deflection of a beam resulting from the application of a certain force. The deflection produced will depend on the acceleration of the charges. By Newton's second law

$$a = \frac{F}{m}.$$

Since the size of F depends on the size of q, the value of a must depend on the ratio of q/m and not on either q or m separately.

It follows from this that while the experiment confirms the existence of mass associated with negative charge of the radiation, it cannot tell us that all the particles are identical and carry the same charge. It only tells us that the ratio q/m is the same for each particle. The density of lead is similarly the same for all specimens of lead. The density is the ratio of mass of sample/volume of sample. We would be wrong to infer that all samples of lead are the same size!

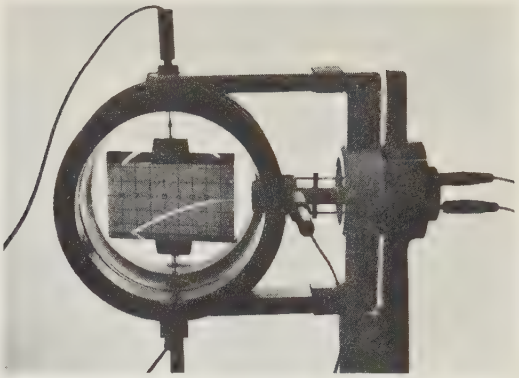


Fig. 43.9 A modern version of the "Thomson" tube. The electron beam is seen deflected in the electric field between the two plates. (Courtesy of Teltron Ltd.)

Some early experimenters expressed doubts about the assumption implicit in our equating the kinetic energy of the charged particles with their loss of electrical energy and looked for a way of measuring more directly the speed of the particles. Thomson produced such an experiment. Figure 43.9 shows a modern (thermionic) version.

In this tube, the magnetic force acting on the thermionic radiation can be counterbalanced, in the centre of the field at least, by an electric force. Under the conditions of balance

$$Eq = Bqv,$$

where E is the electric field produced between the parallel plates. So

$$E = Bv$$

and

$$v = \frac{E}{B}. \tag{43.14}$$

Subsequently, the deflection produced by the magnetic field acting alone can be measured to give the equation

$$Bq = \frac{mv}{r} \qquad \text{from Eq. (43.7).}$$

Then

$$Bq = \frac{m}{r}\frac{E}{B}.$$

So

$$\frac{q}{m} = \frac{E}{B^2 r}. \tag{43.15}$$

E may be measured from the known separation of the plates and the measured potential difference between them.

J. J. Thomson experimented extensively with such radiation and showed that the value of q/m was the same (within the limits of a large experimental error) whether the radiation was produced by cold-cathode discharge, the thermionic effect or by the irradiation of the cathode with ultraviolet light (another discovery of Elster and Geitel). Nor did the ratio depend on the nature of the gas in the tube. This excluded any suspicion that the radiation could be explained in terms similar to those used to explain electrolysis.

We have stressed that the constancy of the ratio q/m for whatever particles constituted the cathode radiation (however derived) implies neither constant mass nor constant charge for the particle. Nevertheless, the work in electrolysis had made it reasonable to suppose that a fundamental unit of charge did exist and that all other charges were integral multiples of this. If q/m is to be constant and q may only change by integral steps, m must also vary by integral steps of mass. It would not be unreasonable to find particles whose charge to mass ratio was q/m, $2q/2m$, $3q/3m$, etc. It is, however, strange that ratios such as $2q/m$ or $q/2m$ did not appear in cathode radiation experiments. Thomson was forced to make the bold hypothesis that, in fact, the particles were alike, all having the same mass and all carrying the basic unit of charge suggested in the electrolysis work.

On this hypothesis the new particle, to which Stoney's word "electron" was soon applied, carries charge of the same magnitude but opposite sign to the charge on the hydrogen ion.

Since hydrogen has the lowest atomic mass of all elements, the ratio of its charge to mass—the so-called specific charge of the hydrogen ion—will have the largest value to be found for any ion. On the assumption that F coulombs liberate 1.08 g of hydrogen, then the specific charge on the hydrogen ion e/m_p is given by

$$\frac{e}{m_p} = \frac{F}{1.08 \times 10^{-3}}\ \text{C kg}^{-1}$$
$$= 8.94 \times 10^7\ \text{C kg}^{-1}.$$

The specific charge on an electron, e/m_e, turns out to be 1.76×10^{11} C kg^{-1} using the best of today's values.

From these two ratios we can calculate the ratio of the mass of an electron to the mass of a hydrogen ion:

$$\frac{m_e}{m_p} = \frac{8.94 \times 10^7}{1.76 \times 10^{11}} \approx \frac{1}{1970}.$$

Thus the hydrogen ion, hitherto the least massive of all particles, is almost 2000 times more massive than the electron.

The way in which Thomson was able to incorporate this new particle into an evolving structure of the atom is told in Unit 9. For our part, its existence as an element of the structure of matter allows us to turn again to the question of the nature of the electric current in solids.

43.6 SOME APPLICATIONS OF THERMIONIC EMISSION

The discovery of thermionic emission led, several decades later, to the development of a major industry—the electronics industry. And although many of the thermionic vacuum tubes which were produced have now been replaced by solid state devices there remains a significant group of tubes in everyday use. This group includes the television tube in the home and the cathode ray tube in the workshop and laboratory.

Freeing the electron provided a charged particle of minute mass (and therefore inertia) which could

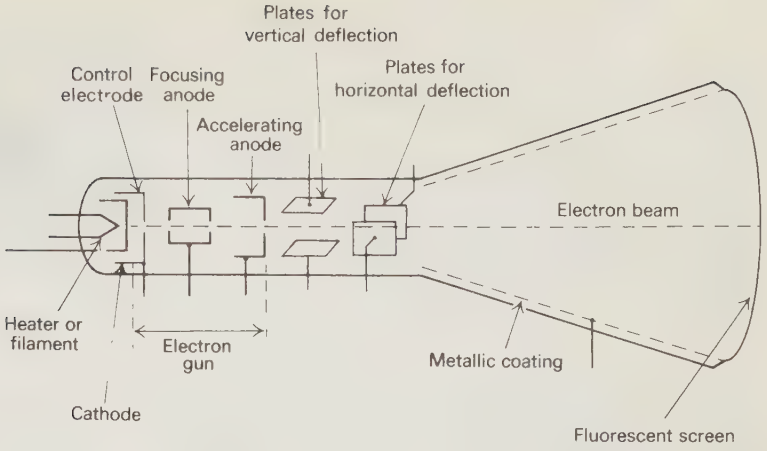


Fig. 43.10 The basic elements of a cathode-ray tube. (From F. W. Sears and M. W. Zemansky, *University Physics*, Addison-Wesley, 1970.)

respond very rapidly indeed to changes in electric and magnetic fields. Moreover, these particles could be given sufficient energy to cause fluorescence to appear when they struck a suitable screen. It was this behaviour which is put to use in the *cathode ray tube* (Fig. 43.10).

Electrons from the filament and cathode complex are accelerated in the electric field towards an anode which is provided with a central hole. In between the cathode and the anode, one or more additional electrodes, each at a suitable potential, serve to concentrate the beam and to overcome the tendency the beam has to diverge as a result of the mutual repulsion of the electrons. This arrangement (which constitutes, with the anode and the cathode, an "electron gun") also produces a converging beam which then travels unimpeded through the tube to strike the fluorescent screen to give a small, bright spot of light.

In the tube shown (Fig. 43.10) the beam passes between two pairs of flat parallel plates which are located between the anode and the screen. A potential difference between the first pair (Y plates) will cause the beam to be deflected in the vertical plane; a potential difference between the second pair (X plates) will cause deflection in the horizontal plane. Using the two pairs together, it is possible to direct the beam to fall on any point on the screen.

It is usual to apply a potential difference which varies with time (in the way shown in Fig. 43.11) to the X-plates. Then the spot moves across the screen in a horizontal direction at a steady speed, for example, 1 ms per cm. When the spot reaches the end of the screen, it is brought back to the starting point very quickly and repeats its travel. If now an alternating potential difference is applied to the Y plates, the spot will trace out a path which displays the wave form of the alternations. If the time base speed is 1 ms per cm, say, and the frequency of the applied alternating potential difference is 5 kHz (i.e., 1 cycle in 0.2 ms) 5 cycles will appear for each cm of trace.

The uses of such a tube which, with the appropriate circuitry, is the heart of the *cathode ray oscilloscope*, include the display of wave forms, the measurement of voltages (both peak-to-peak and direct) direction and range-finding (radar), navigational aids, echo-sounding; indeed almost any phenomenon can be investigated provided that it can be expressed in terms of a change in electrical potential. But, of course, it is the television tube which affords the commonest example of the application of thermionic electrons.

The spot on the tube face is clear evidence that the electrons which make up the beam are energetic. As we shall see in Unit 9, rapid deceleration of the beam of electrons in a suitable target can lead to the



Fig. 43.11 A saw-tooth voltage for a time-base.

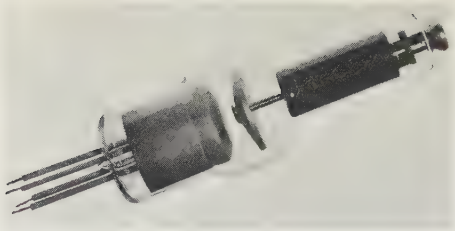


Fig. 43.12 X-ray tube showing the anode construction. (Courtesy of Mullard, Ltd.)

production of extremely short electromagnetic radiation, the X-radiation. With a wavelength in the order of 10^{-10} m this radiation has high penetration and also affects photographic film. It has important effects on some living tissue.

The modern *X-ray tube* is a high vacuum tube containing a thermionic cathode and an anode (Fig. 43.12). The anode presents an inclined surface to the electron beam. The electrons are accelerated by a potential difference in the order of 100 or more kilovolts, so acquiring very considerable energies. On striking the anode much of this energy will cause the anode temperature to rise but a proportion will appear as X-radiation. Evidently then, a major problem in the design of such tubes is that caused by the need to cool the anode. This is usually a massive piece of metal and a good conductor. Cooling water may circulate or, in some designs, the anode is rotated constantly so that a different part of its surface is continuously presented as the target for the beam.

The uses of X-rays are not confined to medical science but extend to the examination of large metal castings and forgings, to the inspection of canned products, the examination of baggage and parcels at, for example, air-ports and, as we have already seen in Unit 4 to the examination of crystal structures.

Chapter 44
A MODEL FOR AN ELECTRIC CURRENT

44.1 The Drude theory of electric conduction in metals
44.2 Thermal conductivity
44.3 The force on a current-carrying wire in a magnetic field
44.4 The Hall effect

44.1 THE DRUDE THEORY OF ELECTRIC CONDUCTION IN METALS

Before we attempt to describe the simple model of electric conduction first put forward by P. K. L. Drude at the turn of the century, we ought first to see what features our model ought to incorporate. Experimentally, they are very few. As we have seen in Unit 6, G.S. Ohm established experimentally in 1827 that the electric current passing through a metal at a fixed temperature was proportional to the potential difference across it (Ohm's law).

Secondly, it has always appeared that good conductors of electricity are good conductors of heat. That this is not a merely superficial observation was made clear by the work of G. H. Wiedemann and R. Franz in 1853 who showed that, at a fixed temperature, the ratio of the thermal and electrical conductivities is the same for all metals.

Experiments show that the rate of conduction of heat (dQ/dt) through a surface of area A in a material is given by

$$\frac{dQ}{dt} = -\lambda A \frac{dT}{dl}, \tag{44.1}$$

where dT/dl is the temperature gradient normal to the surface of area A. The negative sign shows that the energy is being transferred from the higher to the lower temperature. This equation defines the conductivity (λ) of the material under the conditions obtaining. We have seen already (Unit 5) that the rate of conduction of electricity (i.e., the current) through a surface of area A in a metal is given by

$$I = \frac{V}{R} = \frac{A}{\rho L} V,$$

where ρ is the resistivity.

This may be written

$$\frac{dQ}{dt} = \frac{1}{\rho} A \frac{dV}{dl} \tag{44.2}$$

and the similarity with the thermal case is apparent. The electrical conductivity σ is then the reciprocal of the resistivity.

With these observable facts behind us, we will turn to the model proposed by P. Drude. He, in

company with J. J. Thomson and H. A. Lorentz, assumed that some of the electrons which were believed to be the essential constituents of all matter, were somehow freed from their parent atoms in the metallic state, the metal being imagined as a lattice of positive ions in which a "gas" of electrons was confined. It was this "free electron gas" that was to be the seat of the observed electron current when an electric field was applied. We will explore the consequences of this model.

Professor J. M. Ziman in a paper which describes an approach to Drude's theory makes the following comment on the model:

> "It is typical of modern physicists that they will erect skyscrapers of theory upon the slender foundations of outrageously simplified models. We have an array of ions, each with charge enough to bind an electron if left to itself, and an equal number of electrons which seem to be able to move around a bit amongst the ions. So we take the bit between our teeth and assume that the ions are not there at all! Or, rather, we smear them out into a fixed uniform background of positive charge—a sort of jelly—in which the electrons move quite freely. We need this jelly to make our system electrically neutral, so that the electrons will not be driven explosively apart by their coulomb repulsion." [*Ziman, 1963. Reprinted from* Electrons in Metals, *Taylor and Francis, 1963.*]

What then will be the electrical conductivity of such a model. Infinite, surely? Unless we can see some way in which the drift of electrons under an applied field is obstructed in some way. By collision perhaps. Collision with what? The ionic lattice? But we have smeared this out. Nevertheless, temperature effects, impurities, discontinuities in the lattice will all cause the background "jelly" to have irregularities in density. Maybe the electrons will collide with such irregularities. If we have a gas of electrons, we must imagine the electrons within it to be in random motion like the molecules of a real gas. These randomly moving electrons will collide with the irregularities within the lattice.

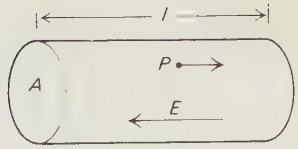


Fig. 44.1

The average time elapsing between successive collisions of a particular electron with an irregularity in the lattice will be characteristic of a given sample of material under specified conditions. We shall follow convention by calling this average time 2τ, where τ itself is the *relaxation time*. This relaxation time will not only differ from material to material but it will also be a function of temperature as well as of other possible variables.

Suppose that an electron P moves under the influence of the uniform field E (Fig. 44.1). It will accelerate until it interacts with some irregularity in the background. This collision may cause it to stop, to move with the field again, to move against the field or to move across the field. The latter case need not concern us except in so far as there exists a component of velocity in the direction of the field or against the field. On average, the electron moves with the field as often as it moves against it. On average, we may regard the electron as brought to rest. And so, immediately after the collision, it will be accelerated by the field yet again.

We can assume that the average time between collisions will still be 2τ even though the electrons acquire additional velocity under the action of the superimposed field. This *drift speed* turns out to be very small when compared with the speeds of thermal motion and it does not affect the average time between collisons of electrons and irregularities significantly. Then the speed v acquired by an electron of charge e and mass m_e after a time 2τ is given by

$$Ee = \frac{m_e(v - 0)}{2\tau} = \frac{m_e v}{2\tau}, \tag{44.3}$$

since, from Newton's second law, the impressed force is equal to the rate of change of momentum and so

$$v = \frac{2Ee\tau}{m_e}. \tag{44.4}$$

Consequently, the average drift speed $\bar{u}$ of the electron under the field E is

$$\bar{u} = \frac{0 + v}{2} = \frac{Ee\tau}{m_e}. \tag{44.5}$$

Let us suppose that our specimen is of length l and of cross-sectional area A^2. If a p.d. V across the specimen gives rise to a current I, we can write

$$E = \frac{V}{l} \tag{44.6}$$

If there are n electrons per unit volume, then

$$I = (\bar{u}A)ne \tag{44.7}$$

since all electrons up to a distance of the average drift velocity, $\bar{u}$, from a surface of area A will cross that surface during the next second. Hence

$$\bar{u} = \frac{I}{Ane}.$$

Substituting these values for $\bar{u}$ and E in Eq. (44.7) we have

$$\frac{I}{Ane} = \frac{V}{l} \cdot \frac{e\tau}{m_e}$$

$$V = \frac{lm_e}{Ane^2\tau} I. \tag{44.8}$$

So Ohm's law applies to the model, and the resistance of the specimen is proportional to the length and inversely proportional to the cross-sectional area just as Ohm found experimentally. Moreover, the resistivity

$$\rho = \frac{m_e}{ne^2\tau} \tag{44.9}$$

Problem 44.1 Estimating the relaxation time τ for copper. While it is not possible to make an independent assessment of τ, it is instructive to calculate an approximate value in terms of our present model. Throughout the succeeding work we shall use copper as an example. The following data will be required:

Density of copper 9×10^3 kg m^{-3}.
Atomic weight 63.
Avogadro's number 6×10^{23} atoms mole^{-1}.
Resistivity 1.6×10^{-8} ohm m
Thermal conductivity 3.9×10^2 W m^{-1} K^{-1}
} Both at 273 K.

a) Since there will be 6×10^{23} atoms of copper per 63 grammes, calculate the number of copper atoms per cubic metre.
b) Assuming that each copper atom provides 2 free electrons (the ion being Cu^{2+}) calculate the electron density.
c) On the assumption that the specific charge of an electron $e/m_e = 1.76 \times 10^{11}$ C kg^{-1} and that $e = 1.6 \times 10^{-19}$ C, calculate τ from the formula derived above.

You will have found a value of about 1.2×10^{-14} s for the relaxation time. What would be the effect on the relaxation time of a rise in temperature of the solid? You should consider first the effect of the temperature rise on the number of irregularities in the lattice.

Does this provide a possible explanation for the rise in the resistivity of metals as temperature rises?

It would be much more satisfactory to be able to predict the value of the resistivity of a substance using the formula (44.9)

$$\text{resistivity } \rho = \frac{m_e}{ne^2\tau}.$$

But there is no independent means of measuring the relaxation time. Nor can we predict quantitatively the way in which the resistivity ρ varies with temperature since we cannot independently measure the way in which the relaxation time varies with temperature.

44.2 THERMAL CONDUCTIVITY

In finding a value for the mean free path of electrons, we assumed that their motion was related to the

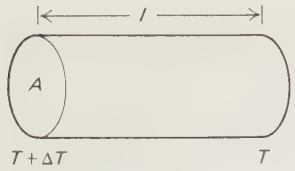


Fig. 44.2

kinetic energy they possessed by virtue of the fact that they were in thermal equilibrium with their surroundings. If there is a difference in the temperature between the two ends of a metal bar, the mean kinetic energy of the electrons at the hotter end will exceed that of the electrons at the cooler end (Fig. 44.2).

At the hotter end where the temperature is $T + \Delta T$ the mean kinetic energy of the electrons will be $3/2\ k(T + \Delta T)$: at the cooler end it will be $3/2\ kT$ (where k = Boltzmann's constant).

We may expect a general diffusion of electrons from the hotter end where the mean speeds will be higher to the cooler end where the mean speeds will be lower. If the temperature difference is maintained in some way, each electron loses $3/2\ k\ \Delta T$ joules on average in the course of drifting from one end to the other. It is as though a force acts in the direction of the drift, under whose influence each electron acquires an average drift velocity $\bar{u}$. So energy transferred to the electron is Fl, where l is the length of the bar, and F the force. The effective force F is given by

$$Fl = \tfrac{3}{2}k\ \Delta T.$$

As before (see Eq. 44.3)

$$F2\tau = m_e v$$

so that

$$\bar{u} = \frac{0 + v}{2} = \frac{F\tau}{m_e}.$$

Substituting for F we obtain

$$\bar{u} = \frac{3}{2}\frac{k\ \Delta T}{l}\frac{\tau}{m_e}.$$

Each electron carries with it energy $3/2\ kT$ which is transported from the hot end to the cold.

Rate of energy transfer associated with n electrons per unit volume moving perpendicularly to a cross-section of area A with velocity $\bar{u}$ is

$$\begin{aligned}\frac{dQ}{dt} &= (\bar{u}An)\frac{3}{2}kT \\ &= \frac{3}{2}\frac{k\ \Delta T}{l}\frac{\tau}{m_e}An\frac{3}{2}kT \\ &= \left(\frac{3}{2}\right)^2 nk^2T\frac{\tau}{m_e}\frac{\Delta T}{l}A.\end{aligned}$$

Comparing this with Eq. (44.1) for the thermal conductivity λ of a material, we see that

$$\lambda = \left(\frac{3}{2}\right)^2 \frac{nk^2T\tau}{m_e}. \qquad (44.10)$$

Hence the ratio

$$\begin{aligned}\frac{\text{thermal conductivity}}{\text{electrical conductivity}} = \frac{\lambda}{\sigma} &= \lambda\rho \\ &= \left(\frac{3}{2}\right)^2 \frac{nk^2T\tau}{m_e}\cdot\frac{m_e}{ne^2\tau} \\ &= \left(\frac{3}{2}\right)^2 \frac{k^2T}{e^2}. \qquad (44.11)\end{aligned}$$

This expression is independent of the nature of the material and this is consistent with the empirical result obtained by Wiedemann and Franz. Furthermore, it appears that the ratio λ/σ is proportional to the absolute temperature T of the material. This fact had been observed by Lorentz in 1872.

Let us look at some numerical values.

$$\begin{aligned}\left(\frac{3}{2}\right)^2\frac{k^2}{e^2} &= \left(\frac{3}{2}\times\frac{1.38\times10^{-23}}{1.60\times10^{-19}}\right)^2 \\ &= 1.58\times10^{-8}\ \text{J K}^{-1}\ \text{C}^{-1}.\end{aligned}$$

Using the values already quoted for copper at 273 K, we obtain

$$\begin{aligned}\frac{\lambda}{\sigma T} = \frac{\lambda\rho}{T} &= \frac{3.9\times10^2\times1.58\times10^{-8}}{273} \\ &= 2.28\times10^{-8}.\end{aligned}$$

Problem 44.2 At 573 K the resistivities and thermal conductivities of copper are

$$\lambda = 3.76 \times 10^{-2} \text{ W m}^{-1} \text{ K}^{-1},$$

and

$$\rho = 3.6 \times 10^{-8} \text{ ohm m.}$$

What is the value of $\lambda/\sigma T$ at this (different) temperature?

The values of λ and ρ at 273 K for sodium are

$$\lambda = 1.35 \times 10^{-2} \text{ W m}^{-1} \text{ K}^{-1},$$

and

$$\rho = 4.2 \times 10^{-8} \text{ ohm m,}$$

and for aluminium are

$$\lambda = 2.38 \times 10^{-2} \text{ W m}^{-1} \text{ K}^{-1},$$

and

$$\rho = 2.45 \times 10^{-8} \text{ ohm m.}$$

What is the value of $\lambda/\sigma T$ in each of these two cases?

How do the values obtained for $\lambda/\sigma T$ compare with each other?

How well do you think they compare with

$$\left(\frac{3}{2}\right)^2 \frac{k^2}{e^2} ?$$

The simple model of a free electron gas predicts a value for

$$\left(\frac{3}{2}\right)^2 \frac{k^2}{e^2}$$

which is consistently about 25% too low when compared with experiment. And yet the model has undoubted success to its credit. For example, the model metal obeys Ohm's law; its resistance rises with temperature; its thermal conductivity is proportional to its electrical conductivity and the ratio is proportional to the absolute temperature; its prediction of this ratio is of the right order of magnitude. The model is surprisingly successful when set against the background of the simplifying assumptions we made at the beginning and this observed error in its prediction of the value of λ/σ.

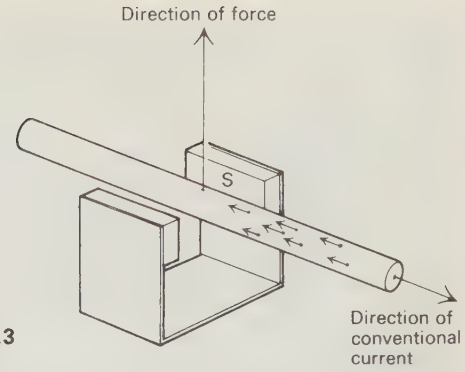


Fig. 44.3

Let us see if the model has any more features which we may look for in the real metal.

44.3 THE FORCE ON A CURRENT-CARRYING WIRE IN A MAGNETIC FIELD

The Drude model of an electric current is a stream of electrons moving with a drift velocity $\bar{u}$ under an applied electric field. How will this stream behave in a magnetic field?

The magnetic force will tend to deflect the stream of negative charge in the direction shown (Fig. 44.3). As the charges become displaced in the wire, electric fields will be set up in it owing to the fact that the displaced negative charge leaves behind it the positively charged ions. The electric field will act in such a direction that the positively charged ions will try to move with the stream of negative electrons. As a consequence, the whole wire moves upwards under the action of the magnetic field.

44.4 THE HALL EFFECT

E. H. Hall, in 1879, showed that when a conductor was placed in a magnetic field and constrained from moving, a p.d. appeared across the dimension of the conductor perpendicular to field and current. The

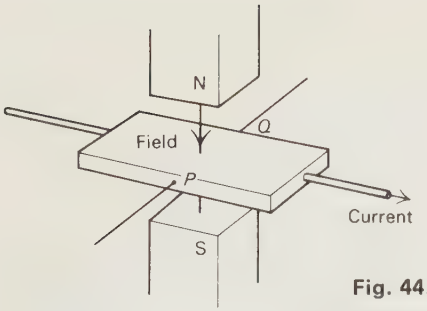


Fig. 44.4

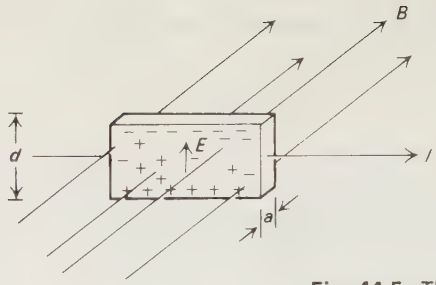


Fig. 44.5 The Hall effect.

p.d. appears across PQ if the materials are arranged as in Fig. 44.4.

The free-electron gas model of a solid readily supplies an explanation of this effect. If the conductor is constrained from moving, the free charges carrying the current will move until the opposing electric field set up by their displacement just balances the magnetic deflection force. Thus if a p.d. V_H appears across a slice of width d, the internal electric field E set up by the displaced charges is

$$E = \frac{V_H}{d},$$

(see Fig. 44.5). The internal field E gives a downward force Ee on each electron in the electron stream. This opposes the magnetic deflecting force $Be\bar{u}$.

$Ee = Be\bar{u}$ ($\bar{u}$ = drift speed of electrons),

$$E = B\bar{u}. \tag{44.12}$$

As before $I = (ad\bar{u})ne$, where a = thickness of slice. Substituting for E and B in Eq. 44.12 we get

$$\frac{V_H}{d} = \frac{BI}{adne}$$

$$\therefore \frac{V_H a}{BI} = \frac{1}{ne}. \tag{44.13}$$

Thus for a given specimen, V_H depends upon a, B, I, and a constant $1/ne$ for the material. The ratio $V_H a/BI$ is called the Hall coefficient for a material.

Apart from providing a qualitative test of the model, the existence of the Hall voltage, as V_H is called, enables two further measurements to be made.

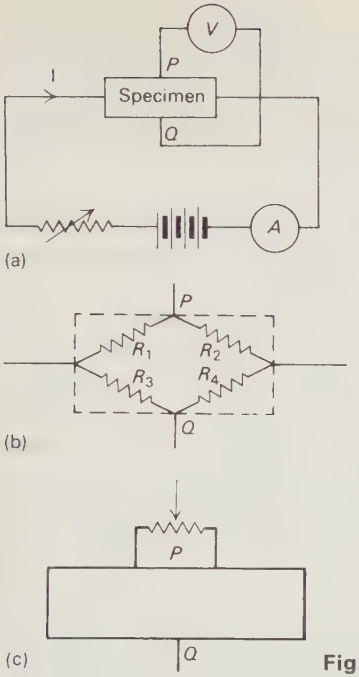
a) The sign of the Hall voltage provides a direct test of the sign of the charge carrier carrying the electric current within a conductor. For the direction of the magnetic deflecting force is independent of whether the conventionally directed current consists of positive charges moving in the direction of the conventional current, or negative charges moving in the opposite direction. But the top of the slice in Fig. 44.5 is negative if negative charges are deflected and positive if positive charges are deflected.

The results of experiments indicate that metals such as copper, aluminium and sodium all have negative Hall voltages in accordance with the electron-drift theory of the electric current.

b) By measuring the size of the Hall voltage obtained when a known current passes through a conductor of measured thickness in a magnetic field, the electron density can be directly determined. This value can be compared with the expected electron density obtained using chemical evidence.

To measure the Hall voltage for copper, a thin foil 0.05 mm thick is covered on each side by thin plastic film and sandwiched between the faces of a pair of cores made of a suitable soft iron which, in turn, form the core of an electromagnet. In this way the foil can be subjected to a field of about 1 tesla. On passing a current of about 10 A through the foil, a p.d. of a few microvolts can be detected and measured across the specimen. A bridge network of resistors has to be used to compensate for potential differences which may be developed within the foil.

The metal foil forms part of a circuit as shown in Fig. 44.6. This is equivalent to the circuit shown at



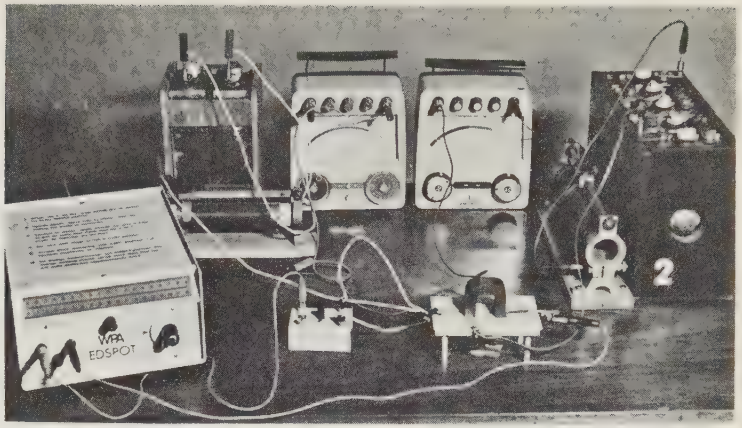


Fig. 44.7 Apparatus for the experimental investigation of the Hall effect. (Photograph by E.J.W.)

Fig. 44.6 Circuit for investigating the Hall effect.

Fig. 44.6b, so far as any potential difference across the specimen between P and Q are concerned. In the absence of a magnetic field through the specimen, the potential difference V_{PQ} between the points P and Q will be zero if

$$\frac{R_1}{R_2} = \frac{R_3}{R_4}.$$

The values of these ratios depend on the exact points of contact at P and Q and are most unlikely to be equal to one another. To compensate a small rheostat (1 to 2 ohms) is placed across the specimen as shown in Fig. 44.6c. By adjusting this rheostat, the potential difference V_{PQ} can be made equal to zero in the absence of the field. Figure 44.7 shows a typical experimental arrangement.

Problem 44.3 The following readings were obtained in a typical experiment: copper foil 0.05 mm thick was subjected to a field of 1.3 tesla, when carrying a current of 8 A. The resultant Hall voltage was 8×10^{-6} V.

Calculate the value of the Hall coefficient for copper. Hence evaluate the electron density. Compare this with the number of copper atoms per cubic metre (see Problem 44.1). Does the ratio number of electrons/atoms for the electron density compare well with the expected number on the basis of chemical valency and electrolysis experiments with copper sulphate?

It still appears that the free-electron gas model gives a good description of the electrical properties of metallic conductors. Before we look at some of the phenomena for which the model has failed to provide an adequate explanation, we will consider one further success.

Since the conduction of heat seems to depend upon electron flow, it follows that a metal subject to a

temperature gradient should exhibit an electrical potential gradient as well. This is a well-known phenomenon—part of a group of phenomena falling under the collective title *thermoelectricity*. This particular effect is called the Thomson effect (named after its discoverer W. Thomson, later Lord Kelvin).

Two of the most notable failures of the free-electron gas model are in its predictions of the heat capacity of a metal, of the variation of resistance with temperature and in some observations made of the Hall effect. The former is perhaps its most spectacular failure.

We saw in Unit 3 how Dulong and Petit showed as long ago as 1819 that the molar specific heat of all metals was about 25 J mol^{-1} K^{-1}. The uniformity of this result can be readily explained in terms of the vibrational energies of the solid ions, rigidly bound together in a lattice. No other source of energy is required. However, the free-electron gas model suggests that each electron has thermal kinetic energy of $3/2\ kT$. Thus each mole of electrons should contribute a further 3/2 R to the specific heat of the metal—about 12.5 J mol^{-1} K^{-1}. Divalent metals will contribute 2 moles of electrons per mole of ions and the specific heat contribution will be 25 J K^{-1} per mole of metal ions. That the electron gas does not contribute measurably to the specific heat of a metal is only too apparent—but the Drude model can provide no explanation of why it does not.

The second failure of the model that we will consider in a little detail concerns the Hall voltage. While many simple metals have a negative Hall voltage (in the sense that the charge carriers are negative), a few show a positive voltage. One such metal is zinc, which is otherwise a typical metallic element. Again the Drude model provides no explanation. Such positive Hall voltages do raise the question as to whether positive charge carriers are responsible for the conduction current—such a deduction would be very difficult to understand in the light of the accepted model of the atom.

C. T. Kettering and G. G. Scott carried out an experiment in 1944 which should give direct evidence for the nature of the charge carrier. A current carrying loop will, as a consequence of the charge movement within it, have a small amount of angular momentum. If the current is suddenly reversed, the sudden reversal of the angular momentum will be compensated by some visible change in the momentum of the suspended loop. The direction of this change will depend on the direction of the change of the angular momentum of the charge carriers. Consequently, the direction of the original motion of the charge carriers can be deduced.

(Can you think of some of the difficulties you might encounter in trying to carry out such an experiment?)

Kettering and Scott experimented with a number of metals, including some with positive Hall voltages. In each case they found that the charge carrier itself was negative.

Thus any modifications to the Drude model do not have to make room for positive and negative charge carriers, but rather have to explain how negative charge carriers can give rise to positive Hall voltages.

The remaining difficulties experienced with the Drude model lie in the prediction of quantitative results (as in the size of the constant in Weidemann–Franz law) and in the *ad hoc* nature of the relaxation time, τ.

However, the qualitative successes of the model are many and its improvement will take the form of modifications rather than of wholesale replacement. This is a story beyond the scope of this book.

Chapter 45
IONIZATION AND ENERGY LEVELS

45.1 The experiment of Franck and Hertz
45.2 An application of the Franck-Hertz tube

45.1 THE EXPERIMENT OF FRANCK AND HERTZ

In Unit 4 and again in the last few pages we have been discussing a model for matter, whether a gas or an electron-gas, in which particles (atoms, molecules, or electrons) collided with one another or with walls or even irregularities without damage. The collisions were regarded as being perfectly elastic.

Under different circumstances, notably in a cloud chamber, or a positive ray tube, or a fine beam tube, we have seen how interactions between particles have caused changes which involve transformations of energy. A collision between an alpha particle and a molecule or between an electron and a molecule may ionize the molecule. Such collisions are probably inelastic.

It is legitimate to enquire why, in most cases of collision between particles, the collisions are elastic when, under other circumstances, the collisions result in ionization. Indeed, what are the conditions which lead to gaseous ionization? Controlled experiments are hardly possible in atmospheric air as we find it in a laboratory: but we can exercise control if we enclose some gas in a tube, at low pressure and examine its properties when bombarded by a stream of electrons. The pioneers in this field were J. Franck and G. Hertz (nephew of H. Hertz).

The apparatus was simple: a long wire filament was mounted along the axis of a cylinder of wire mesh which was itself surrounded closely by a cylinder of metal foil. This assembly was enclosed in a glass vessel filled with the atoms to be studied but at a low pressure.

This is represented diagrammatically in Fig. 45.1.

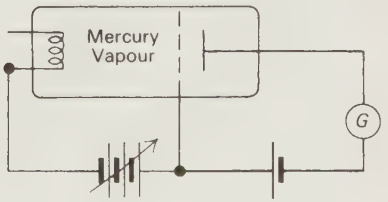


Fig. 45.1 Schematic diagram to represent the Franck-Hertz experiment.

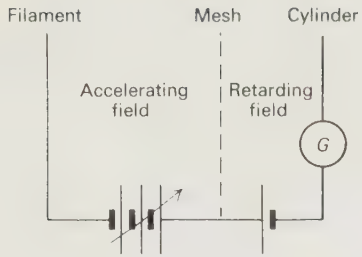


Fig. 45.2 Fields within a Franck-Hertz tube.

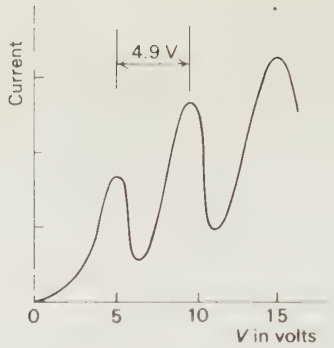


Fig. 45.3 Current/potential graph for a mercury filled tube.

In the experiment, described in 1914, mercury vapour was used at a pressure of about 100 Nm^{-2} (normal atmospheric pressure is 10^5 Nm^{-2}). The equipment was contained in an oil bath at about 380 K.

The filament was heated by the passage of a suitable current and produced thermionic electrons. These could be attracted to the wire-mesh electrode by the application of a small variable potential difference. Electrons moving under the influence of this electric field would, of course, collide frequently with mercury atoms, and many would pass through the wire mesh to fall on the surrounding cylinder. The cylinder was connected to a sensitive galvanometer, its reading providing a measure of the number of electrons arriving each second at the cylinder.

Control of the number so arriving was exercised by applying a small reverse voltage to the cylinder. Some typical field situations (Fig. 45.2) are shown in Table 45.1.

In the experiment itself the retarding potential difference between the mesh and the cylinder was kept at a small fixed value and the accelerating potential difference between the filament and the mesh was raised slowly. The current flowing from the cylinder was recorded. As long as the accelerating potential was smaller than the retarding potential, the current was zero. As the accelerating potential rose above this value, so the current rose in just the same way as the current rises in a vacuum tube. The collisions being elastic, the electrons transferred negligible amounts of energy when in collision with the massive mercury atoms (see Section 7.16). It was just as though the atoms were not there. But when the accelerating potential reached about 5 volts the current through the galvanometer fell almost to zero. The electrons had insufficient energy to cross the gap between the mesh and the cylinder against the retarding field.

Franck and Hertz suggested that the electrons acquired just the right energy to excite the mercury

Table 45.1

P.d. (filament to mesh) in V	Electron energy at mesh in eV	P.d. (mesh to cylinder) in V	Electron energy at cylinder in eV	Current
0	0	−1	No electrons	No current
1	1	−1	No electrons	No current
2	2	−1	1	Current flow
3	3	−1	2	Current flow

atoms in collisions near to the grid. Their energy was all transferred to the excited atoms leaving far too little to overcome the retarding potential.

If the accelerating potential was then raised, the excitation zone in the tube would move away from the mesh towards the filament, so that an electron which had lost its energy in a collision would gain some further energy in the field, pass the grid and possibly reach the cylinder. The current would rise once more (see Fig. 45.3).

Figure 45.3 reveals that the total current flowing reached a larger maximum than before. More electrons were being collected—and these were, presumably, products of the excitation process. At an accelerating potential of about 10 V, the current fell for the second time, when the excitation zone occurred twice in the space between the filament and the mesh.

Excitation occurred in two zones in the space between the filament and the mesh; the electrons suffered two inelastic collisions and the sequence of phenomena was seen to recur again and again as the accelerating potential was further raised. Franck and Hertz concluded that an electron required an energy of 4.9 eV or 7.8×10^{-19} J if it was to excite a mercury atom. Or, to look at it from another point of view, a mercury atom can accept energy in a 'packet' of 7.8×10^{-19} J.

This experiment by Franck and Hertz was the first of a series performed by many workers. In the case of mercury, the excitation at 7.8×10^{-19} J (4.9 eV) and at multiples of this energy which was so marked in the original experiment obscured the complexities of the phenomenon. Using another method devised by Hertz in 1923 and developed by J. C. Morris in 1928, a whole series of excitation levels for mercury was revealed. This series includes ionization (i.e., the removal of an electron from the atom) at an energy of 16.5×10^{-19} J (10.3 eV).

Subsequently the critical potentials of many other gases have been studied. In every case there is a discrete energy packet which atoms can accept and which causes ionization. See Table 45.2.

Table 45.2 Ionization energies

Hydrogen	13.6 eV	or	21.8×10^{-19} J
Helium	24.6 eV	or	39.2×10^{-19} J
Nitrogen	14.5 eV	or	23.2×10^{-19} J
Oxygen	13.6 eV	or	21.8×10^{-19} J
Neon	21.6 eV	or	34.4×10^{-19} J
Argon	15.8 eV	or	25.3×10^{-19} J
Xenon	12.1 eV	or	19.3×10^{-19} J

45.2 AN APPLICATION OF THE FRANCK-HERTZ TUBE

Gas-filled thermionic tubes are used for a number of industrial purposes. Some of these are constructed on the same principles as the Franck-Hertz tube with its three electrodes. Such gas-filled triodes with a cathode, a grid and an anode can perform as quick-acting switches. If the potential difference between the anode and the cathode is high enough, the tube will be conducting with the gas filling it in an ionized state. If now the grid is made sufficiently negative, no electrons will reach the anode; no anode current can flow. If the grid is made positive, current will flow once more. Such tubes, which are called thyratrons, can handle comparatively heavy currents at fairly high voltages.

Small thyratrons usually contain Xenon, Argon, or Helium, and may be used to investigate the effects described by Franck and Hertz. However, for various reasons connected with their design, one cannot expect to get results which match those now generally accepted. If a tube does give such a result then the choice of tube was a fortunate one.

Problem 45.1 Molecules of the gases which make up the air do not appear to ionize one another although we assume that collisions are frequent. Why is this?

The energy of a typical molecule at room temperature (about 300 K) is $3/2\ kT$, where k is Boltzmann's constant. So energy

$$= 3/2 \times 0.69 \times 10^{-23} \times 300 \text{ J},$$

$$\approx 3 \times 10^{-21} \text{ J (about } 2 \times 10^{-2} \text{ eV)}.$$

Now, typically it requires about 20×10^{-19} J to ionize a gas particle. The mean energy of the gas molecules is about 600 times too small at room

temperature and it is unlikely that any molecules in the energy distribution will have enough energy to succeed in causing ionization during a collision.

As we have seen in Unit 4, the energies of the molecules are grouped in the normal distribution pattern so, even when the mean energy is as low as this, a few molecules will have considerably higher energies. If the gas is heated appreciably, by contact with a hot wire, for example, ionization can be detected. A few molecules acquire sufficient energy to ionize others with which they may collide.

Chapter 46
ELECTRICITY, MATTER, AND LIGHT

46.1 Photoelectricity
46.2 Energy levels
46.3 Summary
46.4 The Sun's spectrum
46.5 Absorption of ultraviolet radiation in the atmosphere
46.6 The stimulated emission of radiation
46.7 The helium-neon laser

46.1 PHOTOELECTRICITY

It will be recalled that electrons can be produced within a vacuum tube by heating a filament or by irradiating an electrode with ultraviolet light. This introduces yet another interaction—but this one is between an electromagnetic wave and matter. Photoelectric cells based on this effect are now commonplace and enable us to explore the phenomenon with relative ease. Such cells usually comprise an emitter coated with an alkali metal and a thin wire electrode to collect the electrons (Fig. 46.1). If the collector C is made positive with respect to the emitter S any electrons which may be produced will move to the collector and a small current will flow.

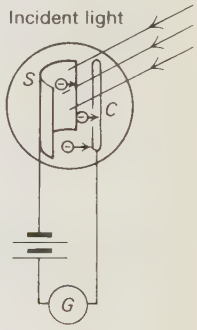


Fig. 46.1 A photo-cell. (From F. W. Sears and M. W. Zemansky, *University Physics*, Addison-Wesley 1970.)

In the light of the results of the Franck-Hertz experiments we are especially interested in the energy with which these electrons are emitted. Suppose that the voltage applied to the cell is carefully reversed. At some reversed voltage the current through the cell will fall to zero because the electrons do not have sufficient energy to cross to the collector against the retarding field. This voltage is a measure of the energy of the most energetic electrons emitted.

What factors might affect the emission and energy of the photoelectrons? Intensity and frequency of the incident light are possibilities.

It was at the beginning of the twentieth century that P. E. A. von Lenard made a systematic study of this effect, which had been discovered by H. R. Hertz thirteen years earlier while working on the electromagnetic waves generated by his spark transmitter.

The facts which Lenard uncovered are listed below. Note that the light is referred to by its frequency (f) rather than by its wavelength (λ). Remembering that $f\lambda = c = 3 \times 10^8$ m s^{-1}, then red light with a long wavelength has a lower frequency than violet light with a short wavelength.

a) The emission of electrons does not occur for light of all frequencies. If below a certain value, called the *threshold frequency*, f_0, then emission will not occur *no matter how intense the light.*
b) For a given frequency f (where $f > f_0$) the kinetic energy of the emitted electrons has a spread of values from zero up to a maximum value, E_{max}, which is proportional to $(f - f_0)$. For any chosen emitter E_{max} has the same value, *no matter how weak or intense the light.*
c) If emission occurs, it does so as soon as the light reaches it (in modern terms, within 10^{-9} s), *no matter how weak the light.*
d) If emission occurs, the number of electrons emitted per second (i.e., the electron current) is proportional to the light intensity.
e) The value of the threshold frequency, f_0, depends upon the material of the surface being illuminated.

Points (a), (b), and (c) are surprising. There is emission if the frequency f is only a little above the threshold f_0 even if the light is very dim; and making the light brighter does not increase the maximum energy of the electrons. It only increases the number emitted. On the other hand, even with the most intense source of light which is available no emission is possible if its frequency is below the threshold, no matter by how little.

It was A. Einstein who, in 1905, related these phenomena to the quantum theory proposed by M. Planck two years earlier. In this model Planck had suggested that the energy of a beam of light was made up of a number of unit 'packets' of energy—the so-called light-quanta or *photons*, and that the size E of the unit was given by the equation

$$E = hf, \tag{46.1}$$

where f is the frequency of the light and h a constant now known as Planck's constant. This constant is very small and has the value 6.63×10^{-34} J s, to 3 significant figures.

Einstein pointed out that the photoelectric emission of electrons could be explained if it was assumed that the quanta of energy penetrate the surface of the metal and that their energy is transferred into kinetic energy of the electrons. If the electron is at the surface (i.e., is ejected) it will have kinetic energy equal to that of the absorbed photon less any energy (Φ) necessary to remove it from the field within the metal, that is

$$\tfrac{1}{2}mv^2 \text{ (max)} = hf - \Phi. \tag{46.2}$$

The term Φ is known as the *work function.*

The model accounts satisfactorily for the experimental results: more photons cause the emission of more electrons; the greater the frequency of the photon, the greater the energy of the emitted electrons.

This must make us reconsider the Franck-Hertz experiment where we observed that an atom could accept energy from an electron provided that there was enough. In the case of mercury, the appropriate energy was 4.9 eV or 7.8×10^{-19} J. The excited atom must eventually return to the unexcited state. If the extra energy is emitted as a photon of light, that photon might have a frequency given by the relationship $E = hf$, that is

$$f = \frac{7.8 \times 10^{-19}}{6.6 \times 10^{-34}} \text{ Hz}$$

$$= 1.18 \times 10^{15} \text{ Hz}$$

which corresponds to a wavelength of 2.5×10^{-7} m. This lies in the deep ultraviolet and reference to tables of line spectra confirms the existence of a line with a wavelength of 2.54×10^{-7} m.

46.2 ENERGY LEVELS

The experiments show that, in this excitation, a mercury atom accepts just a single quantum (7.8×10^{-19} J) of energy and that the line spectrum of mercury includes a line of such a frequency that the photon involved carries exactly the same quantum of energy. In ionization, the mercury atom accepts a

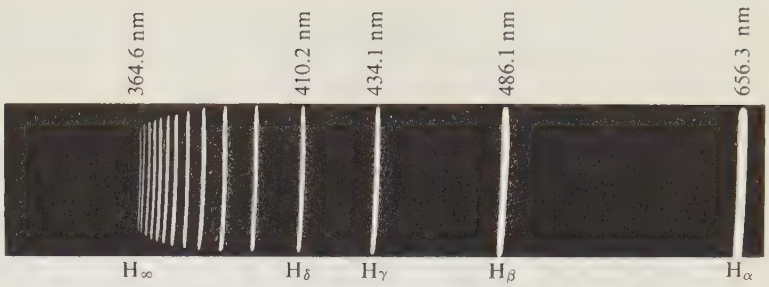


Fig. 46.2 The Balmer series of atomic hydrogen. (From G. Herzberg, *Atomic Spectra and Atomic Structures*, Prentice-Hall, 1937. Reproduced by permission of the publisher.)

single quantum (16.5×10^{-19} J) and the line spectrum includes a line of such a frequency that the photon involved carries exactly this quantum of energy. The atomic model must accommodate these observations. Take the latter first. Let us assume that the neutral atom accepts 16.5×10^{-19} J from an electron which collides with it and that, as a result, an electron is removed from the atom to leave it ionized. When re-combination occurs between this positive ion and an electron, the energy of the incoming electron is just 16.5×10^{-19} J and this appears as a photon of electromagnetic radiation.

What of the other line? Indeed, the line spectrum of mercury contains many other lines; photons of many different energies are emitted by excited mercury atoms. It appears that the electrons in the unexcited atom can accept various discrete energies as a result of collisions and subsequently re-emit that energy in the form of a photon. This has led to the idea of the *energy level*. In the case considered, an electron can exist at certain levels of energy. Subsequently, it may return to a lower energy state and in so doing the atom emits the appropriate photon.

As we have indicated the line spectrum of mercury is a very complex one. So, in accordance with usual practice, let us turn to something simpler—hydrogen. This too is found to possess a complex line spectrum which was extensively studied in the 1880s by J. J. Balmer (see Unit 9). Balmer was a mathematics teacher in Switzerland and was interested in mathematical puzzles rather than physics.

Figure 46.2 shows a photograph of the lines with which Balmer worked and some of the wavelengths are given. This series, which lies almost entirely in the visible spectrum, shows some remarkable regularities. The lines are spaced closer and closer together as we move from the red (long wavelength) to the violet end of the spectrum. And there seems to be a limiting value of wavelength within the series. Balmer developed an empirical relationship to describe this hydrogen series:

$$\lambda = b\left[\frac{n^2}{n^2 - 2^2}\right], \qquad (46.3)$$

where λ = wavelength in nanometres, $b = 364.6$ nm, and n is a positive integer. Each integer, starting with $n = 3$ gives a different line of the series. For the first line (labelled α in Fig. 46.2):

$$\lambda_\alpha = 364.6\left[\frac{9}{9 - 4}\right] = 656.2 \text{ nm}.$$

For the second line (labelled β), $n = 4$ and so the wavelength is:

$$\lambda_\beta = 364.6\left[\frac{16}{16 - 4}\right] = 486.1 \text{ nm}.$$

Table 46.1 shows the results obtained by calculating the wavelengths of seven of these lines up to $n = 9$, and also the values of the wavelengths as measured by experiment. These values are those which Balmer himself reported in the *Annalen der Physik und Chemie* in 1885, and the agreement between the calculated and the experimental values is remarkable. The small discrepancies that do appear in the lower half of the table are due to experimental errors arising

Table 46.1 Balmer's Series in the spectrum of hydrogen

Name of line	n	Wavelength (nm)	
		Calculated	Observed
α	3	656.2	656.2
β	4	486.1	486.1
γ	5	434.0	434.0
δ	6	410.1	410.1
ε	7	396.9	396.8
ζ	8	388.8	388.7
η	9	383.5	383.4

from the use of stars as light sources. More precise determinations have since removed these discrepancies.

This remarkable piece of work provides good reason for naming this series of lines in the hydrogen spectrum after Balmer. We can see how the series tends towards a limiting value which, when $n = \infty$, is 364.6 nm.

Balmer wrote that he was initially aware only of the existence of the first four lines. Imagine his excitement when a friend told him that more lines were known in the spectra of the stars and his satisfaction on discovering that their wavelengths agreed with his prediction. Encouraged by the success of his skill in numerology, Balmer speculated on the existence of hitherto undiscovered lines that would be given by his formula when the 2^2 in the denominator is replaced by such numbers as 1^2, 3^2, 4^2, etc.

Balmer's work inspired others to search for additional lines in the hydrogen spectrum, and to seek similar expressions that would describe the line spectra of elements other than hydrogen. Five years after the publication of Balmer's paper, a Swedish spectroscopist, J. Rydberg, was successful in this latter quest. He was interested in the spectra of the alkali metals, particularly sodium, and found that the complex system of lines in the spectra could be represented by a relationship of the general form:

$$\frac{1}{\lambda} = \frac{R}{(n' + \mu')^2} - \frac{R}{(n + \mu)^2}, \qquad (46.4)$$

where R is Rydberg's constant, μ and μ' are constants both of which are less than 1 and n and n' are positive integers.

This equation reverts to Balmer's formula when $\mu = \mu' = 0$ and $n' = 2$ for then

$$\frac{1}{\lambda} = R\left[\frac{1}{2^2} - \frac{1}{n^2}\right]. \qquad (46.5)$$

You should check that this is the same as Balmer's formula (Eq. 46.3) if

$$R = \frac{2^2}{364.6} \quad \text{or} \quad 0.01097\ \text{nm}^{-1}.$$

It was not until the first decade of the twentieth century that Balmer's speculation about further series of lines within the spectrum of hydrogen bore fruit. C. S. Lyman found lines in the far ultraviolet and F. Paschen reported a series in the near infrared. A total of six such series is now known (see Table 46.2).

Table 46.2 (Wavelengths calculated using Eq. (46.4), putting $\mu = \mu' = 0$.)

Name of series	Date	n'	Values of n	Wavelengths (nm)		Region of spectrum
				First line	Series limit	
Lyman	1906–1914	1	2, 3, 4, . . .	121.5	91.1	Far u.v.
Balmer	1885	2	3, 4, 5, . . .	656.2	364.6	Near u.v. and visible
Paschen	1908	3	4, 5, 6, . . .	1875	820.0	Near i.r.
Brackett	1922	4	5, 6, 7, . . .	4055	1457.6	i.r.
Pfund	1924	5	6, 7, 8, . . .	7460	2277.5	Far i.r.
Humphrey		6	7, 8, 9, . . .	12360	3281.4	Far i.r.

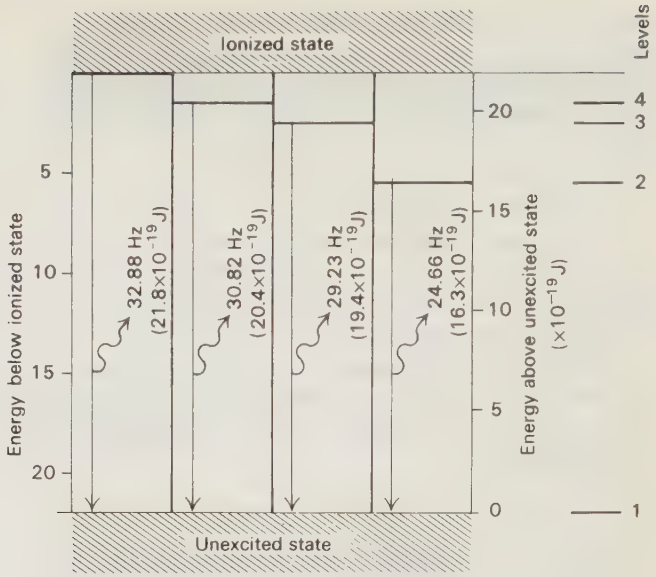


Fig. 46.3 Some energy levels of atomic hydrogen.

Let us now consider the Lyman, Balmer and Paschen series of lines in the hydrogen spectrum in the light of the suggestion that each line represents a different photon with an energy which is associated with the energy of an electron within the atom. Table 46.3 tabulates the frequencies of some of the lines in these series.

If we assume that we may calculate photon energies for each line by applying Planck's formula $E = hf$ we find that these lines correspond to photon energies as given in Table 46.4.

Inspection reveals one familiar number: 21.8 $\times 10^{-19}$ J. This was the energy quoted for the process of ionization by collision in Franck-Hertz type experiments with hydrogen.

If we assume that the electron of an unexcited hydrogen atom exists in a *ground state* of unknown energy which we then take as an arbitrary zero of

Table 46.3 The Hydrogen spectrum

Lyman series frequency $\times 10^{14}$ Hz	Balmer series frequency $\times 10^{14}$ Hz	Paschen series frequency $\times 10^{14}$ Hz
24.66	4.57	1.60
29.23	6.16	2.34
30.82	6.90	2.74
31.56	7.31	2.98
31.97	7.55	3.14
32.21	7.71	
32.36		
–		
–		
Limit 32.88	8.24	3.66

Table 46.4 The Hydrogen spectrum photon energies

Lyman series energy $\times 10^{-19}$ J	Balmer series energy $\times 10^{-19}$ J	Paschen series energy $\times 10^{-19}$ J
16.3	3.02	1.06
19.4	4.10	1.55
20.4	4.57	1.82
20.9	4.85	1.97
21.1	5.00	2.08
21.3	5.11	
21.4		
–		
–		
Limit 21.80	5.46	2.42

energy and then measure other electron energies from this arbitrary zero, we may represent the possible energy states as a series of steps or levels as in Fig. 46.3. In this we have assigned the number 1 to the ground state, 2 to the first excited state at 16.3×10^{-19} J and so on to the limit at 21.8×10^{-19} J. We see that these steps get closer and closer together as we approach the limit and ionization. And we may see how each of the Lyman lines can be associated with the transition of an electron from any one level to the ground state for which $n' = 1$.

What is likely to happen if the transitions occur to the level for which $n' = 2$? Presumably the photons emitted will have energies which correspond to transitions from

$n = 3$ to $n' = 2$

19.4×10^{-19} to $16.3 \times 10^{1-9}$ J $= 3.1 \times 10^{-19}$ J,

$n = 4$ to $n' = 2$

20.4×10^{-19} to 16.3×10^{-19} J $= 4.1 \times 10^{-19}$ J,

$n = 5$ to $n' = 2$

20.9×10^{-19} to 16.3×10^{-19} J $= 4.6 \times 10^{-19}$ J.

These energies correspond very closely to those for the photons of the Balmer series. The model can accommodate this series as well as the Lyman series.

Problem 46.1 Use the energy level model to account for the Paschen series of lines in the hydrogen spectrum (Fig. 46.4).

46.3 SUMMARY

A reasonable model for an atom allows a bound electron to exist in a number of *excited states*. When unexcited (in the ground state) and when excited the electron is bound to the nucleus—in one of a large number of bound states. The gaps between these energy levels get closer together as the atom gets nearer to being ionized; once ionized, the electron can exist with any energy moving freely past the nucleus.

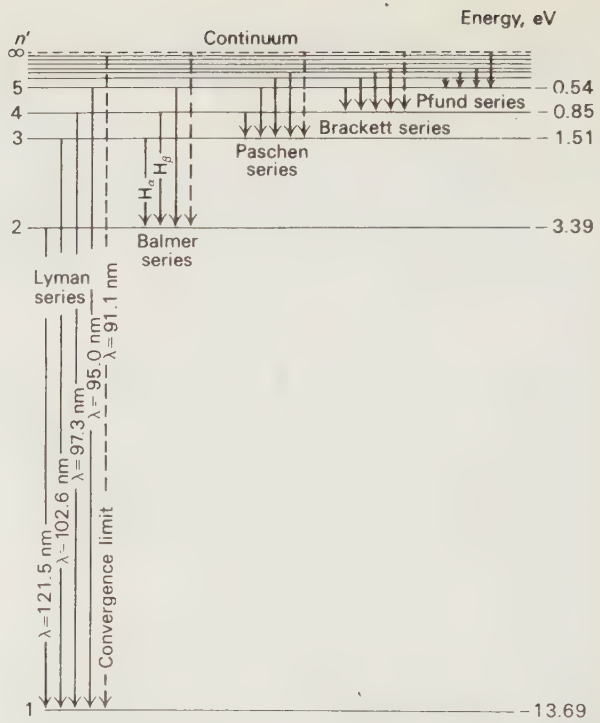


Fig. 46.4 Radiative transitions in hydrogen. (From M. Alonso and E. J. Finn, *Physics*, Addison-Wesley, 1970.)

These energy levels are very well defined and simple numerical relationships exist between them. They are often referred to as *stationary states*.

The emission line spectra of atoms are associated with transitions of electrons from energy level to energy level, the excess energy determining the energy (and therefore the frequency) of the radiation emitted.

Problem 46.2 Figure 46.5 which is a reproduction of a negative, shows what happens when white light is passed through sodium vapour and then examined in a spectroscope. Dark lines appear on a bright field and the wavelengths are identical with those appearing in the emission spectrum of sodium. Account for this absorption spectrum in terms of the idea of energy levels.

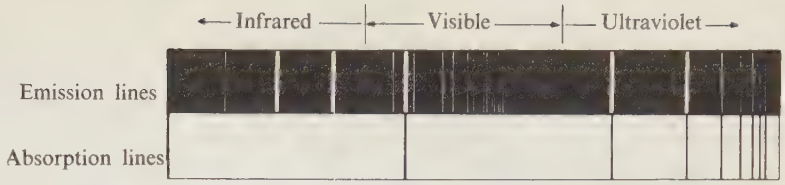


Fig. 46.5 Comparison of the emission and absorption spectra of sodium. (From A. B. Arons, *Development of the Concepts of Physics*, Addison-Wesley, 1965.)

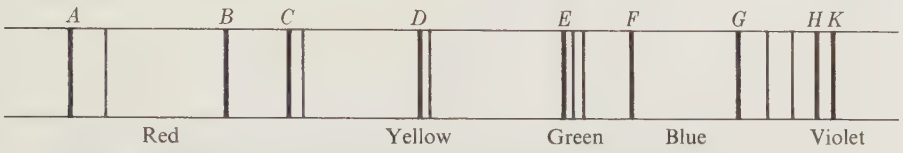


Fig. 46.6 The Fraunhofer dark lines in the visible spectrum of sunlight. Only a few of the prominent lines are shown. (From A. B. Arons, *Development of the Concepts of Physics*, Addison-Wesley, 1965.)

46.4 THE SUN'S SPECTRUM

Figure 46.6 is a splendid example of an absorption spectrum. The body of the sun emits a continuous spectrum. This light passes through a cooler atmosphere and many of the photons are absorbed by the atoms existing in the atmosphere. These excited atoms then emit their characteristic photon. This emission takes place in all directions and the intensity of the light of that frequency in the forward direction is reduced by comparison with the bright continuous background and the line appears dark rather than bright. Such dark lines were observed first by J. von Fraunhofer and are called Fraunhofer lines.

46.5 ABSORPTION OF ULTRAVIOLET RADIATION IN THE ATMOSPHERE

The sun's spectrum includes such ultraviolet radiation as the Lyman α line for hydrogen. Such energetic photons are capable of damaging living organisms and, indeed, it is highly unlikely whether life as we know it could have developed had some mechanism not absorbed this radiation.

This process occurs within the atmosphere and is thought to be photochemical in nature. It causes the dissociation of molecular oxygen. Combination of the resulting atomic oxygen with molecular oxygen forms ozone which also absorbs ultraviolet light strongly. The two processes ensure that few of the high energy ultraviolet photons penetrate below a height of 75 km. Indeed the photons with the highest energies (wavelengths below 100 nm) are absorbed above a height of about 170 km, whilst the Lyman α photons (at 122 nm) penetrate to about 75 km. It will be appreciated that the consequence of such absorption is the development of ionization in the atmosphere and of the so-called ionosphere with its ability to reflect and to absorb radio waves. (Dobson, 1968.)

Major examples of photo-chemical reactions include photosynthesis (the building up of complex organic structures from the synthesis of carbon dioxide and water in living green plants) and the photographic process.

46.6 THE STIMULATED EMISSION OF RADIATION

Figure 46.5 shows what happens when white light passes through sodium vapour. Light of a frequency corresponding to the normal emission frequencies of sodium is absorbed. A radiation quantum corresponding to a frequency in the sodium emission spectrum carries just the right energy to raise an electron in the sodium atom from its ground state to an excited state (Figure 46.7). Consequently radiation

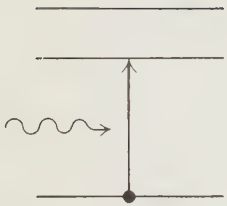


Fig. 46.7 Excitation.

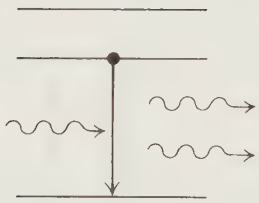


Fig. 46.8 Stimulated emission.

of this frequency is absorbed and dark lines appear in the continuous white light spectrum. Of course, photons of this frequency are emitted when the electron returns to the unexcited state but such emission will take place in all directions so the intensity of the beam in the original direction is reduced.

If, by chance, some of the sodium atoms within the radiation field are already at this higher energy level, interaction with the incident photons can still take place. But now the atom falls to a lower energy level, itself emitting a quantum of energy of the same size as that of the incident photon (see Fig. 46.8). This process is called *stimulated emission.* The interesting feature of this process is that the stimulated quantum of radiation is precisely in phase with the original quantum. Consequently, the on-going radiation is enhanced in energy but remains coherent (see Chapter 23).

Normally energetic atoms fall to lower energy levels of their own accord and at random. This process of spontaneous emission leads to the production of quanta which bear random phase relationships with each other and the radiation is non-coherent.

Einstein showed that, if there were equal numbers of atoms in the ground state and an excited state, an incident photon of the right energy has the same chance of stimulating emission as of being absorbed. If a situation could be produced in which there were more atoms in a high excited state than in a low one, stimulated emission would be more probable than absorption and a beam of highly coherent radiation could be produced. However, this is an improbable state of affairs under normal circumstances and this is fortunate for highly coherent light radiation can be dangerous to the sight.

In normal equilibrium when a beam of light radiation is passing through a suitable medium, more atoms will be in low energy states than in high ones. The reversed state to this is associated with a population inversion and is utilized in the laser. The word *laser* is derived from the initials of the expression "light amplification by the stimulated emission of radiation."

46.7 THE HELIUM-NEON LASER

This is one of the commonest forms of laser. The neon gas is the source of the stimulated emission. The helium gas is the agency which produces a population inversion in the neon.

By chance, one energy level in helium is identical with an energy level in neon and it is also "metastable." This means that excited atoms in this energy level tend to stay in it rather longer than is normal before they return to the ground state.

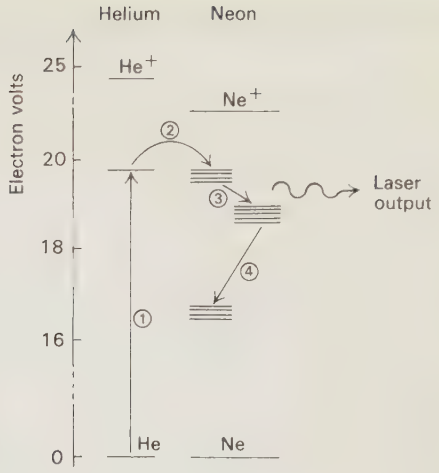


Fig. 46.9 Energy level systems of helium and neon atoms. Successive energy transitions are indicated by the figures in circles. (From D. J. E. Ingram, *Contemporary Physics*, **3**, 435 (1966). Reproduced by permission of the publisher.)

Helium atoms are excited into high energy states by continuous bombardment by electrons in a low pressure gas discharge. These atoms are continuously in collision with each other and with neon atoms. Those which are in the metastable state will, on colliding with a neon atom, have a high probability of transferring this energy to the neon atom so raising it to a high energy level. By suitably adjusting the proportion of helium to neon in the tube it is possible to have more neon atoms in this high energy state than in a lower one. This population inversion will then be maintained in dynamic equilibrium. Radiation quanta, produced by spontaneous emission from neon atoms, will then stimulate others into emission and a beam of highly coherent stimulated radiation will be produced (see Fig. 46.9).

In order to enhance the interaction of the emitted radiation with the energetic neon atoms, the discharge is contained within a tube bounded at one end by a highly reflecting mirror and at the other by a partially reflecting mirror. Thus most quanta traverse the gas many times before escaping in the emitted radiation.

PROBLEMS: Unit Eight

8.1 A voltameter with large platinum plates contains dilute sulphuric acid in which 10^{-2} g of metallic copper has been dissolved. A constant current of 10^{-2} A is then passed through the solution for 25 minutes. Calculate

a) the mass of copper deposited on the cathode,
b) the volume of oxygen liberated at the anode if the pressure is 750 mm Hg and the temperature is 20°C,
c) the additional time for which the same current must be passed before hydrogen is liberated at the cathode.

Oxford and Cambridge Board
A level Physics II, (1961).

8.2 The *electrolytic conductivity* is defined as the conductivity of a column of a solution of the electrolyte 1 m long and 1 m^2 in cross-section. The *molar conductivity* is the product of the specific conductivity and the volume (in m^3) of the solution which contains 1 mole of the electrolyte.

For solutions of strong electrolytes (like copper sulphate) the electrolytic conductivity is almost inversely proportional to the molar concentration. This means that the molar conductivity is almost independent of dilution. For weak electrolytes (like aqueous solutions of acetic acid) the molar conductivity rises as the degree of dilution rises (i.e. as the molar concentration falls).

Write a brief paragraph or two offering an explanation for these differences of behaviour.

8.3 An electrolytic cell 0.1 m wide and 0.1 by 0.02 m in cross-section carries platinum electrodes at each end. The cell is filled with hydrochloric acid of strength 50 mol m^{-3}. A current of 0.5 A passes through the cell. A separate experiment shows that the transport number for the hydrogen ion is 0.83 and that the molar conductivity of the dilute acid is 4×10^{-2} ohm^{-1} m^2 mol^{-1}.

Using this data, answer each of the following questions

a) What percentage of current is carried by the hydrogen ions and what percentage is carried by the chloride ions?
b) Calculate the drift speed of the hydrogen ions and the drift speed of the chloride ions carrying the current.
c) Calculate the resistance offered by the electrolyte to the flow of current through it.
d) Calculate the potential gradient through the electrolyte.
e) The *ionic mobility* is defined as the drift speed of the ions through the field per unit potential gradient. Calculate the ionic mobility of hydrogen ions.

8.4 The ionic mobility is a quantity which should be independent of the current passing through the cell and the dimensions of a particular cell. Assuming that the molar conductivity is Λ ohm^{-1} m^{-1} per mol m^{-3} and that n is the transport number of one of the ions of the electrolyte, repeat the calculation you have just made to show that the ionic mobility of the ion is

$$\frac{n\Lambda}{N_A e},$$

where N_A is Avogadro's number (6×10^{23} mol^{-1}) and e is the charge on the electron.

8.5 A piece of silver foil, 5.2×10^{-2} mm thick carries a current of 10.0 A. When a magnetic field is applied perpendicular to the flow of the current, a transverse e.m.f.

gives a deflection on a galvanometer of 12.2 mm (Hall effect). The galvanometer's sensitivity is given as 0.71 mm/μV.

To find the magnitude of the magnetic field, a 20 turn coil was wound round the core of the electromagnet generating it. The area of the coil was found to be 2.90×10^{-4} m^2. A galvanometer was connected across the coil with a 10^5 ohm resistor in series with coil and galvanometer. On switching on the magnetic field, a throw (corrected for damping) of 30 mm was recorded on the galvanometer. An 0.05 F capacitor raised to a p.d. of 1 volt was found to give the same (ballistic) deflection when discharged through the galvanometer.

Calculate the value of the magnetic field and hence find the number of conduction electrons per cubic metre in silver.

The atomic weight of silver is 108 and its density is 10.5 g cm^{-3}. Assuming that there are 6.02×10^{23} atoms per gramme atomic weight, calculate the number of atoms of silver per cubic metre of the material and hence find the number of free (conduction) electrons contributed by each atom.

8.6 We have seen that the specific charge of an electron (e/m_e) can be determined from a measurement of the deflection of a beam of electrons in a magnetic field of known size, provided the speed of the electrons is known.

a) Write down an expression relating e/m_e to the radius of curvature r of the path of a beam in a field B if the electrons have a speed v.

b) In the text, we described two methods of determining speed v,
   i) by applying a known electric field E,
   ii) by determining the potential energy V lost in an electron gun.

A third technique is available. The beam is allowed to impinge on a small metal target of known mass, M, specific heat capacity, c, and electrical capacitance, C, for a measured time t. The resulting temperature rise ΔT and the rise in potential ΔV of the target are then measured.

If N electrons hit the target in time t, write down expressions for the total energy $N(\frac{1}{2}m_e v^2)$ and the total charge Ne carried by the electrons. Use these two expressions to eliminate N and so derive an alternative expression for V in terms of e/m_e.

Assuming that the electrons were accelerated through a potential of the order of 10^4 V, describe briefly, indicating sizes where possible, the essential design features of the experiment if the target were made of copper. Consider the following points.

How would you measure the temperature rise?
How would you measure the rise in potential?
What would be appropriate values for ΔT and ΔV?
What would have to be the resultant mass of the target and what would be its heat capacity.

8.7 a) Two exactly similar cans are placed on opposite sides of a beam balance. One contains an unknown number of pennies, the other is empty. The mass of the pennies is found by weighing to be 83.7 g.

The can with the pennies is taken off and shaken so that some pennies jump out. Pennies are also thrown at it from a distance and some go in. The mass of the pennies in the can is again found on the balance. This process is repeated three times, and the results (including the first) are: 83.7 g, 65.1 g, 93.0 g, 83.7 g, 37.2 g.

Find the probable mass of one penny, the number of pennies in the can when it was first weighed, and the number in the can when it was last weighed.

b) Explain to Mr. X how the pennies-in-can experiment in (a) resembles the Millikan oil-drop experiment, and how the Millikan experiment supports the view that electricity has a particle property. (He has not heard of the oil-drop experiment before, and he is not interested in the details of the experiment.)

Oxford and Cambridge Board O-level Physics (Nuffield) (1965).

8.8 A number of identical charged drops of known mass used in a Millikan experiment are found to balance at the following voltages applied to the plates:

125, 200, 330, 250, 500, 1 000.

Calculate a possible number of units of electric charge on each of these drops.

8.9 Electrical conductivity is affected by changes in the temperature of the material. Write a short paragraph accounting as far as you can for each of the following. (You may well have to refer to some of the references listed at the end of this unit in order to produce a suitable explanation in each case.)

a) The electrical conductivity of copper falls as the temperature rises.

b) The electrical conductivity of copper sulphate solution rises as the temperature rises.

c) The electrical conductivity of common salt is almost zero at room temperature but rises rapidly when it melts.
d) The electrical conductivity of paraffin wax, which is almost zero at room temperature, is unchanged by heating and melting.
e) The electrical conductivity of air is very low unless either the pressure is reduced, or the potential gradient is raised to 20 000 V m^{-1}.

8.10 Using the simple Drude model for an electric current, show that the charge mobility of the electrons is given by

$$\frac{\sigma}{ne},$$

where σ is the conductivity of the material, n the electron density and e the charge on the electron

Using a data book, calculate the electron mobility in copper, aluminium, silver, and sodium. Comment on your result.

8.11 The terminals of a 2 V cell are connected together by 2 m of copper wire of resistance 1 ohm m^{-1}. Calculate the electron drift speed in the direction of the applied field.

The resistivity of copper is 1.7×10^{-8} ohm m, and its density is 9000 kg m^{-3}. Assume that each copper atom contributes a single conduction electron.

8.12 A thyratron produced commercially for the demonstration of the Franck-Hertz experiment with mercury vapour is shown schematically in Fig. P8.1.

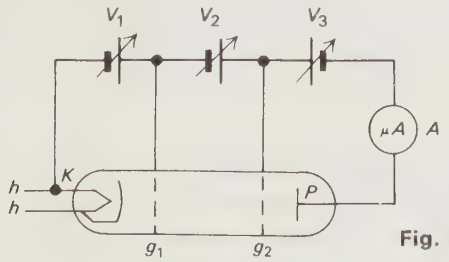


Fig. P8.1

As V_2 is raised from 0 to 20 V, the sensitive ammeter A shows three distinct current minima at intervals of 4.9 V.

a) Explain this.
b) Consider the case when V_1 is 1 V, V_2 is 10 V and V_3 is -1 V. In what region of the space in the tube can the electrons give up energies of 4.9 eV in inelastic collisions? Sketch the graph of current against V_2 in the voltage range 0 to 10 V.
c) Consider now the case when the three voltages are 1 V, 20 V, and -1 V. Sketch the graph which you would expect to obtain in this case.

8.13 Collisions between gas molecules in the air rarely cause ionization. Use the histogram of Problem 4.8 to explain why this is so.

What is the minimum relative velocity of two nitrogen molecules which, in collision, might lead to the ionization of one of them?

8.14 Calculate the speed of an electron which has been accelerated in an electron gun by a field due to a potential difference of 2000 V. You may assume that the electron is emitted with zero speed.

8.15 Calculate the energy associated with each of the following photons:

a) X-radiation at a wavelength of 0.15 nm,
b) ultra-violet radiation at 121 nm,
c) a television transmission at 200 MHz,
d) a radio transmission at 2×10^5 Hz.

8.16 A molecule of silver bromide, a light sensitive compound which is to be found in many photographic films, can be decomposed by about 1.6×10^{-19} J. What is the minimum frequency of the photon which would decompose a molecule of silver bromide?

8.17 Photographic paper of a type known as bromide paper can be handled quite safely in red light but not in white light. How do you account for this?

8.18 A simple laboratory method for obtaining almost monochromatic light is to heat a sodium salt in a bunsen flame. The yellow light emitted comprises, in fact, two discrete wavelengths at 589.0 and 589.6 nm. We shall assume a wavelength of 590 nm and that light of this wavelength is a consequence of the return of electrons from the first excited state to the ground state.

a) What is the photon energy involved?
b) What is the temperature of a gas whose molecules have an average kinetic energy equal to that of the photon energy?
c) The average gas kinetic energy in the gas flame may be taken as about 2000 K. How then is it possible for excitation to occur at all?

8.19 In an experiment with a photocell it was found that electron emission could be prevented by the application of a suitable reverse voltage. This voltage varied with the wavelength of the light used. From the results given in

Table P8.19 plot a graph of "stopping potential" against frequency and use it to determine (a) the threshold frequency and (b) the work-function of the material of the cell cathode. Find a value of Planck's constant.

Table P8.19

Wavelength (nm)	365	436	470	530	589	605
Reverse voltage (V)	2.10	1.60	1.32	1.05	0.80	0.70

8.20 A metal plate of material with a work function of about 5 eV is placed at a distance of 2 m from a weak source of monochromatic light of wavelength about 500 nm. The light output of the light source is 10^{-2} W. Assuming that light is a wave, estimate how long it would take for a single photoelectron to be emitted. You may assume that an individual photoelectron collects its energy over a circular area of the plate centred on the atom from which it will come. This area may be assumed to have a radius of about 10 atomic diameters—say 10^{-9} m. Comment on your estimate.

8.21 In 1909 G. I. Taylor devised an experiment to test whether light photons could be responsible for the diffraction patterns which are usually regarded as a characteristic feature of the wave model for light. This question is concerned with a similar situation.

a) The retina of the human eye can readily detect energy arriving at a rate of about 10^{-12} W. How many photons per second does the retina receive under these conditions if the wavelength of the light used is 600 nm.
b) On average, how far apart will these photons be?
c) Light of this intensity nevertheless produces clear diffraction effects. State the problem which G. I. Taylor was considering.

8.22 According to the wave theory the intensity of the light emitted from a point source varies inversely as the square of the distance from the source. If photons are emitted in random directions from a point source of radiation, show that the number passing through unit area (i.e., the intensity) must also be governed by an inverse square law.

8.23 The spectrum of singly ionized helium is very similar to that of hydrogen. It includes lines with wavelengths of 303, 256, 243, and 237 nm. Compare this series with the corresponding series for hydrogen. If the ionization energy for hydrogen is 21.8×10^{-19} J, how much energy would you expect to supply if you wished completely to ionize this helium?

BIBLIOGRAPHY: Unit Eight

Boorse, H. A., and Motz, L. (1966). *The World of the Atom*, Basic Books.
Volume 1 includes translations of the papers by Franck and Hertz and of Balmer.

Dobson, G. M. B. (1968). *Exploring the Atmosphere*, Oxford University Press.
Includes an interesting account of the interaction between in-coming radiation and the atmosphere.

Feather, N. (1968). *Electricity and Matter*, Edinburgh University Press.
Chapters 4 to 7 are concerned with the behaviour of electrons in matter and provide an extensive account of the subject matter of this unit. The historical development of the electron model makes particularly good reading.

Goodier, J., and Meynell, J. W. (1964). *Electricity*, Mills and Boon.
Many experiments which form the basis of our understanding of the conduction of electricity through electrolytes are described in this book.

Ingram, D. J. E. (1966). "Lasers," *Contemporary Physics*, **3**, 435-452. One of the best short accounts of laser operation to be found.

Millikan, R. A. (1917). *The Electron*, University of Chicago Press.
This is Millikan's own account of his work and contains an interesting account of the historical background as it appeared to Millikan.

(P.S.S.C.) (1968). *College Physics*, Raytheon Educational Press.
Chapters 31 and 33 deal with the photoelectric effect and with the experiments of Franck and Hertz and their outcomes.

Ziman, J. M. (1963), *Electrons in Metals*, Taylor & Francis.
This very detailed account goes far beyond the level of this book, but the development of the "free-electron gas" model of electrons in metals has influenced the development used in the unit and you may care to refer to the account which will be found in the early pages of Chapter 1.

UNIT NINE
Atomic Structure—A Case History

47 The Atomic Model
48 The Nuclear Atom
49 The Nucleus

Chapter 47
THE ATOMIC MODEL

47.1 Introduction
47.2 The early atomists
47.3 The atomic theory
47.4 The genesis of quantum ideas
47.5 X-rays
47.6 The electron—the first chip
47.7 A crisis develops—the photoelectric effect
47.8 Another predicament for the wave theory
47.9 Planck's quantum theory
47.10 Albert Einstein
47.11 Waves? Particles?

47.1 INTRODUCTION

Part of the obituary notice of Lord Rutherford in *The Times*, October 1937, read:

> "In the nineteenth century the astonishing progress of chemical science rested on the immutable character of the elements. When, therefore, the young Professor of Physics in Montreal proposed that, after all, transmutation was not only possible but was actually in continuous progress, he was clearly in need of plenty of evidence to support this most unorthodox hypothesis."

Rutherford made the investigation of the structure of the atom his life's work. In doing this he was aided by the discoveries of other scientists before him and those contemporary with him. These discoveries formed a vast pool of knowledge about the atom. As this knowledge grew, the picture of the atom that emerged demanded hypotheses about the nature of matter and light that were even more unorthodox than was mere transmutation. Scientists were forced to the conclusion that classical physics, i.e., the physics of Newton, of electromagnetics and thermodynamics, had to be modified when it was applied to the internal structure of atoms.

The aim of this Unit is to trace the development of our ideas about the atom. This will include the early ideas of the Greeks, the work of chemists in the eighteenth and nineteenth centuries, the results of spectroscopists, the discovery of the electron and of radioactivity, and the development of successive models of the atom. While it is impossible at this level to treat all the ideas from first principles, it is hoped to show that at least the resulting formulae and ideas are reasonable and founded on experimental data.

47.2 THE EARLY ATOMISTS

In the era 500 B.C. to 50 B.C. the Greek philosophers, or lovers of wisdom, sought to discover rational explanations of natural events that did not involve myths and demons. During these years such men as Leucippus, his pupil Democritus, Epicurus, and

Lucretius developed a model of the atom that is extraordinarily similar to the one produced during the eighteenth and nineteenth centuries by chemists such as Lavoisier and Dalton. To the Greeks, the word *atoma* meant "an indivisible body" and they proposed that all objects consisted of large numbers of these tiny indivisible particles. Lucretius wrote:

> "Since the atoms are moving freely through the void, they must all be kept in motion either by their own mass or on occasion by the impact of another atom. For this must often happen that two of them in their course knock together and immediately bounce apart in opposite directions, a natural consequence of their hardness and solidity and the absence of anything behind to stop them. . . . Those that do not recoil far, being driven into a closer union and held there by the entanglement of their own interlocking shapes, are the atoms that compose firmly rooted rock, the stubborn strength of metal, and the like. Those others that move freely through larger tracts of space, springing far apart and carried far by the rebound, these provide for us thin air and blazing sunlight."

One is justified in wondering why, after such an excellent start, it took another 2000 years for further developments in the atomic theory to take place. There are two main reasons for the decline of the atomic theory; firstly the early atomists were unable to do any experiments to support their theory although they were able to explain many natural phenomena, and secondly, the theory was criticized strongly by Aristotle (384–322 B.C.) who argued that it was impossible for a void or vacuum to exist and that ideas about atoms with continuous motion could not be contemplated. Moreover, he developed a theory based on the "elements" air, earth, fire, and water, and the "qualities" cold, hot, moist, and dry. With this theory, Aristotle could also explain many natural phenomena to the satisfaction of the people of his time for whom it had the advantage that it was based on experiences that were familiar, and it did not involve the use of atoms which could not be observed, or empty space which was difficult to imagine.

Aristotle's authority was immense. He was a pupil of Plato and a tutor of Alexander the Great. He developed the system of logic that we use today—starting from an agreed set of assumptions, and arguing through a series of "If . . . then . . . " steps to a definite conclusion. This system had an immediate appeal in that it seemed to provide answers to some of the puzzles of nature. Also, the people enjoyed the security given by a decisive conclusion. Even today we prefer to feel secure by knowing a fact. It is uncomfortable to be told that light may behave as a wave and that it may behave as a particle.

The scientific treatises of Aristotle were copied, translated, and quoted for centuries. Right up to the time of Galileo in the seventeenth century their authority remained supreme. A further reason for the decline of the atomic theory of matter was that its ideas were considered atheistic and materialistic. Pierre Gassendi (1592–1655), a Priest and philosopher, countered these criticisms by pointing out that it was God who created the atoms and it was He who gave them their motion. The Atomic Age really began in the seventeenth century when Galileo (1564–1642) and Newton (1642–1727) showed that some of Aristotle's ideas about motion were based on false assumptions that had not been tested experimentally. The way was now clear for ideas about the atom to develop, and both Newton and Boyle (1627–1691) described models of a gas involving "particles." (See Unit 1.)

When he was 34, Boyle wrote *The Sceptical Chymist* and in it laid some of the foundations of modern chemistry. His description of an element was very different from that of Aristotle. Boyle wrote that elements were

> ". . . certain primitive and simple, or perfectly unmingled bodies; which not being made of any other bodies, or of one another, are the ingredients of which all . . . perfectly mixed bodies are immediately compounded . . ."

In the eighteenth century, science became recognized and respected. The need to argue from experimental facts was realized and scientists set up laboratories so that they could see for themselves

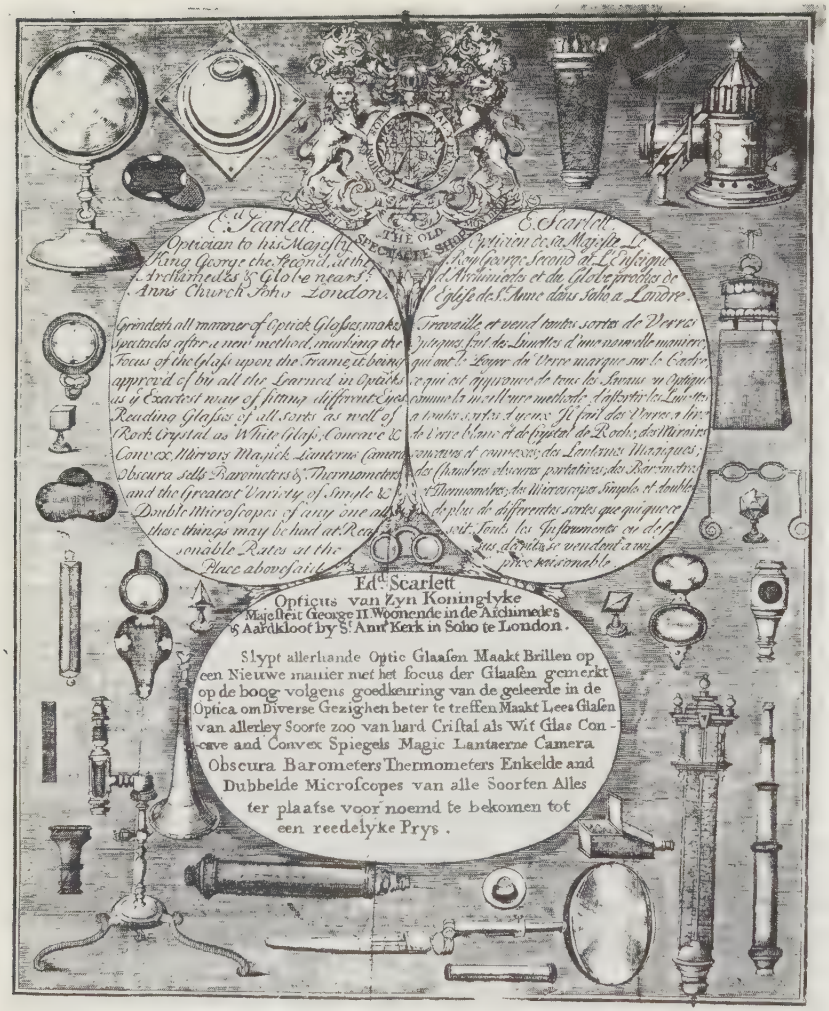


Fig. 47.1 The trade-card of Edward Scarlett, optician to King George II. (Reproduced by permission of the Science Museum, London.)

what really happened instead of just thinking about what ought to happen. Their method of working became more and more quantitative; they measured, weighed, and timed. To have an interest in science was fashionable and many people bought the apparatus that was then available—mirrors and lenses, microscopes, thermometers, balances, air pumps, etc. (See Fig. 47.1.)

Consequently, many advances were made, especially in chemistry. Hales, an English clergyman, invented the technique of collecting gases over water; Black, a Scottish chemist, discovered carbon dioxide; Cavendish, a shy eccentric aristocrat, discovered hydrogen. Scheele in Sweden and Priestley, a Unitarian minister, discovered a number of gases including nitric oxide and oxygen. In France, Lavoisier explained combustion and the composition of water. He arranged the facts of chemistry in a coherent and logical pattern and established the law of conservation of mass. His last book, published in 1789, listed 23 elements. In spite of the fact that his work was comparable in importance to that of Newton, he was,

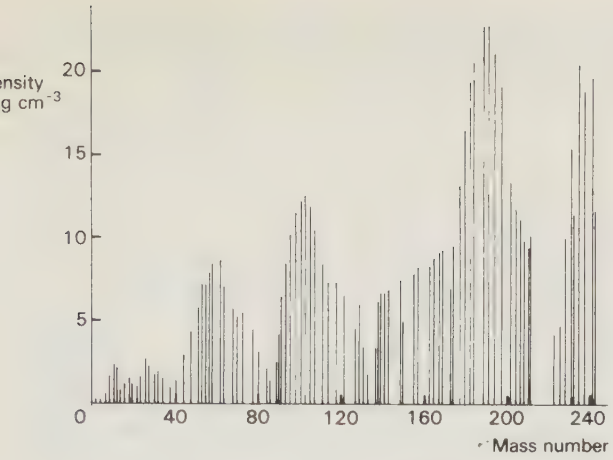


Fig. 47.2 Densities of the elements arranged by mass numbers.

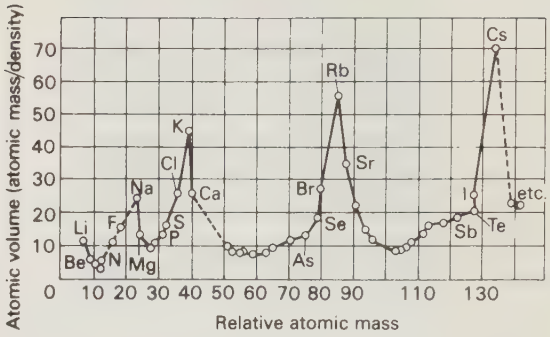


Fig. 47.3 An example of periodicity in the sequence of elements (after Lothar Meyer, but using modern values.) (From G. Holton and D. Roller, *Foundations of Modern Physical Science*, Addison-Wesley, 1958.)

as an aristocrat, a victim of the French Revolution and was sent to the guillotine in 1794 with the words "The Republic has no need of men of science."

47.3 THE ATOMIC THEORY

Dalton (1766–1844) was a village schoolmaster who eventually became a professor at Manchester. He devoted all his energies to science and remained a bachelor "never having had time to marry." He made new assumptions about atoms and explained the facts of chemical combination. In his atomic theory he proposed:

a) Matter consists of indivisible atoms.
b) An element is composed of identical atoms which are different from those of another element.
c) An element can be distinguished by the mass of its atom.
d) The masses of all atoms should be compared to that of the lightest known atom, i.e., hydrogen, which was arbitrarily assigned a mass of unity.
e) Chemical compounds are composed of atoms of elements combined in definite proportions and that the smallest quantity was a *compound atom*—now called a molecule.

Thus, in the first part of the nineteenth century, there was plenty of evidence to support an atomic model, and it was clear that quantitative relationships existed between the atoms of elements.

Element-hunting had become a popular scientific activity and by 1860 a total of 60 elements had been tracked down. Once an element had been found such attributes as its atomic mass, melting and boiling points, its density, the thermal and electrical conductivities were measured. One attribute of particular importance to chemists was the valency or combining capacity of an element to form compounds with other elements. With such a chaos of facts to hand, it must have seemed unlikely that any order or inter-relationships between elements could exist. However, if such a physical constant as density, measured under the same conditions of temperature and pressure is plotted against the mass numbers of the elements, a pattern emerges (see Fig. 47.2). This shows a regular, or periodic, rise and fall of density as one goes through the list of elements in order of mass number. Other properties show similar periodic changes. Figure 47.3 demonstrates the periodicity of atomic volume which is calculated by dividing the atomic mass of an element by its density in the liquid or solid state.

The atomic masses, based on the exact number 12.00000 as the assigned atomic mass of the principal isotope of carbon. ^{12}C, are the most recent (1961) values adopted by the International Union of Pure and Applied Chemistry. The unit of mass used in this table is called *atomic mass unit* (u): 1 u = 1.6604×10^{-27} kg. The atomic mass of carbon is 12.01115 on this scale because it is the average of the different isotopes naturally present in carbon. (For artificially produced elements, the approximate atomic mass of the most stable isotope is given in brackets.)

Group→		a I b	a II b	a III b	a IV b	a V b	a VI b	a VII b	VIII			0
Period 1	Series 1	1 H 1.00797										2 He 4.0026
2	2	3 Li 6.939	4 Be 9.0122	5 B 10.811	6 C 12.01115	7 N 14.0067	8 O 15.9994	9 F 18.9984				10 Ne 20.183
3	3	11 Na 22.9898	12 Mg 24.312	13 Al 26.9815	14 Si 28.086	15 P 30.9738	16 S 32.064	17 Cl 35.453				18 Ar 39.948
4	4	19 K 39.102	20 Ca 40.08	21 Sc 44.956	22 Ti 47.90	23 V 50.942	24 Cr 51.996	25 Mn 54.9380	26 Fe 55.847	27 Co 58.9332	28 Ni 58.71	
	5	29 Cu 63.54	30 Zn 65.37	31 Ga 69.72	32 Ge 72.59	33 As 74.9216	34 Se 78.96	35 Br 79.909				36 Kr 83.80
5	6	37 Rb 85.47	38 Sr 87.62	39 Y 88.905	40 Zr 91.22	41 Nb 92.906	42 Mo 95.94	43 Tc [99]	44 Ru 101.07	45 Rh 102.905	46 Pd 106.4	
	7	47 Ag 107.870	48 Cd 112.40	49 In 114.82	50 Sn 118.69	51 Sb 121.75	52 Te 127.60	53 I 126.9044				54 Xe 131 30
6	8	55 Cs 132.905	56 Ba 137.34	57–71 Lanthanide series*	72 Hf 178.49	73 Ta 180.948	74 W 183 85	75 Re 186.2	75 Os 190.2	77 Ir 192.2	78 Pt 195.09	
	9	79 Au 196.967	80 Hg 200.59	81 Tl 204.37	82 Pb 207.19	83 Bi 208.980	84 Po [210]	85 At [210]				86 Rn [222]
7	10	87 Fr [223]	88 Ra [226.05]	89—Actinide series*								

* Lanthanide series:	57 La 138.91	58 Ce 140.12	59 Pr 140.907	60 Nd 144.24	61 Pm [147]	62 Sm 150.35	63 Eu 151.96	64 Gd 157.25	65 Tb 158.924	66 Dy 162.50	67 Ho 164.930	68 Er 167.26	69 Tm 168.934	70 Yb 173.04	71 Lu 174.97
** Actinide series:	89 Ac [227]	90 Th 232.038	91 Pa [213]	92 U 238.03	93 Np [237]	94 Pu [242]	95 Am [243]	96 Cm [245]	97 Bk [249]	98 Cf [249]	99 Es [253]	100 Fm [255]	101 Md [256]	102 No	103

Fig. 47.4 Periodic table of the elements.

A young Russian professor at St. Petersburg (Leningrad), Dmitri Mendeleyev, sought to systematize the facts of chemistry for a book he was writing in 1868. In his laboratory, Mendeleyev pinned cards on the wall, one per element. On these he noted the mass number and the chemical and physical properties of each element and its compounds. He was struck by the way properties were repeated periodically as one went up the mass number scale. After much rearrangement of his cards, he devised a system of rows and columns in which mass numbers increased from left to right along a row and elements with similar properties were in vertical columns. For example:

Li(7)	Be(9)	B(11)	C(12)	N(14)	O(16)	F(19)
Na(23)	Mg(24)	Al(27)	Si(26)	P(31)	S(32)	Cl(35)
K(39)	Ca(40)	—	Ti(48)	V(51)	etc.	

The first column consists of the alkali metals. These are all soft, have low melting points and low densities —indeed Li, Na, and K, float on water, and if this is tried it will become clear that they are highly reactive, and form an alkali. The second column forms a family known as the alkaline earths. These have densities and melting points higher than those of the alkali metals, and they are much less reactive. Another family with clearly defined properties is in the last column. These, together with bromine and iodine are the halogens and all combine violently with many metals to form white salts—the word halogen means "salt former." They are distinctly non-metallic, poisonous, corrosive, and are potent anti-bacterial agents—the use of chlorine in swimming pools and the process of fluoridation to reduce tooth decay caused by bacteria is well known.

To Mendeleyev the next known element after calcium was titanium but he left a gap below aluminium since titanium had properties more closely resembling carbon and silicon than boron and aluminium. He predicted that this gap would be filled by an undiscovered element. Several gaps like this

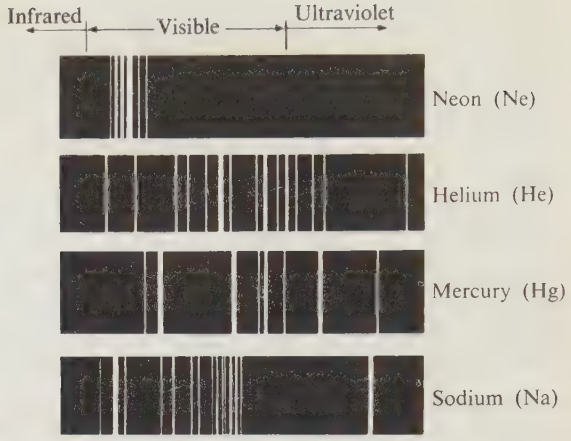


Fig. 47.5 Portions of the bright line spectra of neon, helium, mercury, and sodium (re-sketched from spectrographic records). Note that the lines vary markedly in intensity. (From A. B. Arons, *Development of the Concepts of Physics*, Addison-Wesley, 1965.)

appeared in his periodic table of the elements and such was his faith in periodicity that he predicted with remarkable accuracy the physical and chemical properties of the elements that would eventually fill those gaps.

A modern version of the periodic table is shown in Fig. 47.4. Titanium is no longer under silicon, but heads a column (IVa) of its own, the gap between it and calcium being now filled by scandium. The column on the right contains the rare gases or inert gases. These elements do not readily react with other elements and it was not until 1963 that unstable compounds of xenon and krypton were produced. The elements are here arranged in order of atomic number rather than atomic mass.

As with all advances in science, the periodic table produced more questions than answers. What is it that makes atoms so different, element from element? Why do the elements have such varied properties? Are atoms really indivisible or can they be subdivided into smaller pieces?

The study of matter was originally the province of chemists and it was they who established the idea that matter was made up of atoms. In the twentieth century the investigation of matter has fallen within the field of the physicist. The challenge was to provide a physical explanation of the entire atomic system and to probe the structure of the atom itself. The work has been fruitful—and has contributed to the stock of questions awaiting an answer.

47.4 THE GENESIS OF QUANTUM IDEAS

The study of the spectrum of light from the sun and white hot metals had been performed by Newton, and the resulting continuous band of colours is familiar. With the steadily increasing number of elements grew an interest in the coloured light that each produced when it was vaporized in a flame. Instead of the continuous spectrum associated with white hot solids and liquids, these hot vapours produced a series of coloured lines or bands. As we have seen in Chapter 46, gases at low pressure, made to glow in a tube by using an electrical discharge, also gave a line spectrum. It was soon noted that the positions or wavelengths of a set of lines was a unique characteristic of the element used; a line spectrum is a "fingerprint" which can be used to identify an element. Figure 47.5 shows some typical line spectra.

Spectroscopy, as the study of spectra is called, was established by R. W. Bunsen and G. R. Kirchhoff who provided its major tool—the spectroscope. They gave the first explanation of the Fraunhofer lines in the spectrum of the sun and also discovered the elements rubidium and caesium. Of course, this interest in spectra caused men to wonder why an

element should produce a line spectrum and they (as Mendeleyev had done for chemical properties) sought to find a regular pattern or system which, while not necessarily explaining a reason for the lines, at least would provide a formula for calculating their wavelengths.

As we have seen in Unit 8, J. J. Balmer succeeded in doing this for the hydrogen spectrum in 1885. Others, notably J. Rydberg, who was able to extend Balmer's work so that his form of the equation could describe series of lines in the spectra of other elements, and W. Ritz (1908), who expressed the equation in an alternative form, continued this work.

However useful these empirical formulae were, particularly in predicting new spectral lines and series, and in suggesting where on the wavelength scale one should look for them, they provided no clue to the basic physical mechanism. Perhaps Balmer's greatest contribution to science was that his work showed that a pattern existed. It resulted, not in answers, but in the question "What is the cause of this regular pattern?" Mendeleyev had bequeathed a similar pattern. Any theory or model of the structure of the atom must be able to account for these patterns and regularities.

We have also seen how the late nineteenth-century physicists were interested in discharge tubes and how two models were advanced to account for cathode rays. One of these was a wave model, the other a particle model. The latter gained the day (although surprises were to come later).

This considerable interest in cathode rays resulted directly in three major discoveries; discoveries that led to the recognition of the first "chip" to come from the uncuttable atom, to the transmutation of immutable elements and, eventually, to our present model for the structure of matter. X-rays were chanced upon on 8th November, 1895, to be followed in a few months by the discovery of radioactivity and of the electron.

No one then realized that the following decade would bring a flood of fresh knowledge and the discovery of unexpected phenomena. The age of the quantum and of relativity was near. A crisis was to face the scientific world; when dealing with events on a subatomic scale some of the concepts of classical physics that had been developed so carefully since Galileo and Newton would have to be modified, at least in part.

47.5 X-RAYS

The production of X-rays today has been covered in Unit 8. This penetrating radiation was discovered by accident when W. C. Röntgen in Würzberg was doing research on the nature of the cathode rays. He was repeating some of the experiments of Hertz and Lenard and had enclosed his discharge tube completely with black card to absorb the visible light of the discharge. A translation of his paper was published in *Nature* on January 23rd, 1896. It began:

> "A discharge from a large induction coil is passed through a Hittorf's vacuum tube, or through a well-exhausted Crookes' or Lenard's tube. The tube is surrounded by a fairly close-fitting shield of black paper; it is then possible to see, in a completely darkened room, that paper covered on one side with barium platinocyanide lights up with brilliant fluorescence when brought into the neighbourhood of the tube, whether the painted side or the other be turned towards the tube. The fluorescence is still visible at two metres distance. It is easy to show that the origin of the fluorescence lies within the vacuum tube.
>
> It is seen, therefore, that some agent is capable of penetrating black cardboard which is quite opaque to ultra-violet light, sunlight, or arc-light. It is therefore of interest to investigate how far other bodies can be penetrated by the same agent. It is readily shown that all bodies possess this same transparency, but in very varying degrees. For example, paper is very transparent; the fluorescent screen will light up when placed behind a book of a thousand pages; printer's ink offers no marked resistance. Similarly the fluorescence shows behind two packs of cards; a single card does not visibly diminish the brilliancy of the light. So, again, a single thickness of

tinfoil hardly casts a shadow on the screen; several have to be superposed to produce a marked effect. Thick blocks of wood are still transparent. Boards of pine two or three centimetres thick absorb only very little. A piece of sheet aluminium, 15 mm thick, still allowed the X-rays (as I will call the rays, for the sake of brevity) to pass, but greatly reduced the fluorescence. Glass plates of similar thickness behave similarly; lead glass is, however, much more opaque than glass free from lead. Ebonite several centimetres thick is transparent. If the hand be held before the fluorescent screen, the shadow shows the bones darkly, with only faint outlines of the surrounding tissues."

It was inevitable that X-rays should be discovered since so much work was going on with discharge tubes. Hertz and Lenard must have narrowly missed the discovery themselves; and it is even reported that an English physicist, Frederick Smith, was finding that photographic plates, kept in a drawer near his tube, were continually being fogged. Instead of pursuing the cause of the trouble he merely kept his plates elsewhere and continued his original work!

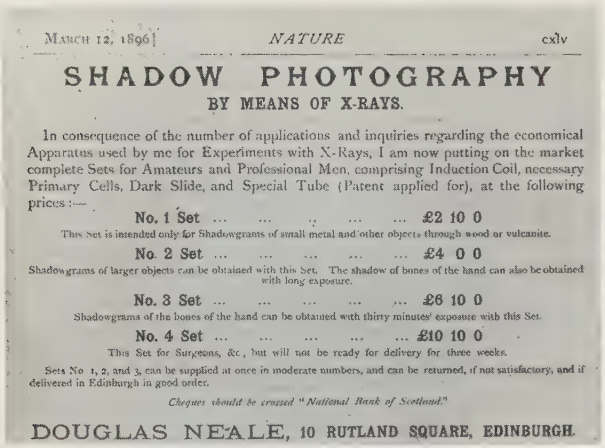
Röntgen, seeing the significance of his original observation, abandoned his other research and, in a matter of seven weeks, discovered almost all the properties of X-rays known to us today. As we have seen, his work was quickly translated and published in scientific journals all over the world. The fascination of the new discovery and especially of the ability it provided to produce shadow photographs of opaque objects captured the imagination of the scientific community. This is hardly surprising for later in his paper Röntgen wrote

> "The justification of the term 'rays,' applied to the phenomena, lies partly in the regular shadow pictures produced by the interposition of a more or less permeable body between the source and a photographic plate or fluorescent screen.
> I have observed and photographed many such shadow pictures. Thus, I have an outline of part of a door covered with lead paint; the image was produced by placing the discharge-tube on one side of the door, and the sensitive plate on the other. I have also a shadow of the bones of the hand (Fig. 47.6), of a wire wound upon a bobbin, of a set of weights in a box, of a compass card and needle completely enclosed in a metal case, of a piece of metal where the X-rays show the want of homogeneity, and of other things.
> For the rectilinear propagation of the rays, I have a pin-hole photograph of the discharge apparatus covered with black paper. It is faint but unmistakable." (Röntgen, 1896.)



Fig. 47.6 Röntgen's photograph of the bones in the finger of a human hand. (From *Nature*, January 23, 1896. Reproduced by permission of the publisher.)

It had never been so easy to try out such a major advance for the necessary Crookes tubes existed in laboratories throughout the world. Within weeks of the announcement, journals were filled with reports of similar experiments and photographs of bones, etc. The medical profession were swift to put X-rays to practical use and apparatus manufacturers jumped on the bandwaggon as Fig. 47.7 shows. What this last figure says about exposure time would horrify any hospital radiologist today. One property of X-rays,



MARCH 12, 1896] NATURE cxlv

SHADOW PHOTOGRAPHY

BY MEANS OF X-RAYS.

In consequence of the number of applications and inquiries regarding the economical Apparatus used by me for Experiments with X-Rays, I am now putting on the market complete Sets for Amateurs and Professional Men, comprising Induction Coil, necessary Primary Cells, Dark Slide, and Special Tube (Patent applied for), at the following prices:—

No. 1 Set £2 10 0

This Set is intended only for Shadowgrams of small metal and other objects through wood or vulcanite.

No. 2 Set £4 0 0

Shadowgrams of larger objects can be obtained with this Set. The shadow of bones of the hand can also be obtained with long exposure.

No. 3 Set £6 10 0

Shadowgrams of the bones of the hand can be obtained with thirty minutes' exposure with this Set.

No. 4 Set £10 10 0

This Set for Surgeons, &c., but will not be ready for delivery for three weeks.

Sets No 1, 2, and 3, can be supplied at once in moderate numbers, and can be returned, if not satisfactory, and if delivered in Edinburgh in good order.

Cheques should be crossed "National Bank of Scotland."

DOUGLAS NEALE, 10 RUTLAND SQUARE, EDINBURGH.

Fig. 47.7 An advertisement from *Nature* for March 12, 1896. (Reproduced by permission of the publisher.)

that only time could discover, was that they could destroy the structure of tissue cells in flesh, bone and blood and produce severe lesions which, like burns, were very difficult to heal. Even worse, genetic aberrations may be caused by the ravaging of reproductive cells by X-rays. No one realized in those early days how dangerous X-rays could be and men unwittingly exposed themselves to maiming and death. Edison's assistant, Clarence Dally, died in 1905 from overexposure resulting from several years of X-ray experiments.

Röntgen's work was acclaimed internationally, but it raised the inevitable question of the nature of the rays. Being undeflected by a magnet they were not charged particles; yet they did not behave like any known electromagnetic wave since they were so penetrating and could not be refracted or reflected. This puzzle remained until 1912 after further discoveries, and the establishment of the quantum theory.

47.6 THE ELECTRON—THE FIRST CHIP

The X-rays did, however, help to solve the question of the nature of the cathode rays. The dispute concerning the appropriate model for these rays has already been discussed. That the "wave model" school could not accept a particle model was due largely to an experiment which failed. Hertz argued that, if the rays were charged particles, then it must be possible to deflect a beam of them with an electric field. He put two parallel plates, connected to a battery, first outside and then inside the discharge tube. No deflection resulted and Hertz thereby concluded that the rays could not be made up of particles.

Then, in February 1896 the ionization of gases exposed to X-rays was reported by both Röntgen, and J. J. Thomson at the Cavendish laboratory in Cambridge. They had found that gases became conducting and allowed electroscopes to discharge. This was highly significant in two ways. Here was evidence that the atom was itself a complex structure involving electrical charges, since atoms of elements like helium and argon could be separated into positive and negative parts. Secondly, the presence of ions in Hertz's discharge tube could explain his failure to deflect cathode rays with an electric field.

Thomson reasoned that the gas pressure in Hertz's tube was not low enough and the cathode rays caused so much ionization that, to use his own words,

> ". . . the positive and negative ions between the plates would slowly diffuse, until the positive plate became coated with negative ions, the negative plate with positive ones; thus the electric intensity between the plates would vanish and the cathode rays be free from electrostatic force."

Accordingly, Thomson had a tube made with parallel plates inside, used a better vacuum pump, and successfully deflected cathode rays.

> ". . . the rays were depressed when the upper plate was connected with the negative pole of the battery, the lower with the positive, and raised when the upper plate was connected with the positive, the lower with the negative pole. The deflexion was proportional to the difference of potential between the plates, and I could detect the deflexion when the potential difference was as

small as two volts. It was only when the vacuum was a good one that the deflexion took place . . ." (Thomson, 1897.)

Thus Thomson had performed a key experiment showing that cathode rays were streams of negative particles.

The story is told of Thomson that he was not a particularly dexterous experimenter and Mr. E. Everett, his laboratory assistant, had to protect the delicate glass tubes he had made from being handled by their designer. However, Thomson's ability to grasp a problem and conceive experimental tests was undoubted. He next assailed the question that arose—if these are charged particles, what is their mass? To throw some light on this, he set about measuring their specific charge.

The principle of his experiment for measuring e/m_e, the charge to mass ratio of the particles, has been covered in Chapter 43. It will be sufficient here to outline the significance of his results. For a range of different gases, hydrogen, air, carbon dioxide, used at various low pressures, he found substantially the same result for e/m_e in every case—values having the same order as the modern result, 1.76×10^{11} C kg^{-1}. Nor did changing the cathode material from aluminium to iron and then to platinum produce any effect on the value of e/m_e. Consequently, Thomson reasoned that the particles forming cathode rays were a basic part of all matter. Moreover, he realized that a knowledge of the mass of the particles would finally decide whether they were whole charged atoms or parts of atoms.

Thomson knew from experiments on the electrolysis of water that 96 500 coulombs of electricity would liberate 1 g of hydrogen. It had been suggested in 1874 that if the atomic theory is correct, the (charge/mass) ratio for single ions in electrolysis must be equal to the (total charge/total mass) ratio for a bulk of ions. Thus for hydrogen ions, (charge/mass) $= 9.6 \times 10^7$ C kg^{-1}.

This was the largest (charge/mass) ratio known until Thomson's work with cathode rays and he had now found a particle with a ratio nearly 2000 times greater. Was this huge value due to the largeness of e or the smallness of m_e? If it was the latter, then a particle had been found having a mass 1/2000 of the mass of the lightest known atom.

Thomson favoured the smallness of the particles, for Hertz and Lenard had shown that they would pass through thin aluminium and about 1 cm of air. He reasoned that this would only be possible with a particle that was small compared with ordinary molecules.

In Thomson's paper announcing the discovery of the electron in 1897, there were no headlines or underlined statements, just a paragraph almost concealed within a mass of print. (Thomson, 1897.)

> "Thus on this view we have in the cathode rays matter in a new state, a state in which the subdivision of matter is carried very much further than in the ordinary gaseous state: a state in which all matter is of one and the same kind; this matter being the substance of which all the chemical elements are built up."

Work on the electron continued at the Cavendish laboratory and a method was devised of causing tiny water droplets to condense on the negative ions (the electrons) produced by X-rays (see Unit 8). Knowing the size of a droplet and the total mass of the vapour cloud, the number of droplets, and therefore ions, follows. If the total charge carried by the cloud is measured the charge on one ion is calculable—hence the charge on the electron. This is all easier said than done but despite the experimental difficulties, Thomson obtained a value for $e = 2.1 \times 10^{-19}$ C. The importance of this was that the value is in the same order of magnitude as the charge on a hydrogen ion. No doubts remained that the high value of e/m_e for cathode ray particles was due to a small value of m_e.

Later work on the charge on the electron by Millikan is described in Unit 8 and the accepted value of e is now 1.6×10^{-19} C.

To confirm that electrons are constituents of all matter, Thomson used electrons produced both by incandescent filaments and by the photoelectric effect, measured e/m_e, and in all cases obtained similar results.

47.7 A CRISIS DEVELOPS—THE PHOTOELECTRIC EFFECT

As we have seen in Unit 8, at the beginning of the twentieth century P. E. A. von Lenard made a systematic study of the photoelectric effect which had been discovered accidentally by H. Hertz in 1887 whilst working on the electromagnetic waves produced by his spark transmitter. A great surprise at that time was that there was emission of photoelectrons if, and only if, the frequency of the light was a little above a threshold frequency even though the light was dim and that raising the intensity of the light did *not* increase the maximum energy of those electrons but only the number emitted. Moreover, it appears that this emission takes place without any delay. This latter point highlights the problem and is well illustrated by a calculation.

Consider a cube of zinc of side 2 cm. This has a volume of 8 cm^3 (8×10^{-6} m^3) and a mass of 64 g, if the density is taken as approximately 8 g cm^{-3}.

The mass number of zinc is 65.4 and so the 2 cm cube contains about one mole of zinc atoms, or 6×10^{23} atoms.

One atom of zinc occupies a volume of

$$\frac{8 \times 10^{-6}}{6 \times 10^{23}} \text{ m}^3 = 13.3 \times 10^{-30} \text{ m}^3.$$

One atom can be contained in a cube of side $(13.3 \times 10^{-30})^{1/3} = 2.4 \times 10^{-10}$ m.

Consequently, if zinc atoms are laid side by side, then in a length of 1 m there would be

$$\frac{1}{2.4 \times 10^{-10}} \approx 4 \times 10^9 \text{ atoms.}$$

Thus a surface 1 m square would contain 16×10^{18} atoms.

A classical physicist would argue that light energy falling on a surface is shared equally among all the atoms in the surface, assuming that the light intensity is constant over that surface.

The rate at which energy falls on a surface can be estimated by considering a torch bulb, placed at the centre of a sphere of radius 1 metre, operating at about 3 V and 0.2 A. If all the electrical energy supplied is converted to light, then the rate at which energy is delivered to 1 m^2 of surface

$$= \frac{\text{power}}{\text{area}} = \frac{3.0 \times 0.2}{4\pi \times 1^2}$$

$$\approx 0.05 \text{ J s}^{-1} \text{ m}^{-2}.$$

Note: the light energy is likely to be less than this since much heat is produced.

0.05 J s^{-1} falls on 16×10^{18} atoms, and so one atom receives energy at the rate of

$$\frac{0.05}{16 \times 10^{18}} = 3.1 \times 10^{-21} \text{ Js}^{-1}.$$

It was known that electrons were emitted with energies up to 5×10^{-19} J, and if classical physics, and its principle of "fair shares for all atoms" held, then each atom would have to wait

$$\frac{5 \times 10^{-19}}{3.1 \times 10^{-21}} \approx 160 \text{ seconds,}$$

before it had stored enough energy to be emitted! This is very different from the 10^{-9} s that an electron requires in practice. Moreover, the classical theory in which energy is transmitted in a steady stream equally across a wavefront could not account for the threshold frequency that existed. This theory predicts that the energy transferred to electrons would depend on the intensity of the wave, and not solely on its frequency.

47.8 ANOTHER PREDICAMENT FOR THE WAVE THEORY

At about the same time, the classical physicists found themselves embarrassed when attempting to explain the distribution of energy from heated black body (or perfect) radiators. We have seen that the spectra of gases and vapours often consist of lines which indicate that the energy is being transmitted at definite wavelengths. However, a hot solid or liquid radiates a continuous spectrum and its energy is spread over a range of wavelengths as Fig. 47.8 shows. If the body is at a high enough temperature, some of the energy is in the visible range of wavelengths and the body glows.

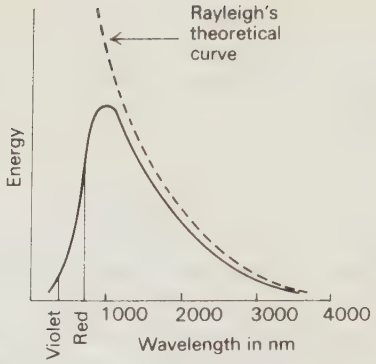


Fig. 47.8 The energy distribution in a continuous spectrum. The solid line refers to a body at 1800 K. Part of the energy is emitted in the visible range. Since there is more energy in the "red" wavelengths than in the violet, the body appears to be red hot. A higher temperature not only gives a higher peak, but shifts the peak to the left so that more "blue" energy is present and the light becomes whiter.

All theoretical attempts to derive an equation for this distribution of energy failed. Some equations fitted the facts at long wavelengths, others agreed at short wavelengths.

The German physicist, Max Planck, deduced an empirical formula that held for all wavelengths. He modestly described his work as ". . . only an interpolation formula found by happy guesswork." Now that he had the correct formula, he set about deriving it from first principles. He concluded that the classical wave theory could not be the basis and arrived, in 1901, at the revolutionary idea of quanta.

47.9 PLANCK'S QUANTUM THEORY

He proposed that energy travels in discrete units, or grains, or packets of energy. The energy carried in each quantum is proportional to the frequency of the radiation, and $E = hf$ where E = energy, f = frequency of the radiation, h = Planck's constant, $= 6.6 \times 10^{-34}$ J s. Thus energy is not radiated in a continuous stream, but in the form of separate concentrated bundles of energy rather like bullets.

Referring to Fig. 47.8, classical theories could not account for the low energies associated with short wavelengths (see the dotted line). Planck explained this by reasoning that at these high frequencies the individual quanta were very energetic, but they were not numerous and so their total energy was small. The existence of large quanta is comparatively rare, as is the chance of meeting a millionaire—a body with a large quantum of cash—in the street.

At a time when the wave theory of light had been firmly established by interference and polarization experiments, and by the electromagnetic wave theory of Maxwell, Planck's theory found little favour. However, as we have already seen, this revolutionary idea of Planck's was put to use in 1905 by one of the world's most famous theoretical physicists.

47.10 ALBERT EINSTEIN

Einstein showed no particular aptitude in his studies at school although he did enjoy learning by himself and devoted much of his time to reading mathematics and science. After graduating in 1900, when he was 21, he took a post for nine years in the Patent Office at Berne as a technical expert, third class. He enjoyed investigating the wide variety of inventions submitted for patents. It suited him well for he found it a comparatively undemanding job that left his mind fresh to develop his major ideas.

Einstein was a friend of the older Planck—when they were in Berlin in later years they often made music together, Planck at the piano and Einstein with his violin. Einstein was familiar with the quantum theory and saw in it a way of explaining the apparent paradoxes of the photoelectric effect.

The success of his application of Planck's quantum hypothesis to the solution of the problem of the

photoelectric effect (see Unit 8) had clearly strengthened the case for the quantum theory. The idea that energy came in discrete packets now had to be taken very seriously.

47.11 WAVES? PARTICLES?

If you are thinking that this quantum idea was a reversion to the particle theory of light that scientists had been systematically disposing of for the previous 200 years, then you can be certain that the physicists of 1905 thought so too. They looked unkindly on anything that "rocked the boat." This was only human. They had long enjoyed the security of knowing that light was a wave and nothing else, and all human beings are heirs to an inheritance of Aristotelian tradition—an appreciation of pure logic and a love for absolute fact.

Even Planck thought that Einstein had overstepped the mark and wrote in 1910:

> "If the photon theory were accepted, the theory of light would be thrown back for centuries to the time when the followers of Newton and Huygens fought one another on the issue of the particle *versus* the wave theory of light. All the fruits of Maxwell's great work were threatened . . . for the sake of a few still rather dubious speculations."

However, these quantum particles at least had a frequency, and so an implied wave, associated with them.

It was hard to accept the proposition that a photon can give up all its energy to an electron, no matter how far the photon had travelled—the photoelectric effect works with light from stars. This idea was as outrageous as asking someone to believe that the energy of a wave produced when a man dives into a swimming pool can all be concentrated on one swimmer at the far end so that he is shot out of the water on to the pool side! It was much more reasonable to expect light to behave like the ripples spreading out from our diver so that all the other swimmers are slightly affected by a tiny part of the expanding wavefront.

The danger that is always present when theorizing with models is to assume that what happens on a large scale will also occur on an atomic scale.

Our difficulties arise when we make dogmatic statements such as "light *is* a wave when we are dealing with lenses and it *is* a particle when it interacts with atoms." It would be better to be more cautious and say that in some circumstances light *behaves* like a wave and in atomic situations light *behaves* like a particle. After all, light is light; why should it be a wave, or a particle?

Our basic problem is that we are compelled to use properties that we can see and experience, in order to try to imagine the physical processes behind invisible atomic phenomena. We know that on a large scale, energy can only be transferred by waves or particles. Our thinking is therefore limited to these two methods.

Later in this section we shall meet further phenomena involving the wave-particle duality. The processes involved can be formulated; whether one's mind can imagine a valid picture of these processes is immaterial and perhaps will always remain beyond our grasp. J. B. S. Haldane, a former mathematician, and later a biologist, summed up the situation when he said:

> "My own suspicion is that the universe is not only queerer than we suppose, but queerer than we *can* suppose."

Chapter 48
THE NUCLEAR ATOM

48.1 The discovery of radioactivity
48.2 Subsequent developments
48.3 The structure of the atom
48.4 Scattering
48.5 Other models of the atom
48.6 Ernest Rutherford
48.7 The scattering experiments of Geiger and Marsden
48.8 The nuclear atom emerges
48.9 Angles of scattering
48.10 Further testing of the nuclear atom model
48.11 Atomic number
48.12 Models of the atom

48.1 THE DISCOVERY OF RADIOACTIVITY

We have seen that the discovery of X-rays resulted indirectly in the identification of the electron, but it also led to Becquerel's work, in France, on the fluorescence and phosphorescence produced by X-rays. Within two months a chance observation caused him to change his research to a new topic, radioactivity, or Becquerel rays as they were then known.

Lord Rutherford, who subsequently spent most of his life working on radioactivity, told the story in a public lecture shortly before his death in 1937:

> "Few of you can possibly realize the enormous sensation produced by the discovery of X-rays by Röntgen in December 1895. It interested not only the scientific man, but also the man in the street, who was excited by the idea of seeing his own inside and his bones. Every laboratory in the world took out its old Crookes tubes to produce X-rays. These old tubes of Crookes showed that cathode rays have the power of causing brilliant phosphorescence in a great number of substances, and it was also observed that X-rays appeared to come from the points which were struck by the cathode rays. This led many people to think that X-rays might be connected with phosphorescence in some way, perhaps that phosphorescent substances might emit X-rays. A number of observers on the Continent did experiments on this subject, among others Henri Becquerel of Paris. His father, a professor before him, had been very interested in phosphorescence, particularly in measuring its duration, and he had also been interested in the rather unusual properties shown by uranium compounds." (Rutherford, 1937.)

Becquerel himself continues the story:

> "A photographic plate was wrapped with two sheets of thick black paper, so thick that the plate was not clouded by exposure to the sun for a whole day. Externally, over the paper sheet, was placed a piece of the phosphorescent substance (a uranium salt), and all were exposed to the sun for

many hours. Upon developing the photographic plate I recognized the silhouette of the phosphorescent substance in black on the negative."

Becquerel then proceeded to show that the radiations could pass through sheets of glass, aluminium, or copper, placed between the photographic plate and the uranium salt. It is clear at this stage that Becquerel still believed the penetrating radiation to be due to phosphorescence induced by sunlight!

The next step was dramatic; the sun failed to shine for a few days, and natural radioactivity had been discovered. Becquerel continues:

> "Some of the preceding experiments were prepared during Wednesday the 26th and Thursday the 27th of February (1896—a leap year), and since on those days the sun appeared only intermittently, I stopped all experiments and left them in readiness by placing the wrapped plates in a drawer of a cabinet, leaving in place the uranium salts. The sun did not appear on the following days and I developed the plates on March 1st (Sunday), expecting to find only very faint images. The silhouettes appeared, on the contrary, with great intensity." (Becquerel, 1896.)

Becquerel was becoming aware that the radiations were not, after all, induced by sunlight. However, being a careful experimenter he wanted to make certain and so kept the uranium salts in an opaque box for two months. When these were tested in a darkroom on a photographic plate he found that the radiations were emitted without noticeable decrease.

You may have the impression that the development of our concepts of science has been very much a matter of luck for Hertz's discovery of the photoelectric effect, X-rays, radioactivity, and later the discovery of penicillin by Fleming were all fortuitous. In fact very little was left to fortune for, in each case, a trained and curious intellect was present. Remember, "Chance favours the prepared mind."

Often in research, an experiment seems to be proceeding well, when an unexpected result crops up. It may be that the design of the experiment is faulty; or that the hypothesis under test is at fault; or that the theory is in need of modification. In such cases a good scientist is never content to let the unexpected remain unexplained. Many discoveries of fundamental importance have been made when trying to tidy up loose ends.



Fig. 48.1 The extraction of radium from uranium ore involved radio-chemists of Madame Curie's day in the process of pipetting by mouth from open dishes of the salts. (Photograph supplied by Professor R. D. Evans, Department of Physics, M.I.T., and reproduced by permission.)

48.2 SUBSEQUENT DEVELOPMENTS

The subsequent study of the Becquerel rays by Marie Curie, her husband Pierre, and Rutherford, will not be covered in detail here. Let it be sufficient to say that the Curies discovered the radioactive elements polonium and radium, and Rutherford investigated the nature of the radiations.

As with X-rays, the dangers of this fresh discovery were initially unknown. The type of method used by Marie Curie for extracting radium from uranium ore is shown in Fig. 48.1 in which it is clear that the early workers in a new field can unwittingly be exposing themselves to hazards; Marie Curie was herself a victim.

Table 48.1

Alpha-particles	Beta-particles	Gamma-radiation
Easily absorbed by matter. e.g., 3–10 cm of air at atmospheric pressure, or 10^{-2} mm of aluminium, thick paper, etc.	Penetrate matter more easily than α-particles, and will pass through 1 m of air or 3 mm of aluminium	Very penetrating. γ-Rays will pass through a few centimetres of lead
Cause intense ionization, e.g., 10^5 ion pairs in 3–10 cm of air	Cause some ionization, e.g., about 100 ion pairs per cm of air	Cause relatively little ionization per centimetre of air
Deflected by magnetic and electric fields with difficulty	Deflected by magnetic and electric fields easily	Cannot be deflected by magnetic or electric fields
Emitted with velocities up to 10^7 m s^{-1}, or 0.1 c	Emitted with velocities up to 0.99 c	Electromagnetic radiation. Travels at c, or 3×10^8 m s^{-1}
Carry a charge $= +2e$	Carry a charge $= -e$	Carries no charge
Mass = 4 × mass of hydrogen atom	Mass = 1/1840 × mass of hydrogen atom = mass of electron	Has no mass as ordinarily understood
Are helium nuclei	Are electrons	Are electromagnetic waves of very short wavelength—shorter than X-rays

By 1909, research, mostly guided by Rutherford, had established the following facts: Three types of radiation can be emitted from radioactive substances, alpha- (α) particles; beta- (β) particles; gamma- (γ) rays. The properties of each of these are listed in Table 48.1.

It was also known that when an atom of a radioactive element decays, it changes into a different element according to the following *displacement* laws.

a) When the atom disintegrates by the emission of an alpha-particle, it turns into an element with chemical properties similar to those of an element two places lower in the periodic table. For example, if radium decays, it does so by alpha emission and becomes radon.

b) When an atom disintegrates by the emission of a beta-particle it turns into an element with chemical properties similar to those of an element one place higher in the periodic table. For example, if an atom of a radioactive type of lead decays, it emits a beta-particle and the resulting atom is bismuth.

48.3 THE STRUCTURE OF THE ATOM

The discovery of radioactivity provided the alpha-particle—the tool with which Rutherford was able to probe the structure of the atom.

An atom is normally uncharged and had been shown to consist partly of light negative particles, the electrons. It was therefore reasoned that the other constituent of an atom was a number of heavier positive charges so that the resultant charge on the atom would be neutral. The question of how the positive and negative charges were distributed arose.

Thomson had thought that the electrons were embedded in a sphere of positive charge like negative raisins in a positive pudding. He wrote:

> "In default of exact knowledge of the nature of the way in which positive electricity occurs in the atom, we shall consider a case in which the positive electricity is distributed in the way most amenable to mathematical calculations; i.e., when it occurs as a sphere of uniform density through-

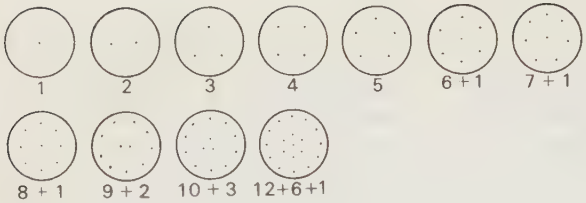


Fig. 48.2 J. J. Thomson's model of the atom. Each circle represents a sphere of positive charge and each dot an electron.

out which the corpuscles (electrons) are distributed." (Thomson, 1907.)

He developed this idea, which had also been suggested by Lord Kelvin, and showed that such an atom would be stable if the electrons were arranged in regular patterns as shown in Fig. 48.2. For example, element 3 is stable because the mutual repulsion between electrons balances the attraction that each one experiences towards the centre of the positive sphere.

Thomson's calculations showed that a ring of 6 electrons would only be stable if an additional one was placed at the centre. An atom with a ring of 10 required an additional 3 for stability; and one with an outer ring of 12 needed an inner ring of 6 plus one electron in the very centre. He noted that certain patterns repeated themselves (e.g., a triangle appears in 3 and 10 + 3 above) and he attempted to relate this periodicity with the periodic table. He also calculated the frequency of vibration of electrons about their equilibrium positions and showed this to be in the order of the frequency of light. He reasoned that an element could be caused to emit light by means of a flame or an electrical discharge because these could set electrons oscillating and emitting electromagnetic radiation. This was a promising idea but it did not predict the line spectra of gases.

Thomson could explain radioactive disintegration too; certain of the arrangements of electrons were only stable if the rings were rotating at a rate that was above a certain critical angular velocity. Radiation of energy from the moving electrons would cause their angular velocity to diminish very slowly, until after a long interval the critical velocity was reached, instability would occur, and what amounts to an explosion would take place during which a part of the atom would be ejected so that what remained would be likely to be stable.

Thus in the first decade of the twentieth century the "raisin pudding" theory seemed quite plausible when it was used to explain phenomena qualitatively; what Thomson needed was some quantitative support from experiments.

48.4 SCATTERING

If a well defined jet of water is directed towards some wire gauze, the jet emerging will not be so well defined—the water will be partly spread out or scattered.

Thomson had already noted examples of scattering in other branches of physics, for example, when a beam of electrons is passed through a thin aluminium foil; when light passes through a gas; when X-rays pass through gases and light elements such as aluminium. These observations had led him to develop a theory of scattering and thence to deduce that the number of electrons in an atom was in the same *order* as its mass number. In fact, the number of electrons is more nearly equal to half the mass number. This value was obtained in 1911 by the English physicist Charles Barkla using the technique of X-ray scattering, by light elements.

Clearly, the atom was yielding some of its secrets, and to provide further avenues for experimentation, Thomson produced a theoretical analysis of the scattering that would result when α- and β-particles passed through thin foils of metal. He assumed the metal atoms had a "raisin pudding" structure and

showed, in particular, that an α-particle would suffer very little scattering from its original path. His model of the atom would gain support if this happened in practice.

48.5 OTHER MODELS OF THE ATOM

In the years from 1900, Cambridge was not the only place where ideas about the structure of the atom were actively discussed. Philipp Lenard, a Heidelberg University physicist, proposed that an atom is made up of pairs of positive and negative charges, called dynamids, that were distributed throughout the volume of the atom. His own observations that electron beams could traverse thin foils of aluminium with little scattering led him to believe that dynamids were exceedingly small and he estimated their radius to be less than 3×10^{-14} m. He deduced that most of an atom's volume consisted of empty space since its radius was about 10^{-10} m.

In 1903, Hantaro Nagaoka read a paper before the Physico-Mathematical Society of Tokyo describing his Saturnian model of the atom. This had been inspired by a mathematical analysis of the stability of Saturn's rings, written nearly 50 years earlier by James Clerk Maxwell, who incidentally, supervised the erection of the Cavendish Laboratory in 1871 and became its first professor. Nagaoka proposed that an atom consisted of concentric rings of electrons rotating about a central massive particle that was charged positively. He assumed that an inverse square law applied to the repulsions between electrons and to the attraction between an electron and the central mass, and showed that such a system would be stable if the attractive force was large. He pointed out that this model is consistent with the small scattering of electron beams but mentioned that such a system must ultimately come to rest because of the exhaustion of energy by radiation. He theorized that the emission of light by an atom is the result of the vibrations of the rings of electrons and that several rings must exist to explain the spectral series. Radioactivity was accounted for by assuming that the oscillations of adjacent rings could interact and, if a case of resonance occurred, the amplitude could become sufficiently large for a ring to break and fly apart with great speed. This could result in the ejection of electrons (β-particles) and some of the central mass (α-particles) as well if the position of the centre of gravity of the atom is to be conserved. Moreover, he discussed the change of resistance of a semi-insulator, like selenium, by exposure to light and reasoned that if this was of the correct frequency, ". . . then it will set the electrons in resonating vibration, break them from the revolving system, and thereby cause the flow of electrons and reduce the resistance of the circuit."

Other theories about the structure of the atom were produced by Lord Rayleigh, Jeans, and Schott; these were developments of Thomson's model that attempted to account for the sharpness of spectral lines. Rayleigh's theory required that within the positive sphere should be electrons that were so numerous that they could be regarded as an electric fluid. Jeans postulated that even opposite charges would repel if they were brought sufficiently close together. Schott achieved stability in his atom model by suggesting that the electron, slowly and continually, increased in radius. In a later paper, Schott produces convincing evidence against Nagaoka's Saturnian theory showing that this system would produce insufficient spectral lines and moreover, the frequencies of the lines would be in an arithmetical progression so that a spectrum would not be a band or a series but a set of lines of constant frequency difference.

48.6 ERNEST RUTHERFORD

To solve such a difficult problem as the structure of the atom clearly required a man of great scientific imagination who was also an experimental physicist of supreme genius. Such was Rutherford. A brief account of his life will give some indication of his stature.

Ernest Rutherford was the son of a flax farmer and was born in 1871 near Nelson in New Zealand. By the age of 22 he had gained his M.A. with a double first in mathematics and physics. Two years later he won a scholarship to Cambridge University and at the Cavendish laboratory J. J. Thomson soon spotted his

ability. In 1897 he changed his research interest to the newly discovered radioactivity and it was at this time that he recognized and named two types of radiation, α- and β-particles. A year later at the age of 27 he went to Canada to become a research professor at McGill University, Montreal. With one of his students he soon discovered that thorium compounds emit an emanation, or gas, which is radioactive with a half-life of about 1 minute, and that this gas would leave on surfaces a deposit with a half-life of about 11 hours.

In 1900 Rutherford went to New Zealand to marry, and on returning to McGill, he continued to investigate the gas from thorium compounds. Working with a young chemist from Oxford, Frederick Soddy, he deduced that transmutation was occurring spontaneously and continuously; i.e., that as an α- or β-particle is emitted, the parent atom is transformed into a different element which in turn could be radioactive and decay into yet another element. This theory was so unorthodox that it met widespread criticism, but as Rutherford pointed out it was based exactly on what was happening in his experiments.

In 1907 when he was 36 and wishing to be closer to developments in England, he returned and became Professor of Physics at Manchester University. There in new, well-equipped laboratories he was to lead a highly productive team, and as his reputation spread, more physicists of high calibre were attracted to Manchester. In 1908 he was awarded the Nobel prize for chemistry, and at about this time showed without doubt that the α-particle was a charged atom of helium. Perhaps his greatest achievement was to inspire the scattering experiments of a colleague, Hans Geiger, and his student Marsden. In interpreting their results, Rutherford was able to deduce the structure of the atom that was at the time the subject of so many theories. As this, and further developments, will be covered later in this chapter, only a list of the major events of Rutherford's career will now be given to complete this account of his life.

1912 Niels Bohr came to Manchester and with Rutherford extended the nuclear model of the atom.

1913 H. G. Moseley worked on X-rays at Manchester and introduced the idea of atomic number.

1914 Rutherford was knighted.

1918 He transmuted nitrogen into oxygen.

1919 Rutherford succeeded J. J. Thomson at the Cavendish laboratory. Working with Chadwick he transmuted other light elements and showed that protons were emitted in the process.

Under Rutherford's direction, the Cavendish saw further advances in physics, including the discovery of the neutron by Chadwick in 1932, and the building of the world's first particle accelerator for nuclear physics research by Cockcroft and Walton in 1932. They used it to accelerate protons to high energies. These accelerated protons were then used to transmute lithium and other elements.

Rutherford was raised to the peerage in 1931, being created Baron Rutherford of Nelson. He died at Cambridge in 1937.

48.7 THE SCATTERING EXPERIMENTS OF GEIGER AND MARSDEN

Rutherford was aware of the various theories about the structure of the atom and had, with Hans Geiger, observed that α-particles would pass through thin foils of heavy metals and that the scattering was very small, being of the order of one degree. He must have noted that this was in agreement with the ideas of his former professor, J. J. Thomson. In 1909, Geiger asked Rutherford if he had any ideas for a small research project that would occupy one of his young students, Marsden, and give him some experience of radioactive methods. Although, as he confessed later, he thought there was not the slightest chance of success, Rutherford replied "Why not let him see if any α-particles can be scattered through a large angle."

The experiment that was devised was beautiful in its simplicity yet it helped solve the puzzle of atomic structure. The apparatus is shown in Fig. 48.3. The results of the experiment were immediate. The very small number of tiny spots of light, or scintillations, seen on the zinc sulphide screen announced that a few α-particles were actually bouncing back off the atoms

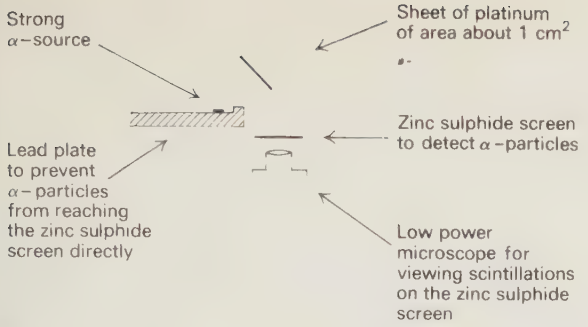


Fig. 48.3 The first experiment of Geiger and Marsden (1910).

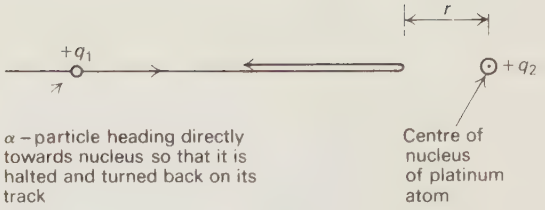


Fig. 48.4 Rutherford's estimate for the maximum "radius" of a nucleus.

of platinum instead of passing through them as everyone had expected. When an excited Geiger reported this to Rutherford he was astonished and recalled later

> "It was quite the most incredible event that has happened to me in my life. It was almost as incredible as if you fired a 15 inch shell at a piece of tissue paper and it came back and hit you."

Geiger and Marsden made three different determinations showing that of the large number of α-particles incident on the platinum sheet, only about 1 in 8000 was reflected, or scattered through angles of more than 90°.

It is of considerable interest that this first experiment which is so crucial to our understanding of the atom, was a rather rough and ready one, especially when it is compared with the later experiments. For instance, in calculating the total number of α-particles reflected, from the number arriving on the tiny screen, they made the assumption that the reflected particles were distributed *uniformly* over a hemisphere with the platinum sheet as its centre. That this happened to be approximately true was due to the fact that the source was not collimated and so a wide range of angles of incidence was possible. Rutherford's comment was "The form of the experiment was not very suited for accurate calculation, but . . .".

What was important though is that Rutherford recognized that the fraction of particles reflected was very small but that, nevertheless, it was real. Fortunately he had a flair for realizing the implications of such small but significant facts.

48.8 THE NUCLEAR ATOM EMERGES

Rutherford considered these remarkable results for about a year, and one day he visited Geiger and declared that he knew what the atom looked like. Most of the mass of an atom is concentrated in a minute positively charged nucleus with electrons circling around it.

In the paper published on the subject in 1911 he recalled the work of Nagaoka whose Saturnian atom was the only model to suggest a nucleus. The rarity of a large deviation of an α-particle led Rutherford to infer that the target by which it was repelled was very small compared with the clear space around the target. Also he reasoned that to deflect anything as fast and massive as an α-particle would require a large force. Assuming that the inverse square law could be applied in this case such a force would arise if the two charges involved could approach to within a very short distance. This implies that the positive charge on an atom was concentrated into a small point rather than being distributed throughout the volume of a sphere. He first calculated how close an α-particle could approach a positive nucleus if it was going to make a head-on collision. Figure 48.4 illustrates the method.

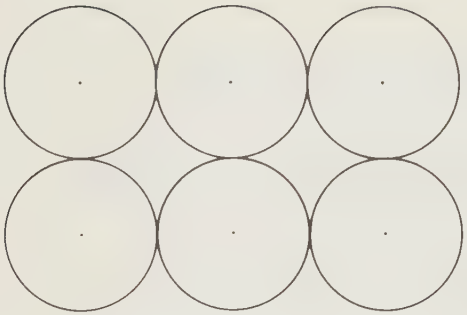


Fig. 48.5



Fig. 48.6

The α-particle initially has a kinetic energy of $\frac{1}{2}mv^2$, and at the point where it is momentarily at rest, all this energy is converted into electrical potential energy. If at this point, the centres of the two charges are separated by a distance r, and the inverse square law applies, then

$$\tfrac{1}{2}mv^2 = \frac{q_1 q_2}{4\pi\varepsilon_0 r},$$

where

$$\frac{1}{4\pi\varepsilon_0} = 9 \times 10^9 \text{ N m}^2 \text{ C}^{-2}.$$

For an α-particle,

$$m = 6.7 \times 10^{-27} \text{ kg},$$

$$v \approx 2 \times 10^7 \text{ m s}^{-1},$$

$$q_1 = 2 \times (1.6 \times 10^{-19}) \text{ C}.$$

For the charge on a platinum atom, Rutherford accepted Barkla's estimate that the number of electrons, and therefore positive charges, is approximately equal to half the mass number. Thus for a platinum atom,

$$q_2 \approx 100 \times (1.6 \times 10^{-19}) \text{ C}.$$

As you can check, with a little arithmetic, the formula gives

$$r = 3.4 \times 10^{-14} \text{ m}.$$

This value cannot be quoted as the size of the nucleus. All that can be said is that the nucleus cannot have a radius greater than 3.4×10^{-14} m, if an inverse square law of force can be assumed to operate at such minute distances. (See Section 48.9.) If, in the calculation, a higher speed of $v = 3 \times 10^7$ m s^{-1} had been used then the corresponding value of r became 1.5×10^{-14} m.

It is now clear, however, that the nucleus is very small indeed compared with the radius of the atom itself, for this is of the order of 1.3×10^{-10} m. (See calculation of the distance between atoms in Section 47.7.)

It is hard to picture a structure in which the outermost fringe has a radius 10 000 times greater than that of the nucleus at its centre. Usually we draw diagrams such as Fig. 48.5 which shows a plane of six atoms, but here the nucleii have radii of about 0.3 mm and the atoms have radii of 10 mm, which is only 30 times greater.

It is difficult to draw a scale diagram of an atom unless a very large piece of paper is used. Inside the square of Fig. 48.6 is printed a very small dot having a radius of about 0.05 mm. If it represents the nucleus of an atom, then the edge of the atom, drawn to scale, will be 0.5 metre away, and the nucleus of the next atom will be a whole metre away.

Once it has been realized that the nucleus is so small compared with the open space around it, one is

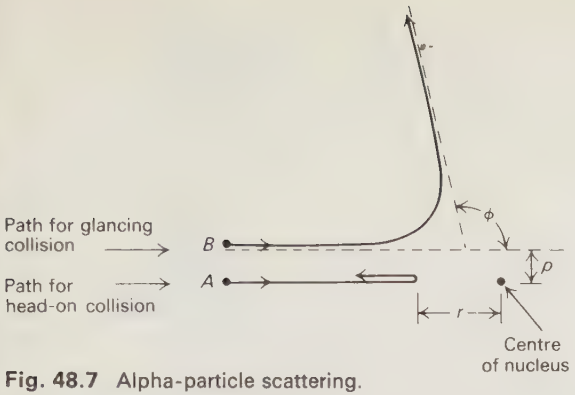


Fig. 48.7 Alpha-particle scattering.

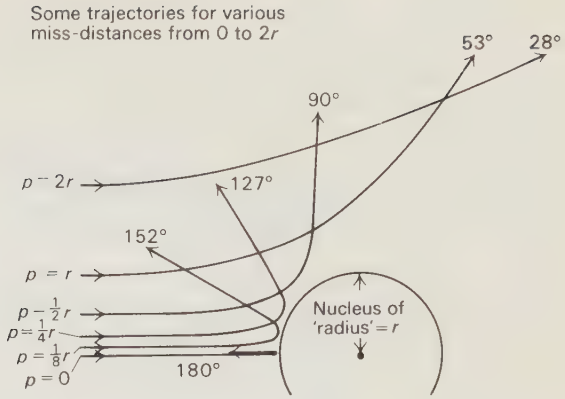


Fig. 48.8

not so surprised that α-particles can pass right through thin foils of metal and only very occasionally suffer a direct collision with a nucleus. What now becomes difficult to understand is that most of the mass of an atom is concentrated in such a tiny volume as a nucleus.

48.9 ANGLES OF SCATTERING

In his paper of 1911, Rutherford showed that if an α-particle did not head directly towards the nucleus then it would undergo a glancing collision and move in a hyperbolic trajectory. In Fig. 48.7 particle *B* is an example, and had it been able to continue on its original course, it would have missed the centre of the nucleus by a distance p. However, it is repelled and finally travels in a direction making an angle ϕ with its original path. It should be noted that if the angle between the asymptotes of the hyperbola is bisected, then the bisector will pass through the nucleus.

Rutherford derived an equation relating r, p, and ϕ, namely,

$$\cot \tfrac{1}{2}\phi = \frac{2p}{r}. \qquad (48.1)$$

Table 48.2

$p =$	$100r$	$10r$	$5r$	$2r$	r	$\frac{1}{2}r$	$\frac{1}{4}r$	$\frac{1}{8}r$	0
$\phi =$	0.6°	6°	11°	28°	53°	90°	127°	152°	180°

Table 48.2 was obtained by putting values in this equation. This gives us an idea of the scattering angles that can be expected for various "miss-distances" or "aiming errors," p, which have been expressed as multiples of the nuclear radius. Some of these values have been used to draw Fig. 48.8 which emphasizes that ϕ is a large angle if, and only if, p is less than r.

The deflection is certainly negligible when $p = 100r$, and, to return to the scale diagram of Fig. 48.7, this means that an α-particle will hardly be deflected unless it falls within 5 mm of that tiny dot. Recall that the radius of the atom on this scale is 500 mm and you will see that there is plenty of opportunity for an α-particle to remain undeflected.

48.10 FURTHER TESTING OF THE NUCLEAR ATOM MODEL

Equation (48.1) cannot be tested directly for it is not possible in an experiment to measure values of the aiming error p. So, Rutherford derived Eq. (48.2) which relates the number of particles scattered on to a small screen to measurable quantities.

$$\Delta N = \frac{kNq_1^2q_2^2nt}{E_k \sin^4 \frac{1}{2}\phi}, \qquad (48.2)$$

where ΔN = number of particles scattered per second on to unit area of a small screen, k = a constant depending on the system of units used and the

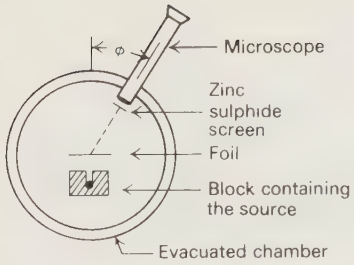


Fig. 48.9 Apparatus used for testing the relationship between scattering angle and the number of particles scattered. The angle ϕ could be varied by rotating the foil and the source relative to the microscope and screen.

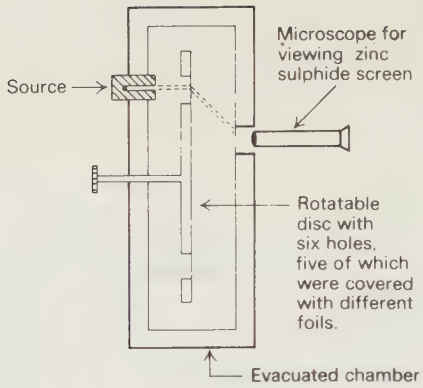


Fig. 48.10 Apparatus used for examining the effect on the number of particles scattered of (a) the thickness of the foil, (b) the mass number of the foil material and (c) the speed of the alpha-particles. In this device the scattering angle was kept constant, only the foils being changed.

dimensions of the apparatus, N = total number of particles falling per second on the foil of metal, q_1 = charge on the α-particle, q_2 = charge on the nucleus, n = number of atoms per unit volume of the foil, t = thickness of the foil, E_k = energy of the α-particle ($= \frac{1}{2}mv^2$), and ϕ = scattering angle.

Note: This formula is true for thin foils only, i.e., foils through which particles can pass with only a small change in speed. Gold leaf is a suitable example even though it is some 2000 atoms thick.

This equation was tested by Geiger and Marsden using more refined techniques than previously. The principle of their apparatus is outlined in Figs. 48.9 and 48.10. Both pieces of apparatus used collimated sources so that a narrow 'pencil' of particles was used. Because the scattering effect is so small, these sources had to be strong (in modern terms, over 100 millicuries, and therefore thousands of times stronger than would be permitted in any teaching laboratory today). Also, the apparatus was evacuated so that complications should not arise from α-particles being slowed down, or even scattered, by collisions with air molecules.

In addition to justifying the assumption that the inverse square law of force applied down to microscopic dimensions these experiments confirmed the major predictions of Eq. (48.2). Since most of the formula had been checked, Geiger and Marsden assumed that the proportionality between ΔN and q_2^2 was valid and deduced a value for the charge on the nucleus. Their value for gold was confirmation that the nuclear charge was given approximately by:

$$(\tfrac{1}{2} \text{ mass number} \times \text{numerical value of the charge on the electron}).$$

This is by no means conclusive since the accuracy stated for this part of their work was 20%. Indeed, later, in 1920, Chadwick using scattering experiments but with improved techniques showed that this relationship was not an exact one, and that nuclear charge was related to atomic number or the position of the element in the periodic table.

48.11 ATOMIC NUMBER

The scattering experiments led to the conclusion that the number of electrons per atom equalled half the atomic mass. In 1911 a Dutch amateur of physics,

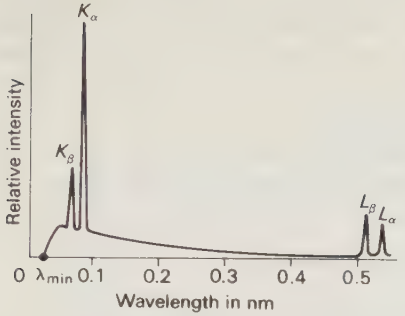


Fig. 48.11 Graph of intensity of X-rays against wavelength for a solid target.

A. van de Broek, suggested in a letter to *Nature* that each element in the periodic system could be characterized by a number of possible, permanent charges of each sign per atom. Two years later, in a further letter, van der Broek indicated that Geiger and Marsden's data for scattering by the atoms of different elements showed better proportionality if the number of α-particles scattered was compared with the square of the number assigned to the element in the periodic table rather than the square of the mass number. He suggested that the nuclear charge was given by the product of this number and the numerical value of the electronic charge. So the atom of gold would consist of a tiny nucleus carrying a charge of $+79e$ (where e is the charge on the electron) with 79 electrons in the nearby space. This number is now known as the *atomic number*.

Confirmation of this suggestion came from several quarters; from the radiochemical evidence developed by F. Soddy and A. Fleck and from the work which was being carried out with X-rays notably by C. G. Barkla and H. G. Moseley. The former had found that when X-rays produced at one target strike a second target, the second target can emit a *characteristic radiation* with a definite wavelength or series of wavelengths as well as scatter the original rays. This characteristic radiation could also be excited directly from the target electrode if the energy of the incident electrons was high enough and it proved to be characteristic of the metal used (see Fig. 48.11). This was, evidently, the X-ray analogue of the optical line spectrum. But Moseley revealed that it was a surprisingly simple spectrum.

Table 48.3 Wavelengths of the K_α and K_β lines in certain X-ray spectra

Element	Atomic number Z	Mass number A	Wavelength (nm)	
			K_α	K_β
Manganese	25	54.93	0.210	0.191
Iron	26	55.85	0.194	0.175
Cobalt	27	58.94	0.179	0.162
Nickel	28	58.69	0.166	0.148
Copper	29	63.54	0.154	0.139

Moseley found that the wavelengths of the lines became progressively shorter as the atomic number of the target increased. Table 48.3 shows this and also indicates that the observed regularities must be related to the atomic number and not to the mass number. Consider, for example, the cases of cobalt and nickel; the latter has the higher atomic number but the lower mass number. The evidence of the wavelengths supports the view that it is the atomic number which is the fundamental measure. Moseley found that the frequency of the K line was given by the empirical relationship

$$f = 2.48 \times 10^{15} (Z - 1)^2 \text{ Hz},$$

where Z is the atomic number.

48.12 MODELS OF THE ATOM

Enough has been said to indicate the state of ferment in which the physics community existed in the years around 1910. Into this ferment of discovery and activity Niels Bohr, a Danish physicist working with Rutherford's team, threw the first attempt at a

quantum model of the nuclear atom in 1913. It is not our intention to develop this model of the atom here; excellent accounts are to be found in, for example, Arons (1965), Baez (1967), Holton and Roller (1958), Rogers (1960). We shall content ourselves with a rather less specific model in which a massive, minute nucleus, carrying positive charges equal in number to the atomic number, is surrounded by a cloud of electrons of the same number. Such an atom is electrically neutral, largely empty space and controlled by Coulomb forces. Each electron in the cloud can exist with a definite energy and, in receiving or losing energy quanta (photons) can change to one of many other energy levels. Our evidence for this lies in the work already discussed in this and the previous Unit.

We shall turn next to consider experimental results which concern the nucleus of the atom.

Chapter 49
THE NUCLEUS

49.1 The composition of the nucleus
49.2 The artificial transmutation of an element
49.3 Natural transmutation
49.4 The search for a neutral particle
49.5 The discovery of the neutron
49.6 The nuclear structure of elements
49.7 Alpha-particle emission
49.8 Isotopes
49.9 Beta-particle emission
49.10 Gamma radiation
49.11 The behaviour and detection of neutrons
49.12 Modern alchemy—a dream come true
49.13 Epilogue
49.14 Notes on the cloud chamber photographs

49.1 THE COMPOSITION OF THE NUCLEUS

E. Marsden completed his work on the scattering of α-particles by thin foils in 1913, and as this method of probing the atoms of elements with α-particles had been so successful, he was soon attacking elements at the "light" end of the periodic table. On the assumption that interactions between nuclei are elastic, it had been reasoned that as a result of a direct hit from an α-particle, a hydrogen atom should recoil from the collision with a speed 1.6 times that of the α-particle itself. This can be readily checked using Eq. (7.7), namely,

$$v_2 = \frac{2u}{\left[1 + \dfrac{m_2}{m_1}\right]},$$

where m_1 and u represent the mass and initial speed of the alpha-particle, and m_2 and v_2 represent the mass and recoil speed of the hydrogen atom.

Furthermore, as the mass of the hydrogen atom is a quarter that of the α-particle, the hydrogen atom will end up with 0.64 of the alpha's original energy. (To verify this see Table 7.5 and Fig. 7.16.)

A hydrogen atom gaining such a speed is likely to lose its electron and to travel as a hydrogen nucleus, or proton. Being charged positively, it should ionize other atoms in its path, steadily losing energy at each encounter until it comes to rest.

It was further calculated that the proton would have a range 4 times greater than that of the α-particle, even though it had only 64% of its energy. It may seem surprising that the particle with the smaller energy should travel further before coming to rest, but it should be remembered that the proton has half the charge carried by an α-particle and consequently it is less able to affect the electrons around the atoms it passes. Also, because of the higher speed of the proton, it spends less time in the vicinity of an atom and has less chance of ionizing it. Because a high speed proton does not ionize so effectively as an α-particle it can travel further before losing all its kinetic energy.

The results of Marsden's experiments were first described in a paper on the structure of the atom by Rutherford:

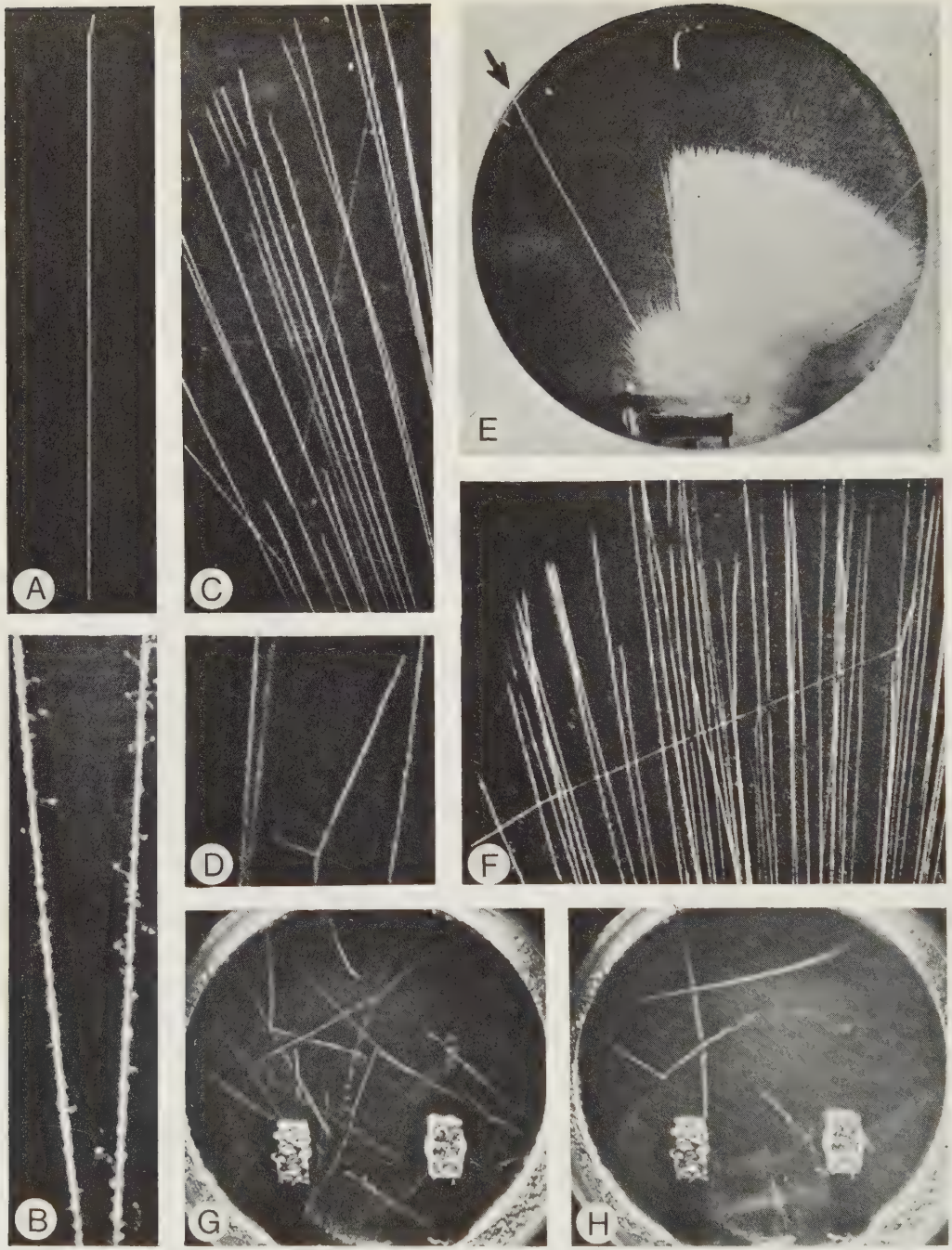
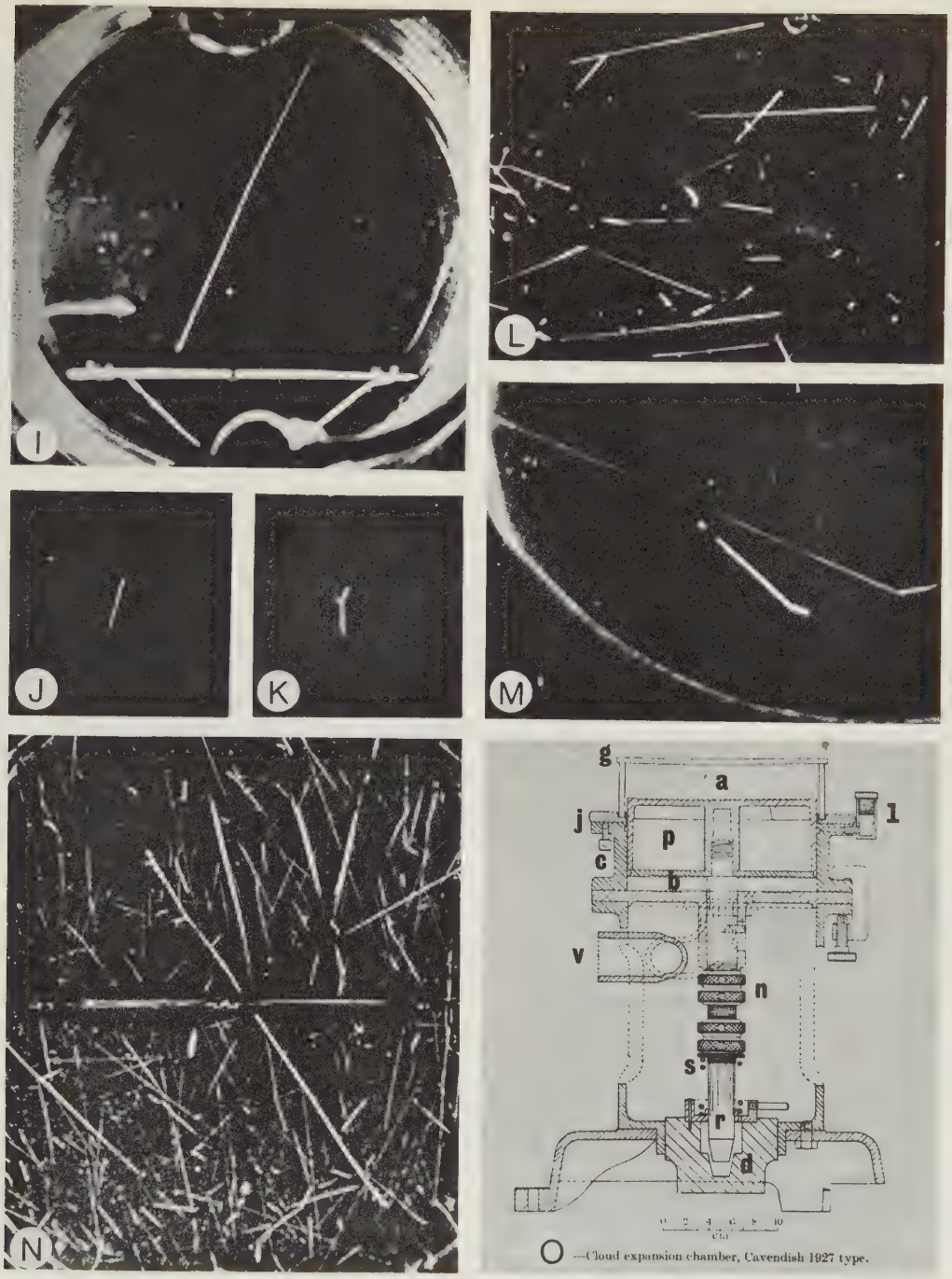


Fig. 49.1 A selection of photographs illustrating work with cloud chambers. See Section 49.14.



O —Cloud expansion chamber, Cavendish 1927 type.

"Mr. Marsden has kindly made experiments for me to test whether the presence of such hydrogen atoms (i.e., protons) can be detected. The general method employed was to place a thin α-ray tube containing about 100 millicuries of purified emanation (radon) in a tube filled with hydrogen . . . the great majority of the scintillations disappeared at about 20 cm from the source, which corresponds to the range of the α-particle in hydrogen. A small number of scintillations, however, persisted in hydrogen up to a distance of about 90 cm. The scintillations were of less intensity than those due to the ordinary α-particle. There appears to be no doubt that the scintillations observed beyond 20 cm are due to charged hydrogen atoms which are set in swift motion by a close encounter with an α-particle." (Rutherford, 1914.)

Thus the existence of the proton had been established and it could take its place alongside α- and β-particles as one of the atomic projectiles that cause ionization.

Some of the features about ionization mentioned above are beautifully illustrated by cloud chamber photographs (Fig. 49.1): long proton tracks can be seen in C and E, and the variation of ionization with speed is evident in A.

During his experiments Marsden had noticed some indications that the radioactive source itself seemed to give rise to protons, but further experiments were interrupted by his departure to New Zealand in 1915 to take up the post of Professor of Physics at Wellington University. Unfortunately he found that the quantity of radium available there was insufficient for the continuation of his work. In England Rutherford realized that this work was important, for if Marsden's suspicions were correct, then this was evidence for the existence of an element, other than helium (in the form of α-particles), in the nucleus of a radioactive substance. World War I had by now broken out and further experiments on radioactivity were put aside. As the war drew to a close, Rutherford resumed his experiments and concluded that, after all, protons were not emitted from the radioactive substances themselves. Although this line of investigation

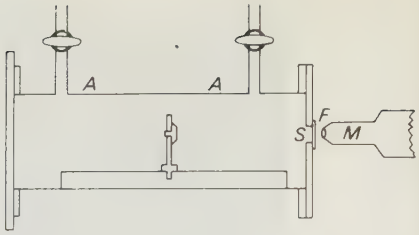


Fig. 49.2

proved fruitless, he made further discoveries during these experiments that made him the first successful alchemist.

49.2 THE ARTIFICIAL TRANSMUTATION OF AN ELEMENT

In his experiments on the scattering of α-particles by atoms of low atomic number Rutherford had observed that α-particles did indeed cause scintillations to occur on a zinc sulphide screen placed beyond the range of the particles in the air or other gas which filled the space between the source and the screen. The particles producing these scintillations had the properties of protons, or H atoms as he called them, for they were positively charged, were deflected by magnetic fields and had the same range and energy as the protons produced by the passage of α-particles through hydrogen itself. Rutherford was concerned to know where these protons came from.

He described his experiment thus:

"The intense source of radium C [in modern terms $^{214}_{83}Bi$] was placed inside a metal box about 3 cm from the end, and an opening in the end of the box was covered with a silver plate of stopping power equal to about 6 cm of air. The zinc sulphide screen was mounted outside, about 1 mm distant from the silver plate, to admit of the introduction of absorbing foils between them. The whole apparatus (Fig. 49.2) was placed in a strong magnetic field to deflect the beta rays. . . . When dried oxygen or carbon dioxide was admitted into the vessel, the number of scintillations diminished to about the number to be expected from the stopping power of the column of gas.

A surprising effect was noticed, however, when dried air was introduced. Instead of diminishing, the number of scintillations was increased, and for an absorption corresponding to about 19 cm of air the number was about twice that observed when the air was exhausted. It was clear from this experiment that the α-particles in their passage through air gave rise to long-range scintillations which appeared to the eye to be about equal in brightness to H scintillations. A systematic series of observations was undertaken to account for the origin of these scintillations."

This series of observations eliminated, in turn, the possibilities that the scintillations were due to the nuclei of nitrogen or of oxygen which had been knocked on, or to the presence of water vapour in the air, or to the small amounts of hydrogen which are present in air, or to the presence of dust in the air. Rutherford goes on

"Since the anomalous effect was observed in air, but not in oxygen, or carbon dioxide, it must be due either to nitrogen or to one of the other gases present in atmospheric air. The latter possibility was excluded by comparing the effects produced in air and in chemically prepared nitrogen. . . . With pure nitrogen, the number of long-range scintillations under similar conditions was greater than in air. As a result of careful experiments, the ratio was found to be 1.25, the value to be expected if the scintillations are due to nitrogen. From the results so far obtained it is difficult to avoid the conclusion that the long-range atoms arising from the collision of α-particles with nitrogen are not nitrogen atoms but probably atoms of hydrogen, or atoms of mass 2. If this be the case, we must conclude that the nitrogen atom is disintegrated under the intense forces developed in a close collision with a swift α-particle, and that the hydrogen atom which is liberated formed a constituent part of the nitrogen nucleus." (Rutherford, 1919.)

It was now clear that protons form a constituent part of a nitrogen nucleus, and that nitrogen nuclei can be disintegrated by being bombarded with α-particles. But what does the nitrogen atom turn into? This was the question that next had to be answered.

While thought was being given to this problem, Rutherford bombarded other light elements, and in conjunction with Chadwick found that disintegration protons could be emitted by all elements from boron to potassium with the exception of carbon and oxygen. A surprise was that some of these protons were emitted with *more* energy than that possessed by the incoming α-particle. This, and the fact that the protons were emitted in all directions and not just *forwards* indicated that the mechanism was not that of an elastic collision, but a hyperelastic one (see Chapter 7). (In, for instance, an elastic collision between billiard balls, a formerly stationary ball will always be knocked forwards or at most, sideways; it can never be knocked backwards in the direction from which the moving ball approached. However, if the stationary ball is a small bomb that detonates as soon as it is struck, a fragment is likely to go in any direction, and the total kinetic energy of the system will be increased.)

The process behind these nuclear disintegrations was made clear in 1925 by P. M. S. Blackett, who realized that α-particles hitting nitrogen atoms in a cloud chamber could produce tracks that would arbitrate between two possible nuclear reactions.

The first of these (i), the knock-on of a nuclear proton, could be described:

An α-particle collides with a nitrogen nucleus. A proton is chipped off, the α-particle continues on its way with less energy, and leaves behind a carbon nucleus recoiling from the collision.

$$^{4}_{2}\mathrm{He} + {}^{14}_{7}\mathrm{N} \rightarrow {}^{1}_{1}\mathrm{H} + {}^{4}_{2}\mathrm{He} + {}^{13}_{6}\mathrm{C}.$$

The second (ii), an alpha capture process, could be described:

An α-particle collides with a nitrogen nucleus to form an unstable nucleus of fluorine which decays, ejecting a proton and leaving a recoiling oxygen nucleus.

$$^{4}_{2}\mathrm{He} + {}^{14}_{7}\mathrm{N} \rightarrow {}^{18}_{9}\mathrm{F} \rightarrow {}^{1}_{1}\mathrm{H} + {}^{17}_{8}\mathrm{O}.$$

A simpler notation for this last reaction is

$$^{14}\mathrm{N}\ (\alpha,\ p)\ ^{17}\mathrm{O},$$

in accordance with the convention that nuclear reactions should be written in the form

$$\text{initial nuclide}\begin{pmatrix}\text{incoming}\\ \text{particle(s)}\\ \text{or quanta}\end{pmatrix},\ \begin{pmatrix}\text{outgoing}\\ \text{particle(s)}\\ \text{or quanta}\end{pmatrix}\text{final nuclide.}$$

The decision between these alternatives depended upon whether an alpha track in a nitrogen filled cloud chamber could be found branching into three tracks (for i) rather than into two tracks (for ii). The cloud chamber provided the only direct method of observing all the particles involved in a collision. Accordingly, Blackett took 23 000 photographs at the rate of 4 a minute, and he examined over 400 000 α-particle tracks. Some showed an elastic collision in which a *complete* nitrogen nucleus was knocked forwards. Blackett (1925) continues the story:

> "But amongst these normal forks due to elastic collisions, eight have been found of a strikingly different type. These eight tracks undoubtedly represent the ejection of a proton from a nitrogen nucleus. It was expected that a photograph of such an event would show an α-ray track branching into three. The ejected proton, the residual nucleus from which it had been ejected, and the α-particle itself, might each have been expected to produce a track. These eight forks, however, branch only into two. The path of the proton consists of a fine straight track, along which the ionization is clearly less than along an α-ray track, and must therefore be due to a particle of small charge and great velocity. The second of the two arms of the fork is a short track similar in appearance to the track of the nitrogen nucleus in a normal fork. Of a third arm to correspond to the track of the α-particle itself after the collision there is no sign. . . . In ejecting a proton from a nitrogen nucleus the α-particle is itself bound to the nitrogen nucleus. The resulting new nucleus must have a mass 17, and provided no electrons are gained or lost in the process, an atomic number 8." (Blackett, P. M. S., 1925.) [This last mention of electrons refers to the idea, held at the time, that electrons existed in a nucleus together with protons. On theoretical grounds, this hypothesis has since been shown to be incorrect and was rendered unnecessary by the discovery of the neutron.]

In a later series of experiments Blackett photographed another 1 100 000 tracks and obtained four more disintegration collisions. One of these is reproduced in Fig. 49.1(F). This is an example of a two-pronged fork with one prong coming *backwards*.

It is worth pondering for a moment on the fact that this was confirmation that Rutherford had detected the first artificial transmutation of an element in 1918. This was the first example of an artificially induced nuclear reaction. Man had at last achieved transmutation of an element, turning nitrogen into oxygen. This may only have occurred one atom at a time and is a far cry from the alchemists' dream of changing lead into gold. The greatness of this achievement lies not in the transmutation itself but in the development by Rutherford and his co-workers of methods of probing the nucleus.

49.3 NATURAL TRANSMUTATION

It should be mentioned that while artificial transmutation was a rarity, natural transmutation had been occurring for thousands of years, ubiquitously and inevitably, during the decay of radioactive substances. Natural transmutation of an element and the idea of a radioactive series can be easily and strikingly demonstrated with a simple diffusion cloud chamber. In such a decay series, whenever an atom ejects an α- or β-particle, it changes into a different element according to the displacement laws (Section 48.2). For example, if an atom of radon-220 (a radioactive gas emitted by thorium compounds) ejects an α-particle, it turns into a different element, polonium-216; and if this too ejects an α-particle, an atom of lead-212 is left. This in turn ejects a β-particle and so the original atom becomes bismuth-212. Part of this series can be demonstrated by injecting into the cloud chamber a tiny puff of radon-220 from a plastic

squeeze bottle containing thorium hydroxide or thorium carbonate. At first the whole chamber is filled with α-tracks darting in all directions, but every 55 seconds (the half-life of ^{220}Rn) the number visible is halved. When the number present has fallen so that individual tracks can be studied easily it will be noticed that frequently two tracks come from the same point to form a V. Examples of this can be seen in photographs G and H, Fig. 49.1. Transmutation has taken place at the point of each V; the original radon atom ejected an α-particle and became a polonium atom, and because this element has a half-life of only 0.16 s there is a high probability of a second α-track occurring very soon after the first. Of course, the actual atoms and α-particles are invisible but the cloud chamber allows us to see the spot where an atom of one element changed into an atom of a different element.

49.4 THE SEARCH FOR A NEUTRAL PARTICLE

Now that the proton had been established as a constituent of the nuclei of many elements, another difficulty arose. How could, for instance, a nitrogen nucleus have a charge equivalent to 7 proton charges yet have a mass equal to that of 14 protons? To provide a solution Rutherford speculated on the existence of a neutral particle having the mass of a proton but without any charge. He mentioned this rarely in public for he disliked making pronouncements that were unsupported by experimental evidence. Indeed, a neutral particle was likely to be difficult to detect since, being uncharged, it could not produce ionization and leave tracks in a cloud chamber, and it would not be possible to deflect it with magnetic or electric fields. It would also have great penetrating power, for, having no charge it could pass very close to a nucleus without effect and so would pass through matter with ease.

J. Chadwick became interested in Rutherford's idea and during the decade from 1920 they were both on the watch for evidence of a neutral particle.

Following the successes of Marsden and Rutherford, the bombarding of various substances with α-particles became a popular occupation of scientists of the early twentieth century, just as "element hunting" had been the sport of those in the nineteenth century. The detectors used were counters, expansion chambers and high pressure ionization chambers.

The next major advance using this technique did not come until 1930 when W. G. Bothe and G. F. Becker in Germany, and H. C. Webster at the Cavendish laboratory, bombarded light elements (Al, Mg, B, Be) with α-particles and discovered that a penetrating radiation resembling γ-rays was emitted. This effect was particularly marked in the case of beryllium and the radiation from it possessed a penetrating power greater than that of any known gamma radiation from a radioactive source. Webster made the further observation that the radiation from beryllium which was emitted in the same direction as the incident α-particles was more penetrating than the radiation emitted in the backward direction. This particularly excited Chadwick for it suggested to him that the radiation consisted of knocked-on particles, and moreover, from its penetrating power, of neutral particles. He suggested that the particles should be passed into a cloud chamber, hoping that despite their lack of charge they would produce tracks. Chadwick wrote later:

> "To our dismay, for we were convinced that a neutral particle of some kind was involved, no such tracks were to be seen. We were very puzzled; we did not know how to reconcile the observations."

This is another classic example of a near-miss of a major discovery. If only they had tried different gases in the cloud chamber, used a stronger source, or even just taken more photographs!

Among those who were investigating the radiation from beryllium were the French physicists Frédéric Joliot and his wife Irène Curie, a daughter of the discoverers of radium. Both were skilled radiochemists and had prepared a very strong source of polonium (of the order of 200 millicuries and at that time the largest ever made). Early in 1932 they announced a remarkable discovery that astonished Rutherford and sent Chadwick hot on the trail of the neutron again.

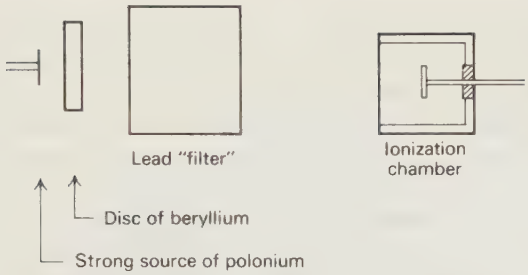


Fig. 49.3 Schematic diagram of the apparatus used by I. Curie-Joliot and F. Joliot (1932).

The polonium was used in conjunction with beryllium to provide a powerful source of "beryllium radiation" which fell on to an ionization chamber with a thin window. The general arrangement is shown in Fig. 49.3.

They found that the radiation was so penetrating that even after passing through 1.5 cm of lead, no appreciable reduction in intensity was observed. The surprise came when a thin sheet of paraffin wax was placed between the lead and the ionization chamber for this caused the intensity of the detected radiation actually to *increase*, in some cases by as much as a factor of two.

The nature of this secondary radiation from the wax was investigated and it was found to be completely absorbed by a sheet of aluminium, only 0.2 mm thick, placed between the wax and the ionization chamber. Such a rate of absorption would be expected if protons were being emitted by the wax, which is a substance rich in hydrogen atoms. This conclusion was soon confirmed directly by observing the tracks of the protons coming from a wax sheet placed in a cloud chamber (see Fig. 49.11).

The fact that alpha-induced radiation from beryllium, ejected protons of very considerable energy from paraffin wax, or other hydrogen compounds (like water or polythene), required an explanation. The Joliots produced a theory based on the idea that the penetrating radiation from beryllium was a form of gamma ray of extremely high energy; but in this theory there were inconsistencies.

Chadwick could not accept their theory on quantitative grounds, for the number of protons actually produced was many orders of magnitude greater than the theory predicted. Also, the energy of a gamma quantum that could eject a proton from paraffin wax would have to be about 50 MeV, and such a large energy could not be accounted for in a reaction involving beryllium and α-particles.

49.5 THE DISCOVERY OF THE NEUTRON

Chadwick soon repeated the Curie-Joliot experiment and from his measurement of the absorption in aluminium deduced that the protons were emitted from the wax with a maximum speed of about 33×10^6 m s^{-1}. He asked a colleague, N. Feather, to investigate the distance that nitrogen nuclei recoiled from the onslaught of the mysterious blows from beryllium radiation. For this he used a nitrogen-filled cloud chamber and obtained photographs, two of which are included as J and K (Fig. 49.1). The maximum track length was only 3.5 mm, yet it allowed Chadwick to calculate that the nitrogen nuclei recoiled with an initial speed of 4.7×10^6 m s^{-1}.

Chadwick now had sufficient information to show that the invisible radiation from beryllium could consist of particles each having a mass about equal to that of a proton. For his calculation he assumed the collisions were elastic and used Eq. (7.7). This equation, rearranged, is:

$$v_2 = \frac{2m_1u}{(m_1 + m_2)},$$

where m_1 = mass of incident particle, u = speed of incident particle, m_2 = mass of nucleus struck, and v_2 = speed of recoiling nucleus.

In the case of a hydrogen nucleus from wax, and using atomic mass units (u),

$$m_2 = 1 \text{ u}, \qquad v_2 = 33 \times 10^6 \text{ m s}^{-1}.$$

$$33 \times 10^6 = \frac{2m_1u}{(m_1 + 1)}.$$

In the case of a nitrogen nucleus,

$$m_2 = 14 \text{ u}, \qquad v_2 = 4.7 \times 10^6 \text{ m s}^{-1}.$$

$$4.7 \times 10^6 = \frac{2m_1 u}{(m_1 + 14)}.$$

Dividing one equation by the other eliminates u which is unknown. Thus,

$$\frac{33}{4.7} = 7 = \frac{m_1 + 14}{m_1 + 1}.$$

Hence,

$$m_1 = 1.17 \text{ u}.$$

Remembering the shortness and difficulty of measuring the track length of the recoiling nitrogen nucleus, errors of the order of 10% are quite possible. (Had v_2 for nitrogen been 4.4×10^6 m s^{-1}, instead, then m_1 would have been unity.)

Thus it was legitimate for Chadwick to conclude that he had at last found a neutral particle (now known as a neutron) and that its mass was nearly the same as that of the proton. However, to be certain, he showed that the observed energies of the protons from wax were consistent with a reaction of the type:

$$^9_4\text{Be} + ^4_2\text{He} \rightarrow ^{12}_6\text{C} + ^1_0n,$$

where ^1_0n represents a particle of mass 1, and of zero charge, i.e., a neutron.

This equation may be written

$$^9\text{Be}\ (\alpha, n)\ ^{12}\text{C}.$$

Finally, by experimenting with boron instead of beryllium, he obtained a more accurate value for the neutron mass and found it to be between 1.005 and 1.008. It is interesting to note that the best methods now available for determining the neutron mass give 1.008665 atomic mass units as compared with the proton mass of 1.007276 u (based on the scale $^{12}\text{C} = 12$).

James Chadwick's long search for the neutron was rewarded with success, and the 1935 Nobel prize for physics. The discovery could possibly have come a year earlier had the powerful polonium source used in Chadwick's later work been available to Webster. Even with the powerful source sending a few thousand neutrons through the chamber per photograph, Feather found that he had to take 1740 pictures to produce only 130 tracks. Little wonder that Webster's 50 photographs, using a weaker source, provided nothing. The discovery of the neutron was the result of the work of different scientists in different countries; it was the consequence of the tradition in science for publishing and sharing knowledge. The story of the neutron is an excellent "case history" in experimental science; it is a good illustration of how scientists operate, how they think about problems, and how they arrive at conclusions. Chadwick's own reminiscences are interesting:

> "I trust that I shall not be misunderstood if I add a postscript to this story. It is unnecessary to record my satisfaction, and delight, that the long search for the neutron had, in the end, been successful. The decisive clue had indeed been supplied by others. This after all is not unusual; advances in knowledge are generally the result of the work of many minds and hands. But I could not help but feel that I ought to have arrived sooner. I could offer myself many excuses: lack of facilities, and so on. But beyond all excuses I had to admit, if only to myself, that I had failed to think deeply enough about the properties of the neutron, especially about those properties which would most clearly furnish evidence of its existence. It was a chastening thought. I consoled myself with the reflection that it is much more difficult to say the first word on any subject, however obvious it may later appear, than the last word—a commonplace reflection, and perhaps only an excuse." (Chadwick, 1962.)

49.6 THE NUCLEAR STRUCTURE OF ELEMENTS

Not only had the existence of the neutron been established, it had also been shown to result from nuclear reactions in a number of light elements and so could be considered to be a part of all nuclei. The neutron was as much a fundamental particle as a proton, and combinations of these two *nucleons* could build the nucleus of any element. For example, a

helium nucleus has a charge of 2 units and a mass of 4, ($^4_2$He). This can be achieved if the nucleus contains 2 protons and 2 neutrons. Likewise a radium nucleus, $^{226}_{88}$Ra, is built up of 88 protons and 138 neutrons. Note that the mass number is given by the total number of nucleons, which is the family name given to protons and neutrons; the atomic number is the number of protons.

49.7 ALPHA-PARTICLE EMISSION

This model of the nucleus allows explanation of the displacement laws relating to natural transmutation. If a nucleus of $^{226}_{88}$Ra emits an α-particle, then since this is a helium nucleus, the radium must lose 2 protons and 2 neutrons, i.e., it loses a charge of 2 and a mass of 4. Consequently, a new element is formed having a mass number 222 and atomic number 86. Element 86 (see Table 47.6) is radon and so the *daughter* product is $^{222}_{86}$Rn, or radon-222. This process can be expressed by the equation:

$$^{226}_{88}\text{Ra} \rightarrow {}^{222}_{86}\text{Rn} + {}^{4}_{2}\text{He}.$$

49.8 ISOTOPES

In Section 49.3 reference was made to radon-220 ($^{220}_{86}$Rn). This and radon-222 ($^{222}_{86}$Rn) are both atoms of the same element for they have the same atomic number, but they differ in mass number. They are examples of *isotopic nuclides* or *isotopes*. The isotopes of a particular element cannot be distinguished chemically for they have the same number of protons and the same number of orbital electrons; their only difference is in the number of neutrons in their nuclei.

The existence of isotopes seems to have been suggested first by Frederick Soddy in 1910. Soddy, a chemist, had played a major role with Rutherford in unravelling the processes involved in radioactive decay, and, in the few years remaining before the outbreak of war in 1914, was to develop this new concept. Soddy's conclusions were summarized in a report in *Nature*:

> "The chemical analysis of matter is not an ultimate one. It has appeared ultimate hitherto, on account of the impossibility of distinguishing between elements which are chemically identical and non-separable unless these are in the process of change the one into the other. But in that part of the periodic table in which the evolution of the elements is still proceeding, each place is seen to be occupied not by one element, but on the average, for the places occupied at all, by no fewer than four, the atomic weights of which vary over as much as eight units. It is impossible to believe that the same may not be true for the rest of the table, and that each known element may be a group of non-separable elements occupying the same place, the atomic weight not being a real constant, but a mean value, of much less fundamental interest than has hitherto been supposed. Although these advances show that matter is even more complex than chemical analysis alone has been able to reveal, they indicate at the same time that the problem of atomic constitution may be more simple than has been supposed from the lack of simple numerical relations between the atomic weights." (Soddy, 1913.)

The existence of isotopes among the stable atoms of the periodic table was detected by J. J. Thomson at about the same time (*Proc. R. Soc.*, **A89** (1913) 1–20). Thomson was able to show that for example, the gas neon was not a simple gas but a mixture of two gases, one of which had an atomic mass of 20 and the other of 22. These two isotopes, $^{20}_{10}$Ne and $^{22}_{10}$Ne are both samples of the element 10, but one has more mass than the other.

Thomson distinguished between them in his positive ray apparatus which was the precursor of the modern mass spectrometer. Developed by F. W. Aston in the years immediately after the war, this has now become an extremely sophisticated tool of industry and research. Modern versions permit a comparison of atomic masses to one part in 10^7 or better and provide a rapid means of complex atomic and molecular analysis. Figure 49.4 is a schematic diagram of an industrial mass spectrometer.

In such a machine the positive ions are produced in an ion source by electron bombardment. Figure

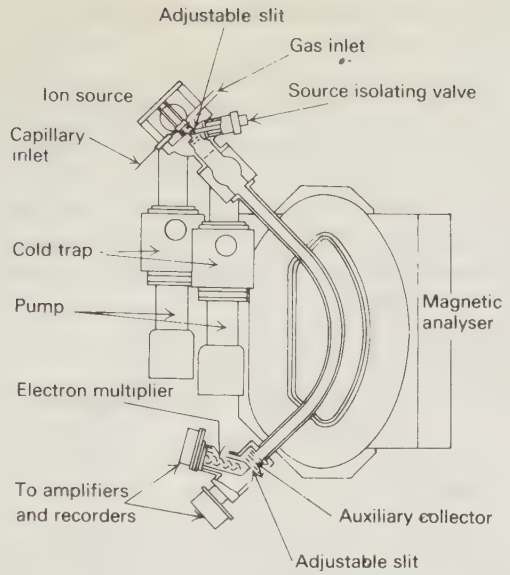


Fig. 49.4 Schematic diagram of the AEI MS12 industrial mass spectrometer. The instrument uses an electron multiplier with a gain of 10^6 to measure the ion current. (Courtesy of AEI Scientific Apparatus Ltd.)

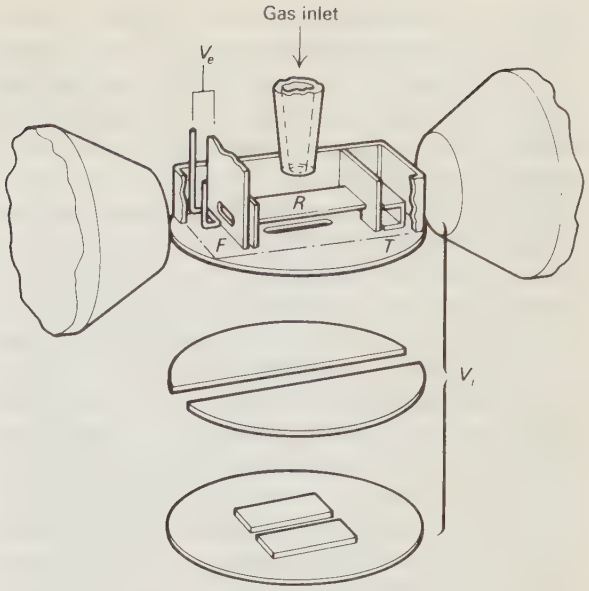


Fig. 49.5 Schematic diagram of an ion source. (From G. P. Barnard, *Modern Mass Spectrometry*, Chapman and Hall, for Institute of Physics 1953. Reproduced by permission.)

49.5 shows such a source. Electrons from the heated filament F are accelerated towards an anode by a voltage V_e which is in the order of 100 V. Those which pass through the hole in this anode will move across the ionization chamber towards the trap T under the guidance of the external magnet system. The sample of material to be investigated, which is in gaseous form, enters the ionization chamber from the top and encounters the stream of electrons. Some of these atoms will be ionized by collision (see Chapters 43 and 45) and will pass with low speeds through a slit in the lower plate of the chamber.

Now they come under the influence of the potential V_i and accelerate towards the lower plate through a series of slits which serves to narrow the ion beam.

The emergent beam, containing a mixture of ions of differing speeds, differing charges, and differing masses enters the field of the electromagnet (magnetic analyser—Fig. 49.4). In this uniform field, the ions will follow a curved path. The radius of this path will depend upon the mass to charge ratio of the ion.

The accelerating voltage V_i determines the energy of the ion in the field and the ratio of mass to charge is given by Eq. (43.9)

$$\frac{m}{q} = \frac{B^2 r^2}{2V_i} .$$

In the case of the mass spectrometer, r is the radius of the path of the ion which is steered by the magnetic field B into the slit of the collector. This radius is, of course, determined in manufacture. The beam of ions can be scanned across the slit by changing either the value of the field B or of the accelerating voltage V_i. The ion current passing the collector slit is amplified before being recorded. For an elementary introduction to the techniques see Moyes (1971).

49.9 BETA-PARTICLE EMISSION

Electrons emitted from radioactive nuclei are known as β-particles. Unlike α-particles which are emitted from the parent nucleus with a definite speed, the β-particles are emitted with a continuous range of

speeds (and therefore energies) from zero up to a maximum. This maximum varies with the nucleus involved. These observations present two problems. In the first place electrons do not appear to be constituent parts of any nucleus and on theoretical grounds cannot, in fact, exist within the nucleus. In the second place the existence of an energy spectrum is difficult to reconcile with the principles of conservation of energy and momentum.

Maybe an explanation for the first problem could be found in the suggestion that a neutron can itself decay into a proton and an electron which, becoming the emitted β-particle, carries away $-e$ charge units. In this case the extra proton within the nucleus causes the atomic number of the daughter product to be one greater than the parent and so the daughter has the chemical properties of an element one place higher in the periodic table. The mass number remains the same, for all that is lost is the tiny mass of the ejected electron. A typical decay might then be written

$$^{214}_{82}\text{Pb} \rightarrow {}^{214}_{83}\text{Bi} + e^-,$$

or in general terms

$${}^{A}_{Z}X \rightarrow {}_{Z+1}{}^{A}Y + e^-,$$

where X and Y represent the two atoms involved, A is the mass number and Z the atomic number of the parent nucleus.

This simple explanation suggests that the emitted electrons, all of which result from the same process, should have identical energies. But, as we have seen, this is not so. It appears that beta decay is a process in which momentum and energy may not be conserved.

In a bold effort to solve the second problem and, at the same time, to retain the laws of conservation of momentum and of energy, W. Pauli suggested that another particle must be involved. To fit the observations, this hypothetical particle must have no charge, very little or even zero rest mass and the ability to carry energy. It was Enrico Fermi who named it "the little neutral one" or *neutrino*. When a low energy β-particle is emitted, the neutrino carries a large amount of energy; a high energy β-particle is accompanied by a low energy neutrino. For a given nuclide the total energy is constant, the neutrino having "balanced the books."

Today we know that such particles exist in two forms—neutrinos and antineutrinos. With such properties they are extremely difficult to detect experimentally but this was achieved by C. Cowan and F. Reines in 1956, twenty-five years after Pauli had suggested their existence.

The equation for the decay of the neutron may now be written

$$n \rightarrow p + e^- + \tilde{\nu}$$

and that for the decay of the nuclide becomes

$$^{214}_{82}\text{Pb} \rightarrow {}^{214}_{83}\text{Pb} + e^- + \tilde{\nu},$$

or in general terms

$${}^{A}_{Z}X \rightarrow {}_{Z+1}{}^{A}Y + e^- + \tilde{\nu},$$

where e^- is the electron (β-particle) and $\tilde{\nu}$ the neutrino. In fact the particle involved is the anti-neutrino and this is indicated by the symbol $\sim$ above the symbol for the particle. In these equations the neutrino ensures that sufficient energy and momentum are carried away for the conservation laws to apply. Historically this incident provides an unusual example of a successful attempt to save a physical law: an attempt which remained an act of faith for 25 years.

49.10 GAMMA RADIATION

Gamma radiation, it must be stressed, does not consist of particles but of electromagnetic photons of nuclear origin. The wavelength is very short, being in the range 10^{-10} m to 10^{-13} m.

Many models of the nucleus have been proposed; all agree that the structure cannot be a rigid one and that it may suffer excitations, and oscillations or changes of shape rather like a blob of mercury that vibrates after being disturbed. As a result of the disturbance caused in a parent nucleus when it emits an α- or β-particle, the daughter nucleus may be formed in an excited state. In returning to a stable or ground state it emits the excess energy as gamma radiation. Some daughter nuclei are formed which are already in the stable ground state and in this case an

α- or β-particle is emitted without any accompanying gamma radiation.

It may be thought that if gamma radiation can remove energy during β-emission, then the introduction of the neutrino is unnecessary. However, this is not the case since some β-emitters produce particles with a range of energies and yet do not radiate γ-rays at all. Also if the neutrino did not exist, then when gamma rays are emitted they would have a large energy when the β-energy was small, and vice versa; since β-particles have a wide range of energy then γ-rays should have a wide range too. However, such a continuous spectrum of gamma energy is not observed during β-emission.

One other important process which leads to the emission of gamma radiation is known as "*K*-electron capture." In this the positive nucleus captures one of the orbital electrons from the innermost, *K*, shell. Consequently a proton is converted into a neutron in the nucleus. This causes the atomic number to decrease by one, and may leave the resulting nucleus in an excited state. It then drops to the ground state with the emission of a gamma ray. When the vacant place in the *K*-shell is filled by another electron X-rays are also emitted. These X-rays may be recognized as they have wavelengths which are characteristic of the element that has been formed.

49.11 THE BEHAVIOUR AND DETECTION OF NEUTRONS

The penetrating power of a particle is governed by how much energy it loses at each interaction as it passes through matter. The discussion in Chapter 7 reached the conclusion that when two masses collide elastically (one being initially at rest) the moving one loses the greatest energy when the two masses are equal. Such is the case when a neutron hits a hydrogen nucleus; if the collision is head-on, the neutron comes to rest and the hydrogen nucleus moves off with the speed and energy of the neutron. See Fig. 49.1, I and L. If the mass of the struck particle is greater than 1 atomic mass unit then the transfer of energy is not so complete. If a helium atom is involved, the neutron only loses 64% of its energy. (Calculated from the equations of Chapter 7.) If lead, having a mass number of 208, is used, only about 0.02% of the energy is lost per collision. Consequently, lead is a very poor absorber of neutrons; the neutrons are simply scattered by the lead, bouncing off the atoms without losing much energy.

On the other hand, substances containing hydrogen, (wax, water, polythene) or substances that have low mass numbers, are very effective absorbers of neutron energy. Because of its high hydrogen content and its ready availability, water is a commonly used neutron shield for nuclear reactors.

The difficulty of detecting an uncharged particle is that is does not itself cause the ionization which is required for the functioning of our usual detectors, e.g., cloud chambers, ionization chambers, scintillators, geiger tubes, etc. The problem is solved if neutrons can knock-on particles that can then produce ionization. Consequently, ionization and cloud chambers that are filled with hydrogen, or even nitrogen, will be able to detect the presence of neutrons. The ionization is not that produced by a neutron, but is that caused by one indirectly. This is why the ionization chambers used in the discovery of the neutron were able to detect the penetrating beryllium radiation.

Another method for detecting neutrons depends on the nuclear reaction that occurs when they are allowed to bombard boron. The reaction is:

$$^{10}_{5}\mathrm{B} + {}^{1}_{0}n \rightarrow {}^{7}_{3}\mathrm{Li} + {}^{4}_{2}\mathrm{He}.$$

Both the lithium nucleus and the α-particle can be detected by the ionization they produce.

This reaction is an example of the way in which transmutation of an element can occur. Because they have no charge, neutrons experience no repulsive force as they approach a target nucleus, which they can enter with comparative ease. (Recall the scattering of an α-particle as it approaches a gold nucleus.) Consequently, neutrons are especially effective particles for the purpose of transmuting elements. Indeed it is now possible to manufacture gold atoms by bombarding mercury with neutrons. The reaction is:

$$^{1}_{0}n + {}^{198}_{80}\mathrm{Hg} \rightarrow {}^{197}_{79}\mathrm{Au} + {}^{2}_{1}\mathrm{H}.$$

(The other atom produced is an isotope of hydrogen

named deuterium, the nucleus of which is called a deuteron.) It must, of course, be realized that the amount of gold produced by this reaction is very small, and a suitable source of neutrons very costly.

49.12 MODERN ALCHEMY—A DREAM COME TRUE

Since the second century alchemists had dreamt of discovering a transmuting agent they called the "Philosophers' Stone." With this it would be possible to turn base metals into gold and produce "the Elixir of Life"—a perfect medicine that would provide everlasting life for mankind. It is an interesting coincidence that they considered mercury to be one of the necessary ingredients of all matter, and often had in their workshops large vats of the metal which was not then particularly valuable as it had few, if any, uses.

Having caused a transmutation to gold, there is a distinct parallel between the neutron and the Philosophers' Stone; but what of the Elixir of Life? Perhaps this can be compared to the peaceful uses of the release of nuclear energy made possible with the advent of the neutron. This source of energy must become more and more important to the continued existence of man, as the world's supply of fossil hydrocarbon fuels approaches exhaustion in the twenty-first century. These fuels are being consumed at an enormous rate; to keep a man in comfort requires an annual expenditure of energy, ranging from the equivalent of 8×10^3 kg of coal in America, to about 3×10^3 kg in Europe. Even to stay alive in one of the underdeveloped parts of the world requires about 400 kg per man. The need for alternative sources of energy from nuclei becomes apparent when one remembers that the world population is not only large, but it is increasing, while the stocks of traditional fuels are running down.

49.13 EPILOGUE

During this case history on the development of the structure of the atom two things should have emerged. First, as models are developed and tested only the fittest survive, but that these often raise more questions than they answer; so the quest for knowledge continues and is regenerative. This is true with the model of the atom developed so far, for this stage is just the beginning of a further series of investigations into the structure of matter that continues to this day. Even with man's present knowledge of wave mechanics and relativity, the end is not in sight. The English quantum physicist Paul Dirac wrote

> "I feel very strongly that the stage physics has reached at the present day is not the final stage. It is just one stage in the evolution of our picture of nature, and we should expect this process of evolution to continue in the future, as biological evolution continues into the future. The present stage of physical theory is merely a stepping stone toward the better stages we shall have in the future. One can be quite sure that there will be better stages simply because of the difficulties that occur in the physics of today." (Dirac, 1963.)

The second point to emerge is that science is the work of many. Quoting an old saying, Sir Isaac Newton said "If I have seen further than others, it is by standing on the shoulders of giants."

This case history began with a quotation about Rutherford; let it end with a sentence he spoke in the course of his last public lecture in 1937:

> "Scientists are not dependent on the ideas of a single man, but on the combined wisdom of thousands of men, all thinking of the same problem, and each doing his little bit to add to the great structure of knowledge which is being gradually erected."

49.14 NOTES ON THE CLOUD CHAMBER PHOTOGRAPHS

Photograph A. This is approximately three times the actual length of the track. The gas in the chamber was air. The track was one of the first ever photographed and the sharpness of it is a tribute to the skill of the inventor of the cloud chamber, C. T. R. Wilson.

A radon-222 atom, at the bottom of the photograph, has ejected an α-particle. The residual polonium-

218 atom recoils and it too produces some ionization; hence the tiny dot at the start of the track. The number of ions, and therefore the number of water droplets produced per cm, increases as the α-particle slows down towards the top of the photograph and spends longer in the vicinity of gas atoms. The increase of interaction time as the speed falls also explains why α-particles so often undergo a large deflection towards the end of their tracks. By looking, in the plane of the page, along the track, two smaller deviations can be seen.

It may just be possible to see tiny projections from either side of the main track especially in its first half. These are better illustrated in photograph B. (Wilson, 1912.)

Photograph B. This is a section of two tracks magnified approximately 80 times; the chamber gas is air.

When a gas atom is ionized, the electron ejected can have sufficient energy to produce further ionization. Such electrons cause tracks, sometimes called delta rays, that can be seen coming from either side of these two α-particle tracks. The tiny blob at the termination of such a track again illustrates the increase of ionization as the charged particle slows down. (Alper, 1932.)

Photograph C (about 4 times full size). Alpha-particles have passed through a chamber filled with hydrogen. One of the alphas has undergone an elastic collision with a hydrogen nucleus, and the knock-on proton leaves a track that is long and which does not display so much ionization per centimetre as an α-particle track. The reason for this is discussed in Section 49.1. The angles made by the branches of the fork with the original direction of the α-particle are consistent with the ratio (mass of alpha-particle/mass of proton) being equal to 4. (Blackett and Lees, 1932.)

Photograph D (5 times full size). This is similar to photograph C except that the chamber is filled with helium and the arms of the fork are at right angles showing that the particles involved have equal masses. (Feather, 1932.)

Photograph E (about full size). A radioactive source is producing mono-energetic α-particles. One half of the source is covered with a thin film of paraffin wax which reduces the track length. The long track from this half is made by a knock-on proton. The wax is rich in hydrogen atoms and an α-particle has collided with one of these. The proton track is longer than an unhindered α-track for the reasons discussed in Section 49.1. (Meitner, 1927.)

This experiment can easily be repeated with a simple diffusion chamber containing a 5μCi americium source covered with a piece of thin polythene sheet, which, like the wax, is also rich in hydrogen.

Photograph F (about 3 times full size). An α-particle is captured when it strikes the nucleus of what is thought to be a nitrogen atom, and the long branch is an example of an ejected proton travelling backwards, i.e., with a velocity component towards the alpha source. The short branch of the fork is attributed to a recoiling oxygen nucleus which results from transmutation. The doubt over the elements involved arises from the fact that the chamber was not filled with nitrogen at the time but with a mixture of 50% argon, 10% oxygen, and 40% hydrogen. The tracks, however, resemble other cases involving nitrogen and it was possible that the argon gas cylinder contained nitrogen as an impurity. Note, again, the increased ionization as the proton slows down. (Blackett and Lees, 1932.)

Photographs G and H (just under full size). These typical α-tracks from radon-220 as seen in a diffusion cloud chamber, help to illustrate the idea of half-life and of natural transmutation. Photograph H was taken 2 minutes after G and clearly there is less activity present; the activity is reduced by about 4 in this time as the half-life is 55 s.

In H there is a good example of two tracks coming from the same point to form a V; two more examples can be seen in G. At the point of each V the transmutation ${}^{220}_{86}\mathrm{Rn} \rightarrow {}^{216}_{84}\mathrm{Po} \rightarrow {}^{212}_{82}\mathrm{Pb}$ has occurred (see Section 49.3). Note the gaps caused in a "young" track when it crosses "older" tracks that have already condensed the available vapour. (Photographs by B. Taylor.)

Photograph I. A thin sheet of paraffin wax is supported in the chamber by two wires. Outside the chamber (below the picture) a strong polonium

source bombards beryllium with α-particles and produces neutrons. These penetrate the chamber but, being uncharged, leave no track. However, on hitting the wax they may eject long range protons. Such a proton track is shown; the start of it is missing, possibly distorted by air currents near the wax. (Curie and Joliot, 1932.)

Photograph J (3 times full size). Neutrons, which leave no track, enter a nitrogen-filled chamber (from below the picture) and knock-on a nitrogen nucleus which, being charged, does leave a track as it recoils from the collision. Actual track length is only 3.5 mm. (Feather, 1932.)

Photograph K (3 times full size). This is another photograph taken with the arrangement used for J. This time the knocked-on nitrogen nucleus has itself collided elastically with a stationary nitrogen atom of the chamber gas. The 90° angle between the forward branches is characteristic of an interaction between equal masses, one being initially at rest, and provides evidence that the particles recoiling from the neutron bombardment are nitrogen nuclei. (Feather, 1932.)

Photograph L (about $\frac{1}{3}$ full size). Neutrons enter, unseen, a hydrogen-filled chamber and their presence is revealed by the protons that are knocked-on elastically producing ionization and tracks. (Hamouda, Halter and Scherrer, 1951.)

Photograph M (full size). This photograph, like F, shows nitrogen being transmuted, but unlike F, the missiles causing this are not visible in the chamber. This indicates that they carry no charge and are neutrons. Only charged particles can interact with the electrons of a gas atom and cause ionization.

As was mentioned in Section 49.11, neutrons can enter a nucleus with ease, compared with an α-particle, a fact noticed by Feather while helping Chadwick to verify the existence of the neutron. On 1740 stereo pairs of photographs Feather found 100 elastic collisons (e.g., J and K) and 30 interactions that re-resulted in transmutation as in M. (When α-particles are used for transmutation (e.g., F) only about 1 in 1000 are successful, the remaining collisions being elastic.)

Photograph M is remarkable for it shows two kinds of transmutation. The upper pair of tracks are those of a proton and a carbon-14 nucleus [$^{14}_{7}\mathrm{N}\ (n, p)\ ^{14}_{6}\mathrm{C}$], while the lower pair are of an α-particle and a recoiling boron nucleus [$^{14}_{7}\mathrm{N}(n,\alpha)^{11}_{5}\mathrm{B}$]. Note the great ionization density along the α-track compared with the relatively sparse ionization of the proton track. (Chadwick, Feather and Davies, 1934.)

Photograph N ($\frac{1}{2}$ full size). This illustrates neutron-induced transmutation of a heavy nucleus and, because the nucleus splits into two roughly equal parts, the process is called nuclear fission. The chamber contains air and has across its middle a screen of gold foil coated with uranium. Neutrons bombarding the foil have caused a uranium nucleus to undergo fission. The total energy released is large, being of the order of 200 MeV which appears mostly as kinetic energy of the fission products (see Problem 2.19). The two "heavy" fragments are charged and leave tracks that can be seen travelling in opposite directions from the sides of the foil. As each carries as many as 50 proton charges the ionization density is very high and the tracks' ends show branches indicating collisions with nuclei of the chamber gas. At the moment of fission, 2 or 3 neutrons are emitted and if at least one of these can be induced to produce fission of another nucleus, a chain reaction results which is self-sustaining. This has not occurred in N, the other tracks being due to recoil nuclei and transmutations caused by neutron bombardment of the gas in the chamber. (Bøggild, 1940.)

Photograph O. This is a Wilson-type expansion cloud chamber designed principally by Dr. Chadwick, and Dr. Nimmo who prepared this drawing. This chamber was used for taking photographs C, D, F, J, K, M. The top and sides (*g*) of the chamber are of glass (unshaded) and are easily removed by separating a greased joint (*j*). A piston (*p*) moves in a cylinder (*c*); liquid from the reservoir (*l*) on the side fills a groove cut in the inside surface of the cylinder and so provides a gas-tight seal between the piston and cylinder. The space (*b*) below the piston can be connected suddenly to a vacuum through the side tube (*v*), the piston descends and this allows the chamber gas in the space (*a*) above the piston to expand rapidly, cool, and

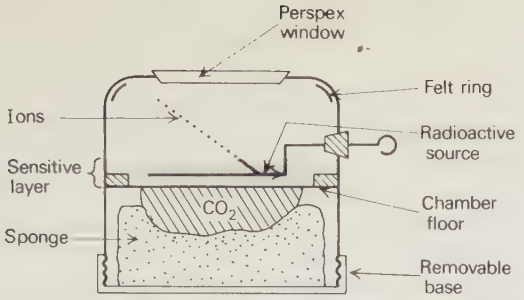


Fig. 49.6 Simple diffusion cloud chamber.

produce condensation on any ions present. An electric "clearing" field is established between the glass top of the chamber and the piston, to remove unwanted ions that may be produced before the expansion takes place; these would otherwise fill the chamber with a fog of water droplets. A spring (*s*) around the piston rod (*r*) raises the piston once (*b*) is re-connected to the atmosphere. Below the end of (*r*) is an oil dash-pot (*d*) which prevents the piston from bouncing and oscillating at the completion of its initial downward movement. Adjustable nuts (*n*) on the piston rod act as a stop to limit the piston's upward travel and to enable a variety of expansion ratios to be used. (Feather, 1933.)

A disadvantage of the expansion chamber is that it is only sensitive for about 0.2 sec after the expansion and then it takes several seconds for it to be re-cycled before another expansion can take place. A type of chamber that permits continuous observation of particle tracks is the diffusion cloud chamber, first suggested by Langsdorf in 1939. It relies on solid CO_2 to provide cooling, rather than on the expansion of a gas. A simple version of a continuous cloud chamber is shown in Fig. 49.6. The transparent case is divided horizontally by a thin black metal plate that forms the floor of the chamber. In the lower half is placed solid CO_2 that is pressed into contact with the underside of the floor by a piece of foam plastic sponge. The top half of the case is the chamber proper and near the Perspex viewing window is a ring of felt soaked with alcohol. While the top of the chamber is at room temperature, say 20° C, the floor is at −78° C, and alcohol vapour evaporating from the felt diffuses downwards. As it does so the gas and vapour is cooled and therefore the degree of saturation of the gas in the chamber increases. At about 1 cm from the floor the saturation is sufficient for condensation to occur on the ions left in the wake of the radiation from a radioactive source, and tracks can be seen in this sensitive layer.

A vertical electric field must be established (by charging the Perspex with a wool cloth) but its function is not that of a clearing field. It is needed to pull ions quickly into the sensitive layer before they have had time to diffuse. Consider first an α-particle that travels horizontally through the sensitive layer; each of the ions it produces will immediately be surrounded by a droplet of alcohol (there are about 10^5 ions per track). Each droplet is of considerable mass compared with the mass of the ion on which it is formed. It falls, with neighbouring droplets, under gravity to the floor of the chamber; there is no relative motion of the droplets because of their great inertias. Thus tracks in the sensitive layer are thin and well defined. However, if an α-particle moves at, say, 45° to the floor, it will first travel up through the sensitive layer where it leaves a sharp track of droplets, but above the layer, a line of ions will be formed that have not yet collected droplets. Because they are "bare" ions they have little mass and can separate easily from one another, consequently, what was a thin line of ions soon becomes wide and diffuse. If an electric field pulls freshly formed ions down rapidly into the sensitive layer before spreading occurs, a sharp track will result. It must be remembered that the consequence of this is that an observed track length is the projection of the true length on to the horizontal plane of the chamber floor.

PROBLEMS: Unit Nine

9.1 Read through Lucretius's description of the behaviour of atoms in Section 47.2 and rewrite it in current scientific language. Discuss a possible connection between the end of the last sentence (. . . "and blazing sunlight"), spectra, and energy levels.

9.2 Imagine you are standing at the edge of an atom (where the outermost electrons are) and are looking in towards the nucleus. Would it appear to be about the same size as an apple held, (a) 3 m; (b) 30 m; (c) 300 m; (d) 3000 m; away?

9.3 How many times bigger than a real nucleus is the dot shown in Fig. 48.6?

9.4 Find, from the text:

a) the mass number of platinum;
b) the mass equivalent to 1 atomic mass unit;
c) the radius of a platinum nucleus. Discuss briefly what is meant by "radius" in this context.

Calculate:

d) the mass of a platinum nucleus;
e) the volume of a platinum nucleus;
f) the density of a platinum nucleus.

If the density of a sample of platinum is 21×10^3 kg m^{-3}, calculate:

g) the volume of a platinum atom;
h) the radius of a platinum atom.

Compare your answers to (c) and (h), and comment on (f).

9.5 A "target" consists of 200 identical nails hammered into the surface of a piece of soft wood, 2 m square. The nail heads act as hard objects that can scatter small bullets fired at the target; those hitting the wood pass through undeflected. Calculate the effective area of a nail head if, out of a total of 10^5 bullets sprayed uniformly over the target, only 500 are deflected.

9.6 A piece of foil measures 0.01 m × 0.01 m and is 2×10^{-6} m thick. It is made of a metal each of whose atoms can be contained in a cube of side 2×10^{-10} m. Calculate:

a) the number of atoms in the surface layer;
b) the number of layers that make up the foil's thickness;
c) the number of nuclei in the foil.

When α-particles bombard the foil, it is found that 1 in 2500 are deflected through an angle greater than 53°. What does this tell you about:

d) the aiming error, p, for these deflected particles;
e) the ratio: $\dfrac{\text{area of the foil}}{\text{total target area of the nuclei}}$,

(Assume that the nuclei are not one behind the other in the line of fire of the particles.) Hence calculate:

f) the target area of a nucleus;
g) an estimate of the "radius" of a nucleus.

9.7 In an α-particle scattering experiment, 1 in 10 000 are scattered through more than 20°. How many would be scattered through more than 20° if:

a) the thickness of the foil was trebled;
b) the original thickness was used but the α-particles have twice the velocity;
c) again the original foil is used but protons bombard the foil that have the same velocity as the original α-particles?

9.8 In an α-particle scattering experiment, how many times more particles would be scattered through 60° than through 120°?

9.9 An experiment is performed with a beam of α-particles that can bombard each of three foil targets in turn. When the gold foil is used, 1400 particles are scattered through 120° in 10^4 s. For the same angle and time, foil *X* scatters 500 particles, and *Y*, 40. The three foils are of such thicknesses that, for each foil, the beam passes the same number of atoms per unit surface area. Calculate the atomic numbers of *X* and *Y*, and suggest a material from which each could be made.

9.10 Calculate the force acting when the centres of an α-particle and a platinum nucleus are separated by, (a) 10^{-13} m, and (b) 10^{-14} m. Compare the answers you obtain with the weights of familiar objects, e.g., a grain of sand, a book, a bicycle, a car.

9.11 What is the acceleration of an α-particle that is at a distance of 10^{-13} m from a platinum nucleus? At what distance would the acceleration of the particle be equal to 10 m s^{-2}? Assume that the only forces acting are due to the charges on the two particles named.

9.12 Calculate the gravitational attraction between an α-particle and a platinum nucleus when their centres are separated by 10^{-14} m. Let $G = 6.7 \times 10^{-11}$ N m^2 kg^{-2}.

Show that the ratio:

$$\frac{\text{coulomb force}}{\text{gravitational force}}$$

has a value of 2.5×10^{35} irrespective of separation.

9.13 What is the energy, expressed in both joules and in MeV, of an α-particle having a speed of 2×10^7 m s^{-1}?

9.14 Calculate the energy, in MeV, of a proton having a speed of 2×10^7 m s^{-1}.

9.15 An α-particle has an energy of 5 MeV. What is its closest distance of approach to a gold nucleus?

9.16 A 5 MeV α-particle and a neutron are initially travelling on the same path towards a gold nucleus. The neutron is undeflected and it misses the gold nucleus by 2×10^{-14} m. Through what angle is the α deflected?

9.17 Estimate the energy per kilogramme of a bullet, and compare it with the energy per kilogramme of an α-particle.

9.18 Explain why J. J. Thomson's plum pudding model of the atom could not account for the occasional large angle of deflection of α-particles that are bombarding a metal foil.

9.19 A proton and a deuteron each have the same energy and are on paths that will lead to a head-on collision with a silver nucleus. Compare their closest distances of approach to the nucleus.

9.20 Goldbeating is an ancient art, being mentioned by Homer and by Pliny who states that 1 ounce (0.02835 kg) of gold can be extended to 750 leaves, each leaf being 4 fingers (about 0.08 m) square. Use this information to calculate the thickness of gold foil, and compare your answer with the wavelength of visible light. Take the density of gold to be 19 300 kg m^{-3}.

9.21 An α-particle with a speed of 10^7 m s^{-1} undergoes elastic collisions with the following stationary particles:

a) an electron;
b) the nucleus of a hydrogen atom;
c) the nucleus of a nitrogen atom.

What is the maximum possible speed of the struck particle in each case, and what percentage of the α-particle's original energy is transferred?

9.22 How much energy does a neutron lose in a head-on collision with:

a) a hydrogen nucleus;
b) a nitrogen nucleus;
c) the nucleus of a lead atom?

What are the desirable properties of a substance to be used as a shield for a beam of neutrons? Explain.

9.23 Why is a neutron difficult to detect? Explain why an ionization chamber used for detecting neutrons is commonly filled with the gas boron trifluoride. (See also Problem 9.45.)

9.24 Since neutrons do not carry any charge, how was it possible for the Curie-Joliots to detect them with an ionization chamber?

9.25 Read Section 49.4. What gas would you have recommended Webster and Chadwick to use in their cloud chamber? Explain the reason for your choice.

9.26 How are neutrons produced, what are their properties, and for what purpose are they particularly useful? How can neutrons be stopped if they don't lose energy by electrical interaction with charged particles?

9.27 Why do the tracks of the fission products shown in photograph N (Fig. 49.1) have branches at the end and

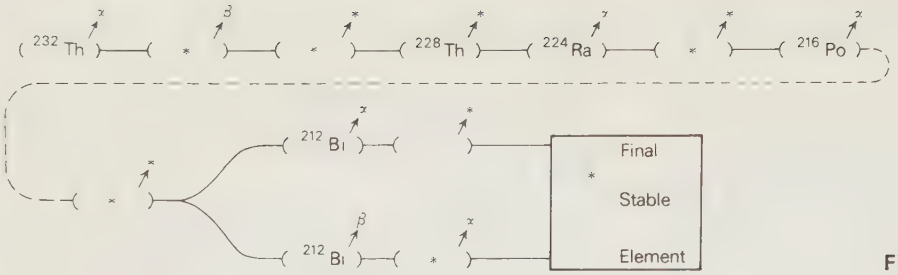


Fig. P9.1

not at the beginning? To help you answer this, read the notes for photograph A.

9.28 A 6 MeV α-particle has a range in air (at s.t.p) of 4.7 cm. In this distance it loses its initial energy by ionizing atoms in its path. If it loses 30 eV at each encounter, how many atoms does it ionize before coming to rest?

Explain why an α-particle travels further if:

a) the pressure of the air is reduced;
b) a gas such as hydrogen is used instead of air.

9.29 A 5 μCi alpha source is placed so that all the particles it emits pass into an ionization chamber containing air which brings the particles to rest. The ionization current, measured with a d.c. amplifier, is 2.96×10^{-9} A. Calculate the number of ions being produced per second, and the number of ions produced by one α-particle. If the ionization potential is 30 eV, calculate the initial energy of an α-particle.

9.30 Refer to Section 49.2. Approximately how long did Blackett's experiment take—days? weeks? months?

9.31 12 kg of the isotope carbon-12 contain 6.023×10^{26} atoms and 1 atomic mass unit (u) is $\frac{1}{12}$ of the mass of a single ^{12}C atom. Use this data to show that 1 u = 1.660×10^{-27} kg.

9.32 What nuclides result if:

Polonium 215 emits an α-particle;
Thallium 207 emits a β-particle;
^{227}Ac emits an α-particle;
^{223}Fr emits a β-particle?

9.33 Supply the missing data (indicated in Fig. P9.1) in the thorium series which is one of the three radioactive decay series that occur in nature. Note that there are two alternative paths towards the end; 36% of the atoms take the upper one. The final stable element is the same whichever path is taken. The upward arrows indicate the particles ejected.

9.34 What is the half-life of a radioactive substance whose activity decays to $\frac{1}{8}$ of the original value in 150s?

9.35 What percentage of the original nuclei of a sample of a radioactive substance are left after

a) 2 half-lives;
b) 5 half-lives;
c) 10 half-lives;
d) n half-lives?

9.36 Radon-220 is a radioactive gas with a half-life of 54 s. What fraction of the mass of a sample of this gas remains as Radon after

a) 108 s;
b) 216 s;
c) 540 s?

9.37 A radioactive isotope, ^{124}Sb, having an initial activity of 2 m Ci has a half-life of 60 days. If it decays for a year, what activity remains?

9.38 The half-life of ^{226}Ra is 1620 years. How long does it take for 10% of a sample to decay?

9.39 A sample of a radioisotope is found to give, at half hourly intervals, the following counts per minute on a scaler: 360, 350, 336, 325, 312, 303, 295, 282, 275. What is its half-life and its decay constant?

9.40 The activity of carbon found in living specimens is 0.007 μCi per kg, due to the ^{14}C present. The charcoal taken from the fire pit of an Indian camp site has an activity of 0.0048 μ Ci kg^{-1}. The half-life of ^{14}C is 5760 years. Calculate the year the camp site was last used.

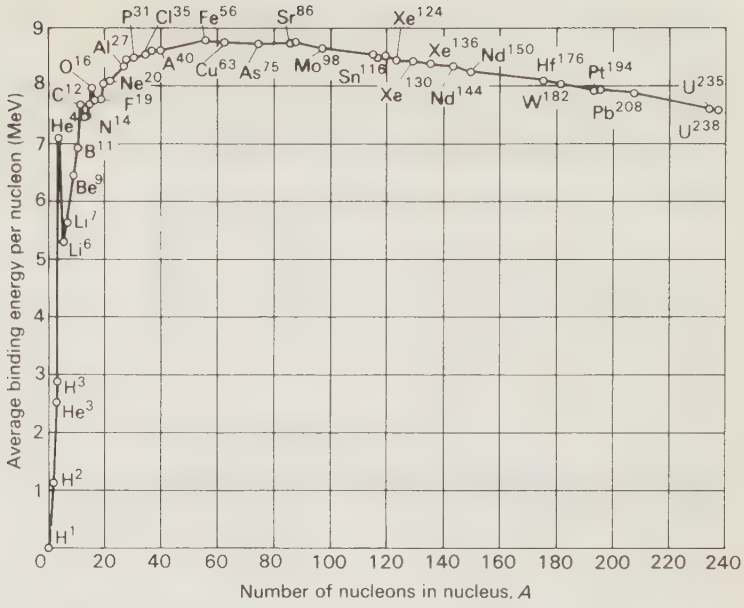


Fig. P9.2 Average binding energy per nucleon as a function of the number of nucleons, A. (From G. Holton and D. Roller, *Foundations of Modern Physical Science*, Addison-Wesley, 1958.)

9.41 Deduce the relationship between half-life and decay constant.

Calculate the half-life of ^{226}Ra if 1 kg of radium emits 3.7×10^{13} α-particles per second and the mass of the radium atom is 3.8×10^{-25} kg.

9.42 Make a copy of the following table and fill in the mass number A, the atomic number Z, the number of protons, P, and the neutron number N.

	^{1}H	^{2}H	^{4}He	^{9}Be	^{12}C	^{27}Al	^{206}Pb	^{226}Ra	^{238}U
A									
Z									
P									
N									

9.43 Complete the following reactions, express each in the abbreviated notation given in Section 49.2, and state each reaction in words:

a) $^{27}_{13}\text{Al} + {}^{4}_{2}\text{He} \rightarrow {}^{1}_{1}\text{H} + (\quad)$;
b) $^{11}_{5}\text{B} + {}^{4}_{2}\text{He} \rightarrow {}^{14}_{7}\text{N} + (\quad)$;
c) $^{24}_{12}\text{Mg} + (\quad) \rightarrow {}^{24}_{11}\text{Na} + {}^{1}_{1}\text{H}$;
d) $^{27}_{13}\text{Al} + {}^{1}_{0}n \rightarrow {}^{28}_{13}\text{Al} + (\quad)$.

9.44 Express the following processes in words:

a) $^{1}_{0}n \rightarrow {}^{1}_{1}p + e^{-} + \bar{\nu}$;
b) $^{27}_{12}\text{Mg} \rightarrow {}^{27}_{13}\text{Al} + e^{-} + \bar{\nu} + \gamma$ ($T_{1/2} = 9.5$ min).

9.45 Complete the following reactions, write each as an equation, and state each one in words. (* = missing data; d = deuteron.)

a) $^{198}\text{Hg}(n, d)^{*}$;
b) $^{10}\text{B}(n, *)^{7}\text{Li}$;
c) $^{*}(n, \alpha)^{11}\text{B}$;
d) $^{196}\text{Pt}(n, *)^{197}\text{Pt}$.

9.46 Show that 1 u of matter is equivalent to 931 MeV of energy, and calculate the energy available if one electron is annihilated.

9.47 a) In a helium nucleus there are two protons and two neutrons; if the protons are separated by 10^{-14} m, what is the electric force of repulsion between them?
b) Calculate the binding energy per nucleon if the mass of an atom of helium $^{4}_{2}$He is 4.00260 u.

9.48 Using Fig. P9.2 which gives binding energies *per nucleon*, find the binding energies for ^{235}U, ^{141}Ba, and ^{92}Kr. Use

these binding energies to show that when the fission of ^{235}U occurs, about 200 MeV of energy is released.

9.49 In a mass spectrometer, hydrogen ions or protons are given an energy of 10 keV and they then enter a uniform magnetic field of 0.5 T in a direction perpendicular to the field. Calculate the radius of the path of these protons in this field.

The hydrogen in the ion source will also contain deuterium which will also become ionized. Calculate the radius of the path of the deuterons. (Mass of a proton = 1.66×10^{-27} kg, charge on a proton = 1.6×10^{-19} C.)

BIBLIOGRAPHY: Unit Nine

In addition to the collections of original sources and other references given in the general bibliography you may wish to consult

Andrade, E. N. da C. (1965). *Rutherford and the Nature of the Atom*. Science Study series, Heinemann.

Baez, A. V. (1967). *The New College Physics*, Freeman.
A 'spiral' approach to Physics culminating in the Bohr theory of the hydrogen atom.

Caro, D. E., McDonell, J. A., and Spicer B. M. (1971). *Modern Physics*, Arnold.
A straight-forward, appealing account.

Frisch, D. H., and Thorndike, M. (1964). *Elementary Particles*, van Nostrand.

Gamow, G. (1961). *The Atom and its Nucleus*, Prentice-Hall.

Gamow, G. (1965). *Mr Tompkins in Paperback*, Cambridge University Press.

Lewis, J. L., and Wenham, E. J. (1970). *Radioactivity*, Longman Topic series.

Lipson, H. S. (1968). *The Great Experiments in Physics*, Oliver and Boyd.

Massey, H. (1960). *The New Age in Physics*, Elek Books.

Romer, A. (1961). *The Restless Atom*, Science Study series, Heinemann.

Schonland, B. (1968). *The Atomists* 1805 1933, Oxford University Press.
A brief history of the ideas and of the men and women who contribute to them.

Specific text references include:

Becquerel, A. H. (1896). *C. r. hebd. Séanc. Acad. Sci., Paris*, **122**, 501.

Blackett, P. M. S. (1925). *Proc. R. Soc.*, **A107**, 349.

Dirac, P. A. M. (1963). *Scient. Am.*, May 1963, 45–53.

Moyes, R. B. (1971). *Sch. Sci. Rev.*, **52**, 879–890.

Röntgen, W. C. (1896). *Nature, Lond.* **53**, 274–277.

Rutherford, E. (1914). *Phil. Mag.*, **27**, 493.

Rutherford, E. (1919). *Phil. Mag.* **37**, 578–587.

Soddy, F. (1913). *Nature, Lond.* **92**, 331.

Thomson, J. J. (1897). *Phil. Mag.* (5), **44**, 297.

Thomson, J. J. (1907). *The Corpuscular Theory of Matter*, Cambridge University Press

Thomson, J. J. (1913). *Proc. R. Soc.*, **A89**, 1–20.

References for the cloud chamber photographs:

Alper, T. (1932). *Z. Phys.*, **67**, 172.

Blackett, P. M. S. and Lees, D. S. (1932). *Proc. R. Soc.*, **A136**, 338.

Boggild, I. K. (1940). *K. danske Vidensk. Selsk. Mat-fys. Medd.*, **18**.

Chadwick, J., Feather, N., and Davies, W. T. (1934). *Proc. Camb. phil. Soc. Math. phys. Sci.*, **30**, 357.

Curie, I. and Joliot, F. (1932). *C. r. hebd. Séanc. Acad. Sci., Paris* **194**, 867.

Feather, N. (1932). *Proc. R. Soc.*, **A136**, 722.

Feather, N. (1933). *Proc. R. Soc.*, **A141**, 196–209.

Hamouda, J., Halter, J., and Scherrer, P. (1951). *Helv. phys. Acta.*, **24**, 217.

Wilson, C. T. R., (1912). *Proc. R. Soc.*, **A87**, 277.

SUMMARY OF TERMS: Unit Nine

Atomic number (*Z*) Number of protons in the nucleus (= number of electrons in the space surrounding the nucleus of a neutral atom).

Neutron number (*N*) Number of neutrons in the nucleus.

Mass number (*A*) $= N + Z =$ total number of nucleons (neutrons and protons) in the nucleus.

Nuclide A species of atoms identical as regards atomic number and mass number, written ${}^{A}_{Z}X$, for example ${}^{238}_{92}\mathrm{U}$.

Isotope Atoms having the same atomic number but different neutron numbers (and hence masses) are called isotopes. Having the same number of external electrons, they are chemically alike and are named after the commonly occurring member of the family, e.g., ${}^{31}_{15}\mathrm{P}$, ${}^{32}_{15}\mathrm{P}$ are isotopes of phosphorus.

Decay law The rate of decay of a radioactive element is proportional to the number of nuclei present, namely

$$\frac{dN}{dt} = -\lambda N.$$

Integration of this equation gives the alternative forms:

$$N = N_0\, \mathrm{e}^{-\lambda t} \qquad \text{or} \qquad \ln \frac{N}{N_0} = -\lambda t,$$

where λ = decay const, N = number of atoms present at time t, N_0 = number present at time 0.

Half-life ($T_{\frac{1}{2}}$) This is the time taken for half of the atoms initially present to disintegrate; it is the value of t for which $N = \frac{1}{2}N_0$.

To calculate N after a given time for a particular nuclide the following relationship is useful:

$$\frac{N}{N_0} = \left(\frac{1}{2}\right)^n \qquad \text{or} \qquad \frac{N_0}{N} = 2^n,$$

where n is time expressed as a number of half-life periods; $n = t/T_{\frac{1}{2}}$.

The activity of a sample of radioactive material is measured in curies. One curie is that quantity of any material in which 3.7×10^{10} atoms disintegrate per second, e.g., 1 g of ^{226}Ra. A convenient sub-division is the micro-curie; 1 Ci = 10^6 μCi.

Note that the value stated for the activity of a source will relate to a date and time which must be specified. The value at any later time can be calculated if the half life or the decay constant of the material is known.

Electron-volt One electron-volt (eV) is the change of kinetic energy of an electron that has moved through a potential difference of 1 V. 1 eV $\approx 1.6 \times 10^{-19}$ J.

Atomic mass unit 1 u $\approx 1.660 \times 10^{-27}$ kg. 1 u is one twelfth of the mass of a single atom of carbon-12 ($^{12}_{6}$C).

Mass and energy Einstein showed that a mass m is equivalent to an amount of energy, $E = mc^2$, where c is the speed of light (3×10^8 m s^{-1}). Thus 1 u of matter is equivalent to 931 MeV. Similarly the rest mass of an electron is equivalent to 0.51 MeV so that the production (or annihilation) of a negative and a positive electron pair involves the disappearance (or appearance) of 1.02 MeV of energy in the form of radiation.

Binding energy When a number of nucleons combine to form the nucleus of an atom, there is an apparent decrease in mass, e.g.:

10 neutrons have a mass of	10.08665 u,
10 electrons have a mass of	0.00549 u,
10 protons have a mass of	10.07276 u,
The total mass is	20.16490 u.

But an atom of the neon isotope $^{20}_{10}$Ne has a mass of 19.99244 u, an apparent mass defect of 0.17246 u. This mass defect corresponds to about 160 MeV, or 8 MeV per nucleon—a typical value for elements above carbon in the periodic table, see Fig. P9.2, page 517. This amount of energy must be supplied if the constituent particles are to be separated again. It thus represents the binding energy which holds the nucleus together.

Nuclear reactions These are described by equations analogous to chemical equations. For example, the first reaction artificially produced was:

$$^{14}_{7}\text{N} + ^{4}_{2}\text{He} \rightarrow ^{18}_{9}\text{F} \rightarrow ^{1}_{1}\text{H} + ^{17}_{8}\text{O}.$$

A simpler notation for the same reaction is:

$$^{14}\text{N}\ (\alpha, p)\ ^{17}\text{O}.$$

This shows the target and product nuclei, and in brackets the bombarding particle and the ejected particle (p = proton).

Many types of reaction are caused by neutron (n) bombardment, e.g., (n, α); (n, p); (n, γ); (n, f). The last represents fission.

Nuclear fission This is a nuclear reaction in which the nucleus of a heavy element (e.g., uranium) splits into two approximately equal parts releasing large amounts of energy and emitting 2 or 3 neutrons in the process.

Chain reaction Under suitable conditions, the neutrons emitted during fission can themselves initiate further fissions and the process becomes self-sustaining. Unless it is controlled, the reaction "runs away" and, if the material is confined, an explosion results.

Critical size If a small volume of uranium is undergoing fission, there is only a small chance that the emitted neutrons will produce further fissions before escaping from the mass of uranium. However, if a larger volume is used, a neutron has to pass more nuclei before escaping and it has a better chance of causing fission. Consider a sphere of uranium of radius r; the chance of fission occurring is proportional to the number of atoms present and so depends on r^3, whereas the chance of escape is governed by the surface area of the sphere and is proportional to r^2. A large value of r increases the probability of a chain reaction. The critical size is that for which the reaction is just self-supporting. Below this critical size, a chain reaction cannot be sustained.

Nuclear fusion This is the process in which nuclei of two light elements are united and the resulting particles have a smaller total mass; the difference in mass appears as energy. The problem of achieving fusion is that the particles must have sufficient kinetic energy to overcome the coulomb repulsion between them. For this very high temperatures are required. (See Problem 7.11).

APPENDICES

Appendix 1
Appendix 2
Appendix 3
Appendix 4
Appendix 5
Answers to Selected Problems
Glossary of Symbols
General Bibliography

APPENDIX 1

THE INTERNATIONAL SYSTEM OF UNITS (SI)

Adopted by the Conférence Générale des Poids et Mesures in 1960, this system is based on six basic units. A complete, coherent system of other units for other physical quantities can be derived from these six without the introduction of numerical factors. The six are:

Metre The metre is the length equal to 1 650 763.73 wavelengths in vacuum of the radiation corresponding to the transition between the levels $2p_{10}$ and $5d_5$ of the krypton-86 atom. (Symbol: m)

Kilogramme The kilogramme is the unit of mass; it is equal to the mass of the international prototype kilogramme. (Symbol: kg)

Second The second is the duration of 9 192 631 770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium-133 atom. (Symbol: s)

Ampere The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible cross-section, and placed 1 m apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} newton per metre of length. (Symbol: A)

Kelvin The kelvin, unit of thermodynamic temperature, is the fraction 1/273.16 of the thermodynamic temperature of the triple point of water. (Symbol: K)

Candela The candela is the luminous intensity, in the perpendicular direction, of a surface of 1/600 000 m^2 of a black body at the temperature of freezing platinum under a pressure of 101 325 N m^{-2}. (Symbol: cd)

A seventh unit, the mole, is also required. This is the unit of amount of substance.

Mole The mole is the amount of substance of a system which contains as many elementary units as there are carbon atoms in 0.012 kg of carbon-12. The elementary unit must be specified and may be an atom, a molecule, an ion, an electron, a photon, etc., or a specified group of such entities. (Symbol: mol.).

APPENDIX 2

PREFIXES FOR SI UNITS

The following prefixes are used to denote decimal fractions or multiples of SI units:

Fraction	*Prefix*	*Symbol*	*Multiple*	*Prefix*	*Symbol*
10^{-1}	deci	d	10	deka	da
10^{-2}	centi	c	10^{2}	hecto	h
10^{-3}	milli	m	10^{3}	kilo	k
10^{-6}	micro	μ	10^{6}	mega	M
10^{-9}	nano	n	10^{9}	giga	G
10^{-12}	pico	p	10^{12}	tera	T
10^{-15}	femto	f			
10^{-18}	atto	a			

Some common examples:

Power generation and utilization: kilowatt (kW), Megawatt (MW), kilovolt (kV)
Electronic circuitry: milliwatt (mW), microwatt (μW), millivolt (mV), microvolt (μV)
Capacitors: microfarad (μF), picofarad (pF)
Resistors: kilohm (kΩ), megohm (MΩ)
Frequency: kilohertz (kHz), megahertz (MHz)
Time: millisecond (ms), microsecond (μs)
Length: millimetre (mm), micrometre (μm), nanometre (nm).

APPENDIX 3

SOME PHYSICAL CONSTANTS

Quantity	*Symbol*	*Value*
speed of light in a vacuum	c	2.998×10^{8} m s^{-1}
permeability of a vacuum	μ_0	$4\pi \times 10^{-7}$ J s^2 C^{-2} m^{-1}
permittivity of a vacuum	ε_0	8.854×10^{-12} J^{-1} C^2 m^{-1}
mass of hydrogen atom	m_H	1.67343×10^{-27} kg
mass of proton	m_p	1.67252×10^{-27} kg
mass of neutron	m_n	1.67482×10^{-27} kg
mass of electron	m_e	9.109×10^{-31} kg
charge of proton or electron	e	1.602×10^{-19} C
Boltzmann constant	k	1.381×10^{-23} J K^{-1}
Planck constant	h	6.626×10^{-34} J s
Avogadro constant	N_A	6.023×10^{23} mol^{-1}
gas constant	R	8.314 J mol^{-1} K^{-1}
Faraday constant	F	9.649×10^{4} C mol^{-1}
gravitational constant	G	6.670×10^{-11} N m^2 kg^{-2}

SOME CONVERSION FACTORS

1 atomic mass unit (a.m.u. or u)	$= 1.660 \times 10^{-27}$ kg
1 electronvolt (eV)	$= 1.602 \times 10^{-19}$ J
1 calorie (cal)	= 4.184 J
1 kilowatt hour (kWh)	$= 3.6 \times 10^{6}$ J
1 degree Celsius (°C)	= 1 Kelvin
A temperature of T °C	= A temperature of $(T + 273.15)$ K
1 curie (Ci)	$= 3.7 \times 10^{10}$ s^{-1}
1 angström (Å)	$= 10^{-10}$ m $= 10^{-1}$ nm

APPENDIX 4

DAMPED SIMPLE HARMONIC MOTION

It was argued in Section 19.1 that the displacement of an oscillator which is damped in such a way that its amplitude decays exponentially can be represented by the expression

$$x = a_0 \, e^{-\Lambda t} \cos \omega t.$$

We now show that this expression is a solution of the equation of motion for a simple harmonic oscillator with viscous damping. Let m be the mass of the oscillating object, k the restoring force per unit displacement, and c the damping force per unit velocity. Then the equation of motion can be written

$$m \frac{d^2x}{dt^2} + c \frac{dx}{dt} + kx = 0.$$

We now proceed to substitute the expression for x into this differential equation. Firstly we find that

$$\frac{dx}{dt} = a_0[-\omega \sin \omega t - \Lambda \cos \omega t]e^{-\Lambda t}$$

and

$$\frac{d^2x}{dt^2} = a_0[(\Lambda^2 - \omega^2) \cos \omega t + 2\Lambda\omega \sin \omega t]e^{-\Lambda t}.$$

Substituting into the differential equation, we then obtain

$$m(\Lambda^2 - \omega^2) \cos \omega t + 2m\Lambda\omega \sin \omega t - c\omega \sin \omega t - c\Lambda \cos \omega t + k \cos \omega t = 0.$$

Now, for this equation to be valid for all values of t, the terms in $\cos \omega t$ and in $\sin \omega t$ must be equated separately. Doing this we get

$$m(\Lambda^2 - \omega^2) - c\Lambda + k = 0,$$

and

$$2\Lambda\omega m - c\omega = 0,$$

hence

$$2\Lambda m - c = 0,$$

therefore

$$\Lambda = \frac{c}{2m}.$$

Then, substituting $c = 2m\Lambda$ in the first of the two equations, we get

$$m(\Lambda^2 - \omega^2) - 2m\Lambda^2 + k = 0,$$

hence

$$\omega^2 = \frac{k}{m} - \Lambda^2 = \omega_0^2 - \Lambda^2.$$

As explained in Section 19.1, in situations where the damping is not very heavy, Λ is very much less than ω_0 and so, to a good approximation, $\omega \simeq \omega_0$.

APPENDIX 5

FORCED SIMPLE HARMONIC MOTION, WITH VISCOUS DAMPING

Let m be the mass of the oscillating object, k the restoring force per unit displacement, c the damping force per unit velocity. The driving force, we suppose, is a function of time only and varies sinusoidally with time at a constant angular frequency ω, so the expression for this force can be written $F_0 \cos \omega t$. The complete equation of motion for the oscillator is therefore

$$m \frac{d^2x}{dt^2} + c \frac{dx}{dt} + kx = F_0 \cos \omega t.$$

Now we know from experimental observation that, in a steady state, the oscillator responds to the driving agency by oscillating with a constant amplitude, a, at the same frequency as the driver; but there is a phase difference, ϕ, between the responding oscillator and the driver. Thus, in the steady state, the motion of the responding oscillator can be represented by $x = a \cos(\omega t + \phi)$. We can obtain expressions for a and ϕ by substituting this expression for x in the differential equation, as follows. Firstly we get

$$-ma\omega^2 \cos(\omega t + \phi) - ca\omega \sin(\omega t + \phi) + ka \cos(\omega t + \phi) = F_0 \cos \omega t.$$

Expanding the terms in $(\omega t + \phi)$ to obtain terms in $\cos \omega t$ and $\sin \omega t$, and then remembering that, for the equation to be valid for all values of t, the terms in $\cos \omega t$ and the terms in $\sin \omega t$ must be equated separately, we can obtain the two equations:

$$-ma\omega^2 \cos \omega t \cos \phi - ca\omega \cos \omega t \sin \phi + ka \cos \omega t \cos \phi = F_0 \cos \omega t,$$

and

$$ma\omega^2 \sin \omega t \sin \phi - ca\omega \sin \omega t \cos \phi - ka \sin \omega t \sin \phi = 0.$$

Eliminating $\cos \omega t$ from the first equation, and $\sin \omega t$ from the second, we get

$$a(k - m\omega^2) \cos \phi - ca\omega \sin \phi = F_0,$$

and

$$(m\omega^2 - k) \sin \phi - c\omega \cos \phi = 0.$$

The second of these equations gives

$$\tan \phi = \frac{c\omega}{m\omega^2 - k}.$$

So here we have an expression for ϕ in terms of the parameters of the oscillatory system. To obtain an expression for a we need to eliminate ϕ from the first of the two equations above. This equation contains $\cos \phi$ and $\sin \phi$, but the expression we have just got is for $\tan \phi$. We can get around this difficulty by using the relationship

$$\cos^2 \phi = \frac{1}{1 + \tan^2 \phi}.$$

Firstly, if we divide the upper of the two equations throughout by cos ϕ we get

$$(k - m\omega^2)a - ca\omega \tan \phi = \frac{F_0}{\cos \phi} = F_0\sqrt{1 + \tan^2 \phi}$$

and then, substituting the expression we obtained earlier for tan ϕ, we obtain

$$a\left[-(m\omega^2 - k) - \frac{c^2\omega^2}{(m\omega^2 - k)}\right] = F_0\sqrt{1 + \frac{c^2\omega^2}{(m\omega^2 - k)^2}}$$

and this yields finally

$$a = \frac{F_0}{\sqrt{(m\omega^2 - k)^2 + c^2\omega^2}}$$

which shows how the steady-state amplitude depends upon the parameters of the system.

The steady-state amplitude is a maximum when the denominator of the expression for a is a minimum. To find what value of ω makes it a minimum we differentiate the expression under the root-sign with respect to ω and equate to zero, thus

$$\frac{d}{d\omega}[(m\omega^2 - k)^2 + c^2\omega^2] = 0$$

hence

$$2(m\omega^2 - k) \times 2m\omega + 2c^2\omega = 0,$$

which yields

$$\omega = \sqrt{\frac{k}{m} - \frac{c^2}{2m^2}}.$$

If there is no damping, $c = 0$, and so $\omega = \sqrt{k/m}$, the natural frequency of undamped oscillation of the system, which we can call ω_0. If there is some damping the amplitude is a maximum when ω is *less* than ω_0.

It was shown in Appendix 4 on damped simple harmonic motion that the decay constant $\Lambda = c/2m$, so the expression for ω which we have just derived can be written

$$\omega = \sqrt{\omega_0^2 - 2\Lambda^2}.$$

The value of Λ^2 in practice is very much less than ω_0^2, and so we can make the approximation $\omega = \omega_0$ in most practical cases.

A graph of the steady-state amplitude against driving frequency produces a curve like that of Fig. 20.5. There it was explained that the width of the resonance peak can be expressed $2\,\Delta\omega$, and we shall now show that $\Delta\omega = \Lambda$, the decay constant for damped free oscillations of the system.

Look again at the expression we derived above for the steady-state amplitude when the driving frequency is ω. It is

$$a = \frac{F_0}{\sqrt{(m\omega^2 - k)^2 + c^2\omega^2}}.$$

Now, the natural frequency (that is, the frequency of free, undamped, unforced oscillations) of the system is given by $\omega_0 = k/m$. We have said above that, if the damping is not too great, the value of ω which makes a a maximum can be taken as approximately equal to ω_0. If we do this, then the expression for amplitude simplifies to

$$a_R = \frac{F_0}{c\omega_0}$$

at resonance, when $\omega = \omega_0$. Notice that we have used the suffix R here: this is to indicate the amplitude *at resonance*.

The expression for the amplitude at frequency ω can be written

$$a = \frac{F_0}{\sqrt{m^2(\omega^2 - \omega_0^2)^2 + c^2\omega^2}}$$

after making the substitution $\omega_0 = k/m$. Now, if $\omega = \omega_0 \pm \Delta\omega$, and $\Delta\omega$ is small compared with ω_0, then we can make these two approximations

$$(\omega^2 - \omega_0^2) \simeq 2\omega_0(\omega - \omega_0) = \pm 2\omega_0\,\Delta\omega,$$

and

$$c^2\omega^2 \simeq c^2\omega_0^2,$$

which enable us to simplify the expression for amplitude to

$$a = \frac{F_0}{2m\omega_0\sqrt{\Delta\omega^2 + \frac{c^2}{4m^2}}}$$

and since, as shown earlier in this appendix, $\Lambda = c/2m$, we can write

$$a = \frac{F_0}{2m\omega_0\sqrt{\Delta\omega^2 + \Lambda^2}}.$$

We can also write the expression for the amplitude at resonance

$$a_R = \frac{F_0}{2\Lambda m\omega_0}.$$

To find the width of the resonance peak we want the value of $\Delta\omega$ which makes

$$a^2 = \tfrac{1}{2}a_R^2.$$

Thus

$$\frac{F_0^2}{4m^2\omega_0^2[\Delta\omega^2 + \Lambda^2]} = \frac{1}{2} \cdot \frac{F_0^2}{4\Lambda^2 m^2\omega_0^2}$$

therefore

$$\Delta\omega^2 + \Lambda^2 = 2\Lambda^2, \qquad \text{and so} \qquad \Delta\omega = \Lambda.$$

This result is quoted in Section 20.2.

ANSWERS TO SELECTED PROBLEMS

Unit 2

2.3 51.2 MJ; 2.56×10^6 N; 16 cm^2.

2.7 3.895 cm high by 3.895 cm diameter.

2.11 (a) 4.2 MJ; (b) 2×10^3 N; (c) 1.4×10^3 N.

2.13 (a) 100 N; (b) 50 kg; (c) 2 m s^{-2}; (d) 2000 J; (e) 1600 J; (f) about 9 m s^{-1}.

2.15 33 kg; 5.25 mm.

2.17 18 m s^{-1}; 9000 N.

2.19 Atom 96: 1.58×10^7 m s^{-1}; 189×10^{-13} J (118 MeV). Atom 140: 1.08×10^7 m s^{-1}; 130×10^{-13} J (81 MeV).

2.21 5.5 m s^{-1}; 16 000 J.

2.23 − 2.14 and 2.86 m s^{-1}; 0.815.

2.27 $\sqrt{\frac{Mg}{Ad}}$; $\frac{Adv^2}{M} - g$.

2.29 600 m s^{-1}.

2.31 0.25 h.

2.35 8×10^3 m s^{-1}; 84 min.

Unit 3

3.1 13.7 Ω.

3.3 Assuming all kinetic energy becomes available to the pellet, the temperature rise of a 1 g pellet of lead is about 40 K.

3.7 8 min.

3.11 21%.

3.15 Efficiency of nth engine is $1/n$.

Unit 4

4.1 0.022 4 m^3.

4.3 (a) 2.7×10^5 N m^{-2}; (b) 5.5 cm.

4.5 500 m s^{-1}.

4.7 3730 J; 3742.5 J; 12.5 J.

4.9 200 W.

4.11 (a) 2260 J; (b) 5650 J; (c) 2260 J.

4.13 3.75 mm.

4.15 54.6 m s^{-1} or 44 Hz.

4.17 (a) 27; (b) 27; (c) 9; (d) 27; (e) 27; (f) 9; (g) 9; (h) $\sqrt{3}$.

Unit 5

5.1 (ii) 0.4%.

5.3 (i) 0, 6, 12; (ii) 3, 9; (iii) 3, 9; (iv) 0, 6, 12; (v) 1, 5, 7, 11.

5.5 (a) 4 s; (b) $\pi/2$ radians s^{-1}; (c) $5\pi/2$ cm s^{-1}; (d) $5(\pi/2)^2$ cm s^{-2}.

5.13 0.64 m.

5.15 89 N.

5.19 $(f_1 - f_2)$.

5.21 For an open pipe the frequencies are 262, 524, 786, 1310, 1572, 1834, 2096, 2358, 2620 Hz.

Unit 6

6.1 (a) Current = 23 A; the fuse would blow; (b) Power consumption is 19% of rate in Britain; (c) 54; (d) 93%.

6.3 (a) 6.2 V; (b) 3.2%.

6.5 97 pF per mm^2.
Area needed is 21 cm^2 and this is too large.

6.7 39.4 J.

6.9 11.9.

6.13 (i) 1 A; (ii) 9 A; (iii) 91 V; (iv) 819 W; (v) 81.9%.

6.15 (ii) Rate decreases; (iii) Rate decreases.

Unit 7

7.1 Between 3 and 4×10^{-7} N.

7.3 If Z is taken to be 50, r to be 10^{-14} m, $E = 7 \times 10^{20}$ N C^{-1}.

7.5 4.4×10^{-2} T.

7.11 23×10^{-15} J; 1.12×10^{9} K.

7.15 1.4×10^{-5} T.

7.17 (a) 4.6×10^{-2} N; (b) 0.12 V.

7.19 0.3 m.

Unit 8

8.3 (a) 83% and 17%; (b) 4.3×10^{-5} m s^{-1}; 0.89×10^{-5} m s^{-1}; (c) 25 Ω; (d) 125 V m^{-1}; (e) 3.5×10^{-7} V m^2 s^{-1}.

8.5 0.994 T; 6.96×10^{28} conduction electrons per cubic metre; 1.2 electrons per silver atom.

8.11 7×10^{-3} m s^{-1}.

8.13 1.4×10^{4} m s^{-1}.

8.15 (a) 1.3×10^{-14} J; (b) 0.164×10^{-17} J; (c) 33×10^{-26} J; (d) 1.3×10^{-28} J.

8.19 (a) 3.2×10^{14} Hz; (b) 2.1×10^{-19} J (1.3 eV); (c) 6.7×10^{-34} Js.

8.21 100 m.

8.23 87×10^{-19} J.

Unit 9

9.3 About 3×10^{9} times bigger.

9.5 10^{-4} m^2.

9.7 (a) 3 in 10 000; (b) 1 in 160 000; (c) 4 in 10 000.

9.9 47, 13, silver, aluminium.

9.11 5.4×10^{26} m s^{-2}; 0.735 m.

9.13 1.33×10^{-12} J; 8.3 MeV.

9.15 4.55×10^{-14} m.

9.17 5×10^{4} J kg^{-1}, 5×10^{13} J kg^{-1}.

9.19 The same.

9.21 (a) 2×10^{7} m s^{-1}; (b) 1.6×10^{7} m s^{-1}; (c) 4.45×10^{6} m s^{-1}; (f) 0.05%, 64%, 70%.

9.29 3 MeV.

9.35 (a) 25%; (b) 3%; (c) 0.1%; (d) $100(\frac{1}{2})^n$.

9.37 0.0294 mCt.

9.39 10.2 h, 1.9×10^{-5} s^{-1}.

9.41 1560 years.

9.47 (a) 2.3 N; (b) 7.1 MeV.

9.49 28.9 mm, 40.75 mm.

GLOSSARY OF SYMBOLS

A	Area of cross-section, area, a virial coefficient, mass number.
a	Acceleration, amplitude, distance, a constant.
B	Magnetic flux density (field strength), a virial coefficient.
b	Distance.
C	Capacitance, heat capacity, a constant, a virial coefficient.
c	Specific heat capacity, speed of light in a vacuum.
d	Displacement, distance.
E	Electric field strength, electromotive force, energy, Young's modulus.
E_K, E_P	Kinetic and potential energy.
$\mathscr{E}$	Induced electromotive force.
e	Elementary charge, electron, exponential.
F	Faraday constant, force.
f	Frequency.
G	Gravitational force constant, shear modulus.
g	Acceleration of free fall, gravitational field strength.
h	Height, Planck constant.
I	Electric current, moment of inertia.
i	Angle of incidence.
K	A constant.
k	A constant, an integer.
L	A constant, a length, self-inductance.
L_s	Energy of sublimation.
l	Displacement, a distance, an integer.
M	Mass, mutual inductance.
m	Mass.
m_e, m_p	Mass of electron, mass of proton.
N	An integer, number of molecules, number of turns, neutron number.
N_A	Avogadro constant.
n	A number of events or items, order, rotational frequency.
P	Power.
p	Momentum, pressure, proton.
Q	Charge, energy transfer as heat.
q	Charge, a number.
R	Molar gas constant, radius of earth, resistance, Rydberg constant.
r	Angle of refraction, radius.
s	Displacement, distance.
T	Temperature, torque, period.
t	Time.
U	Internal energy, electric potential difference.
u	Speed.
V	Electric potential, potential difference, volume.
v	Speed.
$\sqrt{v^2}$	R.m.s. speed.
W	Energy transfer as work.
w	Distance.
X	Reactance.
x	Displacement, distance, a number.
y	Displacement, distance.
Z	Impedance, atomic number.
z	Displacement, distance.

α	Angle, alpha-particle, cubic expansivity.
β	Cubic expansivity.
γ	Surface energy, surface tension, a photon.
ε_r	Relative permittivity.
ε_0	Permittivity of free space (electric force constant).
η	Viscosity.
θ	Angle, temperature.
κ	Boltzmann constant.
Λ	Logarithmic decrement.
λ	Decay constant, mean free path, thermal conductivity, wavelength.
μ	Mass per unit length.
μ_r	Relative permeability.
μ_0	Permeability of free space (magnetic force constant).
ν	A neutrino.
ρ	Charge density, density, resistivity.
σ	Electrical conductivity.
τ	Relaxation time.
Φ	Magnetic flux.
ϕ	Angle, phase difference, work function.
ψ	Electric flux.
ω	Angular frequency, solid angle.

UNIT SYMBOLS

A	ampere
C	coulomb
Ci	curie
°C	degree Celsius
eV	electron-volt
F	farad
g	gramme
H	henry
Hz	hertz
J	joule
K	kelvin
kg	kilogramme
m	metre
mol	mole
N	newton
rad	radian
s	second
T	tesla
u	atomic mass unit
V	volt
W	watt
Wb	weber
Ω	ohm

GENERAL BIBLIOGRAPHY

Original sources

Boorse, H. A., and Motz, L. (1966). *The World of the Atom*, Basic Books.
A compendium of landmark documents in the history of atomic physics in chronological order from Lucretius to Yang and Lee.

Magie, W. F. (1963). *A Source Book in Physics*, Harvard University Press.
Excerpts from original writings covering most aspects of physics.

Wright, S. (Ed.) (1964). *Classical Scientific Papers—Physics*, Mills and Boon.
A collection of famous scientific papers on, mainly, atomic structure.

Other sources

Arons, A. B. (1965). *Development of Concepts of Physics*, Addison-Wesley.

Feynman, R. P., Leighton, R. B., and Sands, M. (1963). *Lectures on Physics*, Addison-Wesley.
Tough, authoritative and stimulating text for undergraduates (and their tutors).

(Harvard Project Physics). (1968). Student texts and Readers: Holt, Rinehart and Winston.

1. Concepts of Motion.
2. Motion in the Heavens.
3. The Triumph of Mechanics.
4. Light and Electromagnetism.
5. Models of the Atom.
6. The Nucleus.

(Nuffield). (1967). *Physics Teachers' Guides* (5 volumes), Longmans/Penguin.
Designed for teachers of physics to the Nuffield "O" level course in the United Kingdom.

(Nuffield). (1972). *Advanced Level Physics*, Teacher's Guides and Student's Books, Penguin Education.

(P.S.S.C.) (1968). *College Physics*, Raytheon Educational Press.
This includes the text for the P.S.S.C. course in physics for high schools in the United States together with the advanced topics supplement.

Resnick, R., and Halliday, D. (1966). *Physics*, Wiley.

Rogers, E. M. (1960). *Physics for the Inquiring Mind*, Princeton University Press.
The text for a one year undergraduate course which has had a considerable influence on other courses in physics.

Royal Society (1969). *Symbols, Signs, and Abbreviations*. Royal Society, London.

Sears, F. W., and Zemansky, M. W. (1970). *University Physics*, 4th edn, Addison-Wesley.

INDEX

Absolute zero of temperature, 89
Absorption bandwidth, 181
Absorption of ultraviolet radiation, 463
Absorption spectrometer, 181
  spectrum, 181, 462
Acceleration, 39
  central, 66
  centripetal, 66
  due to gravity, 48, 343, 353, 366
  in harmonic motion, 152, 158
Accumulator, lead, 251
Action and reaction pairs, 42, 341
Activity, radioactive, 520
Adiabatic process, 80, 85
Air table, 25
Alkali metals, 476
Alpha particle, 487
  emission, 506
  scattering, 490
Alternating current, 313
  generator, 332
Alternator, 332
Ammeter, 235, 284, 294
  calibration of, 236
  resistance of, 284
Ampère, A. M., 288, 346
Ampère's law, 404
Ampère, the, 290, 352, 522
Amplification, 255
Amplifier, audio, 254
Amplitude, 145, 158
Analogy, for the electric current, 237, 242
  for electric charge, 268
Angle, phase, 151
  of incidence, 207
  of reflection, 207
  of refraction, 210, 211
  of shear, 133
Angstrom unit, 524
Angular frequency, 148
Angular momentum, 55
  conservation of, 55
Angular velocity, 55, 150
Anion, 427
Anode, 427
Anode radiation, 439
Anti-neutrino, 508
Antinodal line, 212
Antinode, 203
Antiphase, 150
Arago, D. F. J., 288
Archimedes' principle, 161
Aristotle, 473
Armature, 293, 297
Aston, F. W., 506
Atom, 12, 106
Atomic mass, 475
  unit, 476, 520, 524
Atomic model, 8, 106, 472, 489, 495
Atomic number, 494, 506, 519
Atomic theory, 475
Atomists, the early, 472
Attenuator, 282
Audio-amplifier, 254
Avogadro's hypothesis, 106
Avogadro's number, 98, 105, 429
  determination of, 105, 431, 437

Balance, torsion, 367
Ballistic galvanometer, 171, 270
Balmer, J. J., 459, 478
Balmer's series, 460
Bandwidth, absorption, 181
Barkla, C. G., 488, 495
Barometer, mercury, 4
Battery, storage, 252
Beats, 223
Becker, G. F., 503
Becquerel, H., 485
Beta particle, 487
  emission, 507
  energy spectrum of, 508
Betatron, 399
B field (see Magnetic field)
Binding energy, 520
Biot-Savart law, 371, 372
Black-body radiation, 482
Black, J., 83, 474
Blackett, P. M. S., 501
Blue of the sky, 210
Bohr, N., 490, 495
Bohr atom, 495
Boltzmann distribution, 103
  constant, 104
Bothe, W. G., 503
Bouguer, P., 349
Boyle, R., 2, 8, 473

Boyle's law, 9, 97, 98
Boys, C., 367
Bragg, Sir Lawrence, 123, 125
Bragg law, 124
Brahé, T., 364
Bridge circuits, 265
Brittle behaviour, 113
Brownian motion, 99
Brushes, 296
Bunsen, R. W., 477

Caloric theory, 83
Calorimetry, 83
Campbell, N. R. (quoted), 351
Candela, the, 522
Cantilever, 177
Capacitance, 271, 272
Capacitor, 271, 316
  and dielectric, 276
  discharge of, 273
  electrolytic, 276
  paper, 275
  parallel plate, 271, 275
  uses of, 277
Carnot, N. L. Sadi, 85
Carnot cycle, 87
Carnot engine, 88
Carnot refrigerator, 88
Carnot theorem, 88
Cation, 427
Cathode, 427
Cathode radiation, 439
Cathode rays, 438, 478
Cathode ray oscilloscope, 386, 443
  tube, 443
Cavendish, H., 367, 474
Cell, dry, 251
  reversible, 251
  unit, 118
Central acceleration, 66, 365
Central force, 65, 66
Centrifugal force, fictitious, 66
Centripetal force, 65, 66, 365
Chadwick, Sir James, 501, 503, 504
Chain reaction, 520
Change, rate of, 18
  reversible, 86
Characteristic of a circuit element, 278
  of a diode, 279
Charge, 238, 268, 428
  equality of, 344
  induced, 344
  measuring, 270
  of the electron, 434
  unit of, 238, 352
Charles' law, 97
Chladni's plate, 177, 205
Choke, 325
Circuits, bridge, 265
  integrated, 232
  oscillatory, 327
  simple, 235
  smoothing, 325
  time-constant of, 275
  tuned, 337
  water analogue, 268
Circuit elements, 278
  in combination, 281
Circuit problems, solving, 257
Circular motion, 65
  and frames of reference, 66
Clausius, R. J., 429 (quoted), 102
Cloud chamber, diffusion type, 513
  expansion type, 513
Cockcroft, Sir J. D., 490
Coefficient, virial, 97
Coherence, 218, 464
Collimation, 103
Collision, 29, 30
  elastic, 13, 30, 56, 504
  hyperelastic, 13, 56
  inelastic, 13, 30, 56
Commutator, 296
Components of a vector, 23
Compressibility, 111
Compton, A. H., 437
Computer, 233
Conduction of electricity, 136, 445
  in electrolytes, 429
  in gases, 437
Conduction of heat, 137, 445
Conductivity, electrical, 137, 245, 445
  molar, 466
  thermal, 137, 445, 447
Conductors, linear, 243
  non-linear, 278
  ohmic, 243
Conservation, of angular momentum, 55
  of energy, 54, 80, 95, 508
  of momentum, 34, 61, 95, 508
Conservative fields, 389
  forces, 54
Conversion factors, 524
Copper, crystalline structure of, 125
  losses, 331
Coulomb, C. A., 288
Coulomb forces, 288
Coulomb, the, 238, 352
Couple, 294
Cowan, C. and Reines, F., 508
Crack, 131
Creep, 114
Critical potential, 455
  size, 520
Cross product, 293
Crookes, Sir William, 438
Crookes' dark space, 438
Crystal, 116
  hexagonal close packed, 118
  face-centred cubic, 118
  rotating, 125
Crystalline structure, 116
Curie, Marie and Pierre, 486
Curie, Irene and Joliot, F., 503
Curie, the, 520, 524
Current balance, 288, 346
  simple, 292
  standard, 294
Current, electric, 235, 248, 426, 445
  alternating, 313
  conventional direction of, 237, 348, 439
  model of, 426, 445
  unit of, 290, 352, 522
Cycle, 144

Dalton, J., 12, 106, 475
Dalton's atomic theory, 475
Damped oscillations, 166, 295, 525
Damping, critical, 171, 295
  degree of, 166
  loss of energy due to, 172
  viscous, 167, 526
Decay, exponential, 168, 275
  law, 519

Deformation, permanent, 131
plastic, 131
Deuteron, 510
Dielectric material, 275
Dielectric constant, 276
Diffraction, 121, 208
at an aperture, 208
at an edge, 208
X-ray, 121, 210
Diffraction grating, 216
Diffusion, 104
cloud chamber, 513
Diode characteristic, 279
rectifier, 278
semi-conductor, 279
thermionic, 279
zener, 280
Dipole field, 390
Dipole moment, 390
Dirac, P., 510
Discharge of capacitor, 273
Dislocation, edge, 133
Dispersion, 197, 211
Dispersive medium, 197
Displacement, 19, 145
laws, 487, 506
vectors, 19
Dissipative forces, 54
Drift speed, 446
Driver, 174
Drude, P. K. L., 445
Drude theory of electric conduction, 445
Dry cell, 251
Ductile, 113
Ductility, 133
Dulong and Petit law, 83, 135, 452
Dynamid, 489
Dynamo, 298
disc, 305
simple, 303

Earth, gravitational field of, 48, 67, 342, 349, 366
magnetic field of, 346
potential well near, 381
Eddy-current, 331
losses, 331
Edge dislocation, 133
Efficiency, 85, 87
of a Carnot engine, 88
E field (see Electric field)
Eigen frequency, 202, 204
Einstein, A., 401, 406, 414, 418, 458, 464, 483
photo-electric equation, 458
mass-energy relationship, 418
special theory of relativity, 414
Einstein, A., and Infeld, L. (quoted), 406
Elastic collisions, 13, 30, 56, 504
Elastic deformation, 119
limit, 112
moduli, 112
Electric charge, 238, 268, 343, 351, 357, 402, 426, 434
induced, 344
measuring, 270
unit of, 238, 352
Electric conduction, 136, 445
in electrolytes, 429
in gases, 437
Electric current, 235, 248, 351, 426, 445
alternating, 313
conventional direction of, 237, 348, 439
model of, 426, 445
unit of, 290, 352, 522
Electric field, 343, 344, 346, 358, 401
model of, 360
of a dipole, 390
energy stored in, 389
Electric flux, 361, 402, 404, 410
Electric force constant, 360
Electric motor, 295, 331, 336
Electric and magnetic fields, directions of induced, 404
Electrical conductivity, 136, 245, 445
Electrical devices, linear, 243
Electrical energy, 231, 238, 248, 251, 299
Electrical oscillation, 325
Electrical resistance, 243
Electrode, 427
Electrolysis, Faraday's laws of, 106, 427, 429
Electrolysis of water, 428
Electrolytes, 427
Electrolytic conductivity, 466
Electromagnetic field, Maxwell's model for, 401
Electromagnetic forces, 288
induction, 301, 306, 393, 398, 402
oscillations, 407
wave, 406, 407, 412
Electromagnetism, 288
Electrometer, 358
Electromotive force, 248, 249
back, 298
comparison of, 264
direction of, 396
induced, 310, 393
peak value, 304
Electron, 351, 433, 442, 480
charge of, 434
mass of, 442
specific charge of, 442
Electron-gun, 443
Electronic hypothesis, 433
vacuum tubes, 232
Electron spin resonance, 182
Electron-volt, the, 388, 524
Electrostatic force, 288, 343, 358
between two charged particles, 358
Electroscope, gold leaf, 386
Elster, J., and Geitel, H., 439
Emission, stimulated, 463
Emission spectrum, 462
Energy, 46, 248
binding, 520
conservation of, 54, 80, 95, 508
internal, 80, 110
kinetic, 48
potential, 48, 53, 54, 79, 163, 376, 385, 390
storage of, 47
in a capacitor, 277
in an electric field, 389
in a gravitational field, 48
in a magnetic field, 396
in a rotating body, 54
strain, 128
surface, 129
Energy density, 390, 398
distribution, 481
levels, 458, 496
of a resonant system, 179

Energy transfer, in change of shape, 49
in change of speed, 50
in collisions, 56
quantitative examination of, 52
Energy transformation, 46
in electrical circuits, 238
Energy transmission by particles, 213
by waves, 213
electrically, 230
Engine, Carnot, 88
ideal reversible, 87
Enrichment factor, 105
Equipotential lines, 376
surfaces, 376
Equation of motion for a simple harmonic oscillator, 158
Equation of state, 97
Equilibrium, thermal, 77
Equilibrium state, 76
Excitation, 455, 458
Excited state, 462
Expansion, thermal, 113, 135
Explosion, 28
Exponential decay, 168, 275

Face centred cubic, 118
Falsifiability, 7
Farad, the, 273
Faraday, M., 230, 301, 303, 350, 427, 431
Faraday constant, 430
dark space, 438
lines of force, 303, 401
Faraday's laws of electrolysis, 106, 427, 429
of electromagnetic induction, 393, 398, 402
Fatigue, 114
Feather, N., 504, 505
Fermi, E., 508
Ferrites, 331
Ferranti, de, 231
Ffestiniog power station, 47
Field, 341
conservative, 389
electric, 343, 346, 358
between parallel plates, 362
comparison with gravitational field, 367
direction of, 346
intensity of, 345, 361, 385
gravitational, 36, 48, 342, 346, 364
comparison with electric field, 367
direction of, 346
intensity of, 343
magnetic, 346, 369
direction of, 347
intensity of, 346
of flat coil, 371
of long solenoid, 369, 372
of long, straight conductor, 403
rotating, 332
of force, 341
potential, 388
vector, 343
Field strength (intensity), 343, 345, 355, 361, 385
Field winding, 299
First law, Newton's, 33
First law of thermodynamics, 81
Fission, 512
Fleck, A., 495
Fleming's left hand rule, 348, 439
Flux, 356
and the inverse square law, 356
electric, 361, 402, 404, 410
density, 361
magnetic, 303, 356, 374
density, 346
model, 361, 402
Force, 38, 44
between current-carrying conductors, 288
central, 65
centrifugal, 66
centripetal, 66
conservative, 54
dissipative, 54
electrostatic, 341, 343
fields of, 341
gravitational, 36, 341, 342, 364
inter-molecular, 13
inter-particle, 108
lines of, 349
magnetic, 341, 346
of strong interaction, 341
of weak interaction, 341
on conductors in magnetic fields, 449
unbalanced, 33
unit of, 40
Foucault, J., 210
Fourier analysis and synthesis, 197, 198
Fracture, brittle, 113
Frames of reference, 61
accelerating frames, 63
and circular motion, 66
conservation of energy in, 62
intertial frames, 62
rotating frames, 63
Franck, J. and Hertz, G., 453
Franz, R., 445
Fraunhofer lines, 463, 477
Free electron gas model, 446, 452
Free fall, 342
Frequency, 145, 325
angular, 148
characteristic, 326
eigen, 202, 204
fundamental, 204
Fresnel, A., 208
Friction, reduction of, 24
viscous, 434
Fundamental frequency, 204

g, 48, 343, 353, 366
Galilean relativity, 414
Galileo, G., 3, 36, 146, 414, 473
Galvanometer, ballistic, 171, 270
damping, 171
moving coil, 300
Gamma rays, 487, 508
Gas constant, 97, 104
Gaseous ionization, 453
Gas, a model for, 11, 94
laws, 97
thermometer, 78
Gassendi, P., 473
Gay-Lussac, J. L., 97
Gauss' law, 361
Geiger, H. and Marsden, E., 490
Generation of electricity, 47, 230
Generator, a.c., 252
Glass transition temperature, 113
Gold leaf electroscope, 386

Glossary of symbols, 531
Gravity, force of, 6, 47
  acceleration due to, 48, 343, 366
Gravitation, Newton's law of, 6, 364
Gravitational constant, 349
  measurement of, 367
Greek science, 5
Griffith, A. A., 131
Grimaldi, F. M., 208
Ground state, 461
Group velocity, 197
Growth of current in inductive circuit, 311

Haldane, J. B. S., 484
Hales, S., 474
Half-life of oscillatory system, 169
  of *R-C* circuit, 275
  of radioactive nuclei, 519
Hall effect, 449
  coefficient, 450
  voltage, 452
Halogens, 476
Harmonic oscillator, 158, 159
Heat, 12, 80
  and temperature, 89
  and work, 79
  energy model for, 79, 83
  of sublimation, 128
Heat capacity, 82
  molar, 83, 135, 452
  specific, 83, 135, 452
Heat engine, ideal reversible, 87
  Carnot, 88
Heating effect of the current, 235
Helmholtz, H. L. F. von, 431
Henry, the, 309
Hertz, G., 453
Hertz, H. R., 438, 457, 482
Hertz, H. R. and Lenard, P. A. R. von, 438, 479
Hertz, the, 145
Hexagonal close packed crystal, 118
Hirn, G. A., 54
Hittorf, J. W., 430
Hooke, R., 4
Hooke's law, 5, 112, 127
Hopkinson, J., 230
Huygens, C., 208
Huygens' wave model, 298
Hyperelastic collision, 13, 56
Hypothesis, 7, 106

Ideal gas, 98
  equation of state of, 104
  scale of temperature, 78, 90
Image, 207
Impedance, 202, 255, 321, 324
Induced charge, 344
Inductance, mutual, 309
  self, 309
Induction, 305
  electromagnetic, 301, 306
  electrostatic, 344
  in a magnetic field, 392
  mutual, 309
  self, 309
Induction motors, 332
  linear, 333
Inductors, 316, 318
  in a.c. circuits, 318
  reactance of, 320
Inelastic collisions, 13, 30, 56
Inertia, 33
  moment of, 55
Inertial frame of reference, 414
Infrared spectroscopy, 181
Insulators, 343
Integrated circuit, 232
Intensity of electric field, 345, 361, 385
  of gravitational field, 343
  of magnetic field, 346
Interactions, 24, 28
  hard, 30
  in two dimensions, 62
Interference, 121, 212
  of light, 213
  of ripples, 212
  of sound, 223
Interference patterns and crystal structure, 121
Internal energy, 80, 110
Interparticle force, 109
Inverse square law, 355
  flux and, 356, 393
  for electric charges, 360
  for masses, 364
  Newton's test of, 366
  range of application, 388
Involute, 66
Ion, 427
Ionic mobility, 466
  model, 427
Ionization, 427, 455, 458, 463
Ionosphere, 463
Irreversible process, 86
Isochronism, 146
Isochronous oscillation, 146, 158
Isolated system, 34
Isothermal process, 85
Isotope, 506, 519
  separation by diffusion, 104

Jeans, Sir James, 489
Joliot, F., 503
Joule, J. P., 54, 79, 80, 84
Joule, the, 46
Joule-meter, 239

Kapitza, P., 291
K-electron capture, 509
Kelvin, the, 89, 522
Kelvin scale of temperature, 89
Kepler, J., 2, 5, 6, 364
Kepler's laws, 2, 6, 364
Kettering, C. T. and Scott, G. G., 452
Kilogramme, international prototype, 36, 522
Kilowatt-hour, 239
Kinetic energy, 48, 50, 51, 59
  and elastic collisions, 57
  and inelastic collisions, 57
Kinetic model of a gas, 94
Kinetic theory of gases, 104
Kirchhoff, G. R., 477
Kirchhoff's first law, 237, 289
Kirchhoff's second law, 255

Laminations, 328
Laser, 464
Laser clock, 415
Law, physical, 2, 5
  of gravitation, 364
  of radioactive decay, 5
  of reflection, 5
Laws, displacement, 487, 506
  of chemical combination, 106
  of electrolysis, 106, 427, 429

of electromagnetic induction, 302, 393
of mechanics, 414
of motion, 33, 39, 41, 68
of thermodynamics, 77, 81, 86
L-C circuit, 321
Lenard, P. A. von, 482, 489
Length, contraction of, 416
proper, 416, 417
Lenz's law, 295, 396
Light, coherence of, 218, 464
diffraction of, 208
dispersion of, 211
interference of, 213
monochromatic, 211
quanta, 458
reflection of, 207
refraction of, 210
speed of, 411, 414
wavelength of, 217
Lightning flash, 437
Limit, elastic, 112
Linear air-track, 24
Linear expansion, 113, 135
Linear momentum, 34, 36, 42
Linear motor, 333
Line spectrum, 459, 477
Lines of force, 349, 350
Load, electrical, 252
Logarithmic decrement, 169
Longitudinal wave, 186, 189, 204
Loop laws, 404
Lorentz, H. A., 446
Loschmidt's number, 429
L-R circuit, 318
Lucretius, 473
Lyman, C. S., 460
Lyman series, 461, 463

Magnetic effect of the current, 235, 288
Magnetic field, 346
direction, 347
energy stored in, 396
exploration of, 370
flux model for, 374
intensity (B), 293, 346
near a long straight conductor, 404
of a circular coil, 371
of a solenoid, 372, 374
strength, 292, 346
Magnetic flux, 293, 393
density, 293, 346, 369, 393
lines of, 302, 402
loops of, 402
Magnetic force constant, 371, 374
Magnetic lines of force, 402
Magnets, superconducting, 244
permanent, 346
Maltese cross tube, 439
Mapping electric fields, 349
gravitational fields, 349
Marsden, E., 490, 497
Maskelyne, N., 349
Mass, 33, 34
atomic, 106
conservation of, 474
defect, 520
energy and, 418, 520
gravitational, 36
inertial, 36, 414
measurement of, 35
number, 495, 506, 519
rest, 418
spectrometer, 506
unit, 36
Matching, 253
Matter, structure of, 12
Maximum power theorem, 254
Maxwell, J. C., 103, 340, 350, 401, 489 (quoted), 341, 350, 431
Maxwell's corkscrew rule, 348
distribution, 103
equations, 402, 405
model of the electromagnetic field, 401
Mean free path, 101
Mechanics, statistical, 90
Mendeleyev, D., 476
Metals, 114
thermal and electrical properties of, 134, 447
Metre, the, 522
Micro-miniaturization, 232
Microphone, 255
Millikan, R. A., 342, 434
Millikan experiment, 435
M.K.S. system of units, 291
Models, 5, 6, 117, 213, 350
for a gas, 11, 94
for a solid, 108, 114
for light, 208, 213
for the atmosphere, 3
for the atom, 8, 106, 495
for the electric current, 426
for the electric field, 360
for the electromagnetic field, 401
mathematical, 6
particulate, 12
Modulus, rigidity, 132
shear, 132
Young's, 112, 128
Molar conductivity, 466
Molar heat capacity, 83, 135, 452
Mole, the, 83, 105, 522
Molecular speeds, 103
Molecules, 12, 106
compound, 106
Moment, dipole, 390
of inertia, 55
Momentum, 34
and force, 42
angular, 55
changes in, 36
conservation of, 34, 55, 61, 95, 508
Monochromatic light, 211
Moon's period of rotation, 366
Morris, J. C., 455
Moseley, H. G., 490, 495
Motion, accelerated, 39
circular, 64, 65, 66, 382
harmonic, 152, 158
planetary, 364
satellite, 68, 364
wave, 199
Motor, electric, 298
a.c., 331
series-wound, 299
shunt-wound, 299, 336
Moving coil meters, 285, 294, 300
Musicians, problems for, 224
Mutual inductance, 309

Nagaoka, H., 489
Nanometre, 459, 523
National grid, 231, 253
Neutrino, 508
anti-, 508

Neutron, 504, 509
number, 519
Newton, Sir Isaac, 6, 33, 208, 414, 473, 510
Newton, the, 41
Newton-metre, the, 46
Newton's cradle, 60
first law, 33
law of gravitation, 6, 364
second law, 39, 41
third law, 41
Nodal lines, 212
Nodes, 203
Non-linear device, 335
Non-metals, 114
Non-periodic motion, 144
Normal, 207
Normal distribution, 103
modes of vibration, 204
Nuclear atom, 491
fission, 520
fusion, 520
reactions, 502, 520
structure of elements, 505
Nucleon, 505
Nucleus, composition of the, 497
models of, 508
Nuclide, 519
Numerical computation for the discharge of a capacitor, 273
gravitational potential, 377
growth of current in an inductive circuit, 311
motion of an oscillator, 156

Object, 207
Oersted, H. C., 288, 346
Ohm, G. S., 243
Ohm, the, 243
Ohmic conductors, 243
devices, 243
Ohm-meter, 244, 287
Ohm's law, 243, 447
Onnes, K., 244
Optical axis, 214
path difference, 215
Orbiting laboratory, 25
Order of spectra, 216
Organ pipes, 224
Oscillation, 145
damped, 166, 525
forced, 173, 526
fundamental modes of, 225
normal modes, 185
steady state, 174
Oscillator, 407
computing the motion of, 155
coupled, 183, 186
harmonic, 158, 159
Oscillatory circuit, 327
Oscilloscope, 386, 443
Overtone, first, 226

Parallelogram of vectors, 20
Parallel-plate capacitor, 271, 275
Particles, alpha, 487, 506
beta, 487, 507
neutral, 503
string of, 191
Particle models, for cathode rays, 478, 480
for light, 210, 213
for matter, 12, 94
Particles, waves and models, 213
Paschen, F., 460
Paschen series, 461
Pattern, 118
Pauli, W., 508
Peak value of a.c., 314
Pendulum, 146
compound, 159
simple, 159
Pendulum clock, 162
Period, 145
Periodic motion, 144
Periodic table, 476, 477
Periodic time, 145
Permeability of free space, 375
Permittivity of free space, 276, 360
relative, 360
Perrin, J. B., 105, 439
Perrin's method for determination of N_A, 105, 439
tube, 440
Phase, 149
angle, 151
change on reflection, 200
difference, 151, 316
lag, 151
lead, 151
relationships, 322
velocity, 197
Phases of matter, 108
Phasor, 317
Philosopher's stone, 510
Photography, multiflash, 27
stroboscopic, 27
Photochemical reaction, 463
Photoelectric cell, 457
as a timing device, 26
Photoelectric effect, 457, 482
Photon, 458
Physical constants, 524
Planck, M., 458, 483
Planck's constant, 458, 483
quantum theory, 458, 477, 483
Planetary motion, Kepler's laws of, 2, 5, 6, 364
Plastic yield, 112
Plimpton, S. J. and Lawton, W. E., 362
Plucker, J., 438
Pohl, R. W., 386
Polarity, 236
Polarization, 277, 344
Polymeric solids, 113
Popper, K. (quoted), 7
Population inversion, 464
Positive ray apparatus, 506
Potential, critical, 455
electric, 385, 390
gravitational, 376, 377, 380
Potential energy, 48, 53, 54, 79, 163, 376, 385, 390
Potential change, 379
difference, 240
divider, 261
gradient, 379, 385
wave, 408
well, 381
Potentiometer, 262
linear, 262
slide wire, 263
Powder method of X-ray analysis, 125
Power, 46, 248
absorption, 180
factor, 324
loss, 253

ratings, 248
theorem, 254
Poynting's vector, 413
Prefixes for S.I. units, 523
Pressure, 12, 94, 97, 98
Principle of superposition, 199
Problems for musicians, 224
Projectile motion, 67, 353
Proton, 500, 506, 511
Pucks, 25
Pulse, 144, 187
circular, 205
of electric field, 412
straight, 205

Quadrature, 150
Quantity of charge, 238
Quantum theory, 458, 477, 483
Q-factor, 179

Radiation, black body, 482
characteristic, 495
electromagnetic, 213, 407
gamma, 487, 508
Radioactive substance, 487
Radioactivity, 485, 506, 507, 508
Random distribution, 13, 103
Random walk, 100
Rate of change, 18
of momentum, 41
Ratio of conductivities, 448
Rayleigh, Lord J. W., 489
R-C circuit, 321
Reactance, 320, 323
Reaction, 42, 341
chain, 520
Rectification, 279
Reed switch, 272
Reference, frames of, 61
Reflection, 199
at a barrier, 206
grating, 216, 218
hard, 200
law of, 207
partial, 201
phase change on, 200
soft, 200
Refraction of light, 210
of ripples, 210
Reines, F., 508
Relative permittivity of free space, 272, 276
of a medium, 360
Relative velocity, 22
Relativity, special theory of, 64, 414
general theory of, 64
Galilean, 414
Relaxation time, 446
Resistance, 241, 243, 244
Resistances in series, 267
in parallel, 267
Resistivity, 244, 445, 447
temperature coefficient of, 246
Resistor, 261
light-dependent, 263, 281
photo-sensitive, 281
swamping, 286
thin film, 247
Resonance, 175
electron spin, 182
mechanical, 175
nuclear magnetic, 182
spectra, 178
Resonance curve, width of, 178
Responder, 174
Rest mass, 418
Resultant, 20
Reversible cell, 251
Reversible process, 86
Rigidity modulus, 132
Ripple tank, 196, 205, 212
Ripples, 195
diffraction of, 208
dispersion of, 197, 211
interference of, 212
propagation of, 205
reflection of, 206
refraction of, 210
Ritz, W., 478
R-L circuit, 320
Röntgen, W. C., 438, 478, 480
Root mean square values, 315
Rotation, 149
energy stored in, 54
Rotor, 297, 332
Rumford, Count, 83
Rutherford, Lord E., 8, 472, 485, 489, 510
Rutherford, E. and Geiger, H., 434
Rutherford model of the atom, 8, 490
Rydberg, J., 460, 478
Rydberg's constant, 460

Satellite, 67, 364
Scalar quantity, 19
Scale, well tempered, 225
Scattering, of particles, 488, 490
of waves, 209
Scholastic philosophy, 7
Scott, G. G. 452
Search coil, 303, 370
Second, the, 522
Second law of Newton, 39, 41
of thermodynamics, 86
Selectivity, 179
Self-energy, 381
Self inductance of a solenoid, 398
Self induction, 396
Seismic wave, 189
Semi-conductor, 114
Sensitivity of a galvanometer, 300
Shear modulus, 132
Shearing stress, 132
Sign convention for circuit calculations, 251
for potential, 379
Simple harmonic motion, 152, 155, 158, 163
damped, 166, 525
forced, 173, 526
Simple harmonic oscillator, 152
acceleration of, 152
computing the motion of, 155
energy of, 163
equation of motion, 158
Simple pendulum, 159
Single phase supply, 333
Sinusoidal motion, 147, 154
waves, 189
Sinusoidal oscillator, 152
equation of motion, 158
S.I. units, 41, 291, 522
Size, critical, 520
Smoothing circuit, 325
Snell's law, 5
Soddy, F., 490, 495 (quoted), 506
Solenoid, 372, 374, 396, 398

Solid, a model for, 108, 114
as ordered structure, 120
electrical properties of, 110, 447
physical properties of, 110
polymeric, 113
thermal properties of, 110, 447
Sound, 213, 222–4
beats, 223
diffraction, 208, 209, 222
interference, 223
speed, 98, 222
stationary wave, 204
Space, free, 211
permeability of, 375
permittivity of, 276, 360
Specific charge, 442, 481
measurement (Thomson's method), 441
of electron, 442, 481
of hydrogen ion, 442, 481
Specific heat capacity, 83, 135
Spectrometer, absorption, 181
infrared, 181
mass, 506
Spectroscope, 477
Spectrum, 211, 216
absorption, 462
continuous, 211
line, 459, 477
of the sun, 463
order of, 216
X-ray, 495
Speed, 19
drift, 446
mean square, 96
measurement of, 26
molecular, 103
of compression wave, 188
of light, 411, 414
of sound, 98, 222
of transverse pulse, 193
Spring of the air (Boyle), 8
Squirrel cage rotor, 332
Standing wave, 202, 204
State, ground, 461
equilibrium, 76
excited state, 462
metastable, 465
stationary, 462
Step-by-step solution of differential equations, 156, 273, 311, 377
Statistical mechanics, 90
Stern, O., 103
Stiffness, 111
Stokes' law for viscous friction, 434
Stoney, J., 351, 433
Strain, 112, 127
energy, 128
gauge, 265
Strength, 111
Stress, 112, 127
breaking, 131
concentration, 131
Stress-strain behaviour, 112
Stringed instruments, 204, 226
Stroboscopic photography, 27
Sublimation, heat of, 128
Summary of terms, 519
Superconduction, 244
Superposition, principle of, 199
Surface energy, 129
tension, 129
Surfaces, formation of, 128
Système International d'Unités, 41, 292, 522

Taylor, Sir G. I., 133, 469
Television tube, 443
Temperature, 12, 76, 79
absolute zero of, 89
and kinetic theory, 104
dependent properties, 76
effects of, 113
glass transition, 113
ideal gas scale, 78, 90
Kelvin scale, 77, 88
scale, 77
thermodynamic scale, 88, 89
Tensile strength, 112, 130
Tension, surface, 129
Tesla, N., 292
Tesla, the, 292, 346, 394
Terminal velocity, 40, 434
Theory, 5
and models, 136, 402
atomic, 475, 489, 495
caloric, 83
Drude's, 445
kinetic, 104
of the atmosphere, 3, 8
of gravitation, 6, 364
of relativity, 64, 414
quantum, 483
wave, 213, 482, 484
Thermal conductivity, 135, 447
Thermal efficiency, 87
Thermal equilibrium, 77
Thermal expansion, 113, 135
Thermionic emission, 439
applications of, 442
Thermionic tubes (gas filled), 455
Thermionic vacuum tubes, 279, 442
Thermistor, 280
Thermodynamics, 90
first law, 81
second law, 86
zeroth law, 77
Thermoelectricity, 452
Thermometer, constant volume gas, 77
Third law of Newton, 41
Thomson, Sir J. J., 434, 438, 439, 442, 446, 480
Thomson effect, 452
Thomson's method for specific charge, 441
model of the atom, 487
Thought experiment, 79
Three-dimensional close packing, 117
Three phase electrical supply, 332
Threshold frequency, 458
Thyratron, 455
Ticker tape timer, 26
Time, 415
dilation, 416
effect on behaviour of materials, 113
proper, 415, 417
relaxation, 446
Time constant of *R-C* circuit, 275
Timing methods, 26
Torque, 296
Torricelli, 3, 8
Torsion balance, 367
Townley, R., 9
Townsend, J. S., 434
Transformers, 308, 328, 329
Transient, 174
Transistor, 232, 334

Transmission grating, 216
Transmutation of elements, 500, 502
Transport number, 430
Transverse wave, 189, 191
Triangle of vectors, 20
Triple point of water, 77
Trolley, dynamics, 24
Tsunami, 196
Tuned circuit, 337
Tuning fork, 204

Ultimate speed, 414, 417
Ultraviolet radiation, 213, 458, 460
  absorption in the atmosphere, 463
Uncalibrated ammeter, 236
Unit cell, 118
Units, systems of, 41, 291, 522
Universal gas constant, 97
Universal gravitational constant, 365

Valency, 106
Van der Broek, A., 495
Van de Graaff generator, 362
Variation of resistance, with temperature, 452
Vector field, 343
Vectors, 19
  addition of, 19, 20
  and navigation, 22
  components of, 23
  cross-product, 293
  resolution of, 23
  rotating, 317
  signs and, 22
Velocity, 19
  angular, 55, 150
  group, 197
  of a pulse, 193
  of water waves, 195
  phase, 197
  terminal, 40, 434
Very high frequency oscillations, 407
Vibration, 144
  damped, 166
  complex mode, 204
  longitudinal, 186
  normal modes, 204
  of air columns, 222
  of strings, 202
  transverse, 191
Virial coefficient, 97
Viscous damping, 167
Volt, the, 240, 385, 426
Voltage, 240
  breakdown, 279
  current characteristic, 278
  lag, 322
  measurement, 282
  r.m.s., 324
Voltameter, 428
Voltmeter, 386
  multi-range, 285
  resistance of, 283
  transistor, 284
  valve, 283
Von Laue, M., 122
Von Laue's method, 125
Volume expansion, 97

Walton, E. T. S., 490
Water analogy for the electric current, 237, 242
Water, specific heat capacity of, 83
Watt, the, 46, 248
Wave guide, 411
Wavelength, 190
Wave machines, 189
Wave model, 213, 480
  for cathode rays, 478, 480
  for light, 210, 480
Wave velocity, 195, 197
Wave-particle duality, 484
Waves, behaviour of, 199, 205
  diffraction, 121, 208, 210
  interference, 121, 212, 213, 223
  propagation, 205, 208
  reflection, 199, 200, 206, 207
  refraction, 210
Waves, longitudinal, 186, 204
  stationary, 202, 204
  transverse, 189, 191
  travelling, 186, 189, 191
Water waves, 194
Weber, the, 393
Webster, H. C., 503, 505
Weight, 42
Weightlessness, 67, 365
Westgren, A., 106
Wiedemann, G. H. and Franz, R., 445, 448
Wilson, C. T. R., 434, 510
Wilson, H. A., 434
Wilson cloud chamber, 512
Wind instruments, 204, 224
Work, 44, 79, 80, 87
  and energy, 44
  units of, 45
Work function, 458
Work hardening, 131, 133

X-rays, 124, 438, 478
  applications of, 444
  crystallography, 125, 218
  diffraction, 125
  spectra, 495
  tube, 444

Yield, 112
Young's modulus, 112
Young's slits, 213
Young, T., 208, 213

Zartmann, I. F., 103
Zener diode, 280
Zero, absolute, 89
Zeroth law of thermodynamics, 77
Ziman, J. M. (quoted), 446

Printed in Great Britain at the University Press, Cambridge CDE798

SELECTED "WORLD STUDENT SERIES" ECONOMY EDITIONS OF OUTSTANDING UNIVERSITY TEXTBOOKS FOR STUDENTS OF PHYSICS

FEYNMAN LECTURES ON PHYSICS
By R. P. Feynman and R. B. Leighton, *California Institute of Technology,* and M. L. Sands, *Stanford University*

Vol. 1: Mainly Mechanics, Radiation, and Heat	*519 pp, 342 illus*
Vol. 2: The Electromagnetic Field	*525 pp, 449 illus*
Vol. 3: Quantum Mechanics	*376 pp, 192 illus*

FUNDAMENTAL UNIVERSITY PHYSICS

Vol. I Mechanics	*463 pp, 399 illus*
Vol. II Fields and Waves	*530 pp, 546 illus*
Vol. III Quantum and Statistical Physics	*598 pp, 350 illus*

PHYSICS *760 pp, 843 illus*
All by Marcelo Alonso, *Organization of American States,* and Edward J. Finn, *Georgetown University*

COLLEGE PHYSICS, Third Edition *1024 pp, 765 illus*
UNIVERSITY PHYSICS, Fourth Edition *671 pp, 96 illus*
All by Francis W. Sears, *Dartmouth College,* and Mark W. Zemansky, *the City College of the City of New York*

MECHANICS, Third Edition
By Keith R. Symon, *University of Wisconsin* *639 pp, 168 illus*

INTRODUCTION TO NUCLEAR PHYSICS
By Harald Enge, *Massachusetts Institute of Technology* *592 pp, 258 illus*

PHYSICS FOR BIOLOGY AND PREMEDICAL STUDENTS
By Desmond M. Burns and Simon G. G. MacDonald, *University of Dundee* *614 pp, 372 illus*

ADDISON-WESLEY ORIGINAL SCIENCE PAPERBACKS

SPACE, TIME AND RELATIVITY
By Rolf Nevanlinna, *Academy of Finland* *158 pp*

VECTOR ALGEBRA AND MECHANICS: THEORY, PROBLEMS AND SOLUTIONS
By R. Capildeo, *University of London* *268 pp*

EXPERIMENTAL MEASUREMENTS: PRECISION, ERROR AND TRUTH
By N. C. Barford, *University of London* *143 pp*

PROBLEMS AND SOLUTIONS IN GENERAL PHYSICS FOR SCIENCE AND ENGINEERING STUDENTS
By S. G. G. MacDonald, *University of St. Andrews* *276 pp*

PARTICLES AND THEIR INTERACTIONS
By J. G. Powles, *University of Kent at Canterbury* *260 pp*

ADDISON-WESLEY PUBLISHERS LIMITED
LONDON · READING, MASSACHUSETTS · AMSTERDAM · SINGAPORE · SYDNEY · TOKYO

201 08607 7