FOREWORD TO THE SERIES

BY

F. C. BROWN

Editor: Senior Science Master,
The Perse School, Cambridge

Science advances, these days, at a bewildering pace and teachers of science are forced to adopt fresh viewpoints and new techniques. The last few years have witnessed a fresh impetus to the teaching of science. The Association for Science Education has produced highly stimulating proposals for new syllabuses, the Nuffield Foundation is generously sponsoring an exhaustive survey of new teaching material and methods, and many Examination Boards are making considerable revision to their syllabuses in science.

Such re-appraisal is invigorating to all concerned, but the world of textbooks must also take account of such changes. This is a complete new series of texts, in physics and chemistry, which accepts the challenge of this revolution in teaching methods and is designed to meet—both now and for the next few years—the needs of schools and of the student about to proceed to University.

One of my tasks, as Editor of the series, has been to seek out gifted and enthusiastic teachers able to communicate the modern approach, new developments, the basic simplicity and symmetry underlying so much of their subject, and the elegance and vital importance of practical work integrated with theory. Whatever criticisms are made of these books, they will not be condemned as 'the mixture as before'—each has been written afresh out of the author's experience.

In each subject a related set of texts is to be published. Each set will comprise:

1. A pair of books designed to cover a complete course comfortably up to the standard of G.C.E. O level. Each book is richly leavened with elegant and practicable experiments for both student and teacher.

2. A theoretical book giving a thoroughly up-to-date treatment of the subject up to G.C.E. A level, or any comparable examination.

3. A theoretical book taking the subject a stage further than the A level book and, therefore, well suited to the requirements of the student who is proceeding to University or contemplating such examinations as the A.R.I.C.

4. A practical book of experiments matched to both the advanced books described above. This will contain both standard experiments and a wide selection of more original problems.
FOREWORD

Although the texts are related to each other, each text is complete in itself. We hope that teachers and students, having found satisfaction in one text, will turn confidently to others in the series for an extension of their interests and activities.
PREFACE

This book is intended for students who have completed a General Certificate of Education Advanced level course in Physics and wish to continue their study of the subject in greater depth. It is therefore suitable for those preparing to take the G.C.E. Special papers; for candidates for University Entrance Scholarship examinations; and for any student who intends to take a University course in Physics. Although written primarily as part of a series, the book is self-contained; all that is assumed is that the student has covered an Advanced level course similar to that advocated by the Association for Science Education, including a grounding in atomic and nuclear physics.

In writing the book I have had in mind that some students pursue the subject well beyond the Advanced level stage in the Sixth Form, whereas many leave school with only the minimum knowledge required and need to undertake further reading during the vacation before going up to University. I hope that this volume will be found suitable for both categories.

The subject matter may be grouped under two headings: (a) topics which have already been studied, treated here in greater depth, (b) new topics chosen for their interest and relevance in present day physics. Chapters 1–6 inclusive and Chapter 9 cover the necessary extensions in classical physics required by students preparing for the examinations mentioned previously. Chapters 11 and 12 serve the same purpose in modern physics. The chief reason for including Chapter 12 is to lead the student to a deeper understanding of the periodicity of atomic structure, a topic which must of necessity be treated very arbitrarily at lower levels. Chapters 7, 8, 13 introduce topics which are new to the reader and it is hoped that he will be encouraged to consult the books mentioned in the bibliography for further detail.

In all sections in which electrical quantities are involved care has been taken to state clearly the units used. Where a choice between units has had to be made the c.g.s. system has been used; this occurs in certain sections in Chapters 9 and 12, but the student familiar with m.k.s. units should also be able to follow these sections if he makes use of Tables 10.2 and 10.3.

The purpose of Chapter 10 is to enable the reader who has followed an A level electricity course in c.g.s. units to find out about the m.k.s. system so that he may be able to read the more advanced texts written in these units. Given the present situation in electricity teaching it is certainly an advantage for a student to be conversant with both of these systems of units.
PREFACE

I am grateful to Mr F. C. Brown, the Editor of the series, for his help and encouragement during the writing of this book, and to Dr P. F. Scott, Fellow of Sidney Sussex College, Cambridge, who read through the typescript and made several valuable suggestions and criticisms. My thanks are also due to my wife for preparing the typescript; and to three of my former pupils, J. Clarke, S. R. Coslett and J. M. R. Hatfield, for their help in completing and checking it.

I am indebted to the International Nickel Company (Mond) Limited for permission to reproduce the version of the Periodic Table used in Table 12.2.

D. W.
CONTENTS

Foreword v
Preface vii

1 ANGULAR MOTION AND GRAVITATION 1

Angular Motion
1.1 Angular Velocity 1
1.2 Angular Acceleration 1
1.3 Kinetic Energy of Rotation, Moment of Inertia 2
1.4 Moments of Inertia of Certain Bodies 3
1.5 Parallel Axes Theorem 3
1.6 Perpendicular Axes Theorem 3
1.7 Kinetic Energy of a Rolling Body 4
1.8 Angular Momentum 6
1.9 Equations for Angular Motion 6
1.10 Angular Simple Harmonic Motion. Oscillation of a Rigid Body suspended by a Torsion Wire 8
1.11 Oscillation of a Rigid Body about a Horizontal Axis 9
1.12 Compound Pendulum 9
1.13 The Reversible Pendulum. Kater's Pendulum 12

Gravitation
1.14 Newton's Law of Gravitation 13
1.15 Gravitational Field and Potential 13
1.16 Potential due to a Point Mass 14
1.17 Relation between Field and Potential 15
1.18 Uses of the Concept of Gravitational Potential 15
1.19 Gravitational Potential and Field due to a Thin Spherical Shell 16
1.20 Fields Produced by a Thick Spherical Shell and a Solid Sphere 18
1.21 Escape Velocity 19
1.22 Variation of $g$ with Altitude and Depth 20
1.23 Variation of $g$ with Latitude 21

2 SURFACE TENSION AND VISCOSITY 24

Surface Tension
2.1 Introduction 24
2.2 Definition 24
2.3 Surface Energy 24
2.4 Contact between Liquid and Solid. Angle of Contact 25
CONTENTS

2.5 Pressure Difference across a Surface of any Shape 26
2.6 Applications of the Formula for Pressure Difference across a Curved Surface 27
2.7 Vapour Pressure Above a Curved Surface 27
2.8 Effect of Surface Tension on Condensation and Evaporation of Drops 29
2.9 Effect of Surface Tension on Boiling 29
2.10 Effect of Electrically Charged Particles on Condensation 30

Viscosity 30
2.11 Introduction 30
2.12 Coefficient of Viscosity 30
2.13 Streamline and Turbulent Flow 32
2.14 Critical Velocity, Reynolds’ Number 32
2.15 Flow of Liquid through a Tube, Poiseuille’s Formula 33
2.16 Application of Poiseuille’s Formula to the Determination of the Viscosity of Water 35
2.17 Cylinder in a Rotating Liquid 37
2.18 Determination of the Coefficient of Viscosity of a Very Viscous Liquid 39
2.19 Viscosity of Gases 40
2.20 Measurement of the Coefficient of Viscosity of a Gas 41

3 OSCILLATIONS AND WAVES 43

Oscillations 43
3.1 Simple Harmonic Motion 43
3.2 Damped Simple Harmonic Motion 44
3.3 Forced Simple Harmonic Motion – No Resistance 47

Waves 49
3.4 Wave Motion 49
3.5 Harmonic Waves 49
3.6 The Equation of a Harmonic Wave 50
3.7 The Equation of Wave Motion 52
3.8 Stationary Waves 53
3.9 Beats 54
3.10 Interference 55
3.11 Vector Method for the Addition of Waves 56
3.12 Dispersion 56
3.13 Group Velocity 57
3.14 Modulation 58

4 WAVE PROPERTIES OF LIGHT 61

Introduction 61
4.1 Light as a Wave-motion 61
4.2 Huygens’ Construction 61
## CONTENTS

4.3 Optical Path 62
4.4 Derivation of the Thin Lens Formula 62
4.5 Fermat’s Principle 63
4.6 Fermat’s Principle Applied to Refraction and Reflection at a Plane Surface 64

### Interference

4.7 Double-slit Fringes 65
4.8 Intensity Distribution in Young’s Fringes 66
4.9 Effect of Finite Width of Source Slit 67
4.10 Conservation of Energy 69
4.11 Interference Effects Produced by Thin Films 70
4.12 Thin Films – Reflected Light 71
4.13 Thin Films – Transmitted Light 73
4.14 White Light Fringes 74
4.15 Non-reflecting Coatings – ‘Blooming’ 75
4.16 Wedge Fringes 75
4.17 Newton’s Rings 75
4.18 Classification of Fringes 77
4.19 Interferometry 78

### Diffraction

4.20 Fresnel and Fraunhofer Diffraction 80
4.21 Fraunhofer Diffraction at a Single Slit 80
4.22 Diffraction by a Rectangular Aperture 83
4.23 Resolving Power 83
4.25 Resolving Power of a Microscope 85
4.26 Diffraction Grating 87
4.27 The Resolving Power of a Diffraction Grating 90
4.28 Echelon Gratings 91
4.29 Fresnel Diffraction. Half-period Zones 91
4.30 Fresnel Diffraction at a Circular Aperture 94
4.31 Fresnel Diffraction at a Circular Obstacle 95
4.32 Zone Plate 95
4.33 Cylindrical Wave-fronts 96
4.34 Fresnel Diffraction at a Straight Edge 97
4.35 Rectilinear Propagation 99

## 5 HEAT TRANSFER

### Conduction

5.1 Coefficient of Thermal Conductivity 101
5.2 The Wiedemann–Franz Law 101
5.3 Flow of Heat along an Unlagged Bar 102
5.4 Conduction through Composite Slabs 104
CONTENTS

5.5 Conduction through the Walls of Pipes 106
5.6 Conduction through the Wall of a Hollow Sphere 108
5.7 Conductivity of Liquids and Gases 108

Convection 108
5.8 Introduction 108
5.9 Forced Convection. Newton's Law of Cooling 109
5.10 Natural Convection 109

Radiation 110
5.11 Thermal Radiation 110
5.12 Definition of Terms 110
5.13 Black-body Radiation 110
5.14 Kirchhoff's Law 111
5.15 Stefan's Law 112
5.16 Energy Distribution in the Black-body Spectrum 113
5.17 Wien's Laws 114
5.18 The Rayleigh-Jeans Formula 114

6 KINETIC THEORY 116

Gases 116
6.1 Pressure of a Gas 116
6.2 Deductions from the Pressure Formula 117
6.3 Number of Molecules Crossing a Plane in a Gas 118
6.4 Maxwell's Law of Distribution of Velocities 120
6.5 Equipartition of Energy 121
6.6 Specific Heats of Gases 121
6.7 Variation of Specific Heats of Gases with Temperature 123
6.8 Mean Free Path 124
6.9 Viscosity 125
6.10 Thermal Conductivity 127

Liquids 128
6.11 Brownian Motion 128
6.12 Perrin's Experiments. Determination of Avogadro's Number 129

Solids 131
6.13 Atomic Heat. Dulong and Petit's Law 131

7 THERMODYNAMICS 133
7.1 Introduction 133
7.2 The First Law of Thermodynamics 133
7.3 Work Done during an Expansion 133
7.4 Reversible and Irreversible Changes 135
7.5 Cycles of Changes 136
CONTENTS

7.6 Carnot Cycle 137
7.7 The Carnot Refrigerator 138
7.8 The Second Law of Thermodynamics 139
7.9 Carnot’s Theorem 140
7.10 Kelvin Absolute Temperature Scale 141
7.11 Relation between the Kelvin Scale and the Perfect Gas Scale 142
7.12 Entropy 144
7.13 Reversible Processes 146
7.14 Entropy Change in an Irreversible Cycle 146
7.15 Irreversible Processes – The Principle of Increase of Entropy 147
7.16 Unavailable Energy 148
7.17 Theoretical Use of Reversible Cycles 149
7.18 The Clausius–Clapeyron Equation 149
7.19 Surface Energy 151

8 LOW TEMPERATURES 153
8.1 Introduction 153

**Liquefaction of Gases** 153
8.2 Early Methods 153
8.3 The Joule–Kelvin Effect 155
8.4 The Hampson and Linde Processes 156
8.5 The Claude and Heylandt Processes 157
8.6 Helium Liquefiers 158

**Production and Measurement of Very Low Temperatures** 158
8.7 Liquid Helium Under Reduced Pressure 158
8.8 Adiabatic Demagnetization 159
8.9 Measurement of Low Temperatures 160

**Some Low-temperature Phenomena** 161
8.10 Liquid Helium 161
8.11 Solid Helium 163
8.12 Superconductivity 163

9 FURTHER TOPICS IN ELECTRICITY 165

**Electrostatics** 165
9.1 Gauss’ Theorem 165
9.2 Applications of Gauss’ Theorem 166
9.3 Capacitance of a Coaxial Cable 167
9.4 Force on the Surface of a Charged Conductor 167
9.5 The Attracted Disc Electrometer 168
9.6 The Quadrant Electrometer 169
## CONTENTS

9.7 Use of the Quadrant Electrometer to Measure a Very Small Current 170

### Electromagnetism

9.8 Magnetic Field on the Axis of a Circular Coil 171
9.9 Magnetic Field on the Axis of a Solenoid 172
9.10 Self-induction 173
9.11 Mutual Induction 173
9.12 Theory of the Ballistic Galvanometer 174

### Electrical Circuits

9.13 Kirchhoff’s Laws 176
9.14 D.C. Circuit containing Capacitance and Resistance 178
9.15 D.C. Circuit containing Resistance and Inductance 180
9.16 Discharge of a Capacitor through an Inductance 181
9.17 A.C. Circuit containing Resistance only 182
9.18 A.C. Circuit containing Capacitance only 182
9.19 A.C. Circuit containing Inductance only 183
9.20 A.C. Circuits containing either $L$ or $C$ with $R$ 183
9.21 A.C. Circuit containing $L$, $C$ and $R$ in Series 184
9.22 Algebraic Treatment of the $L$, $C$, $R$ Series Circuit 185
9.23 Use of Complex Numbers in A.C. Circuit Problems 186
9.24 Power in A.C. Circuits 187
9.25 The Transformer 188

## 10 ELECTRICITY IN RATIONALIZED M.K.S.

### UNITS

10.1 Introduction 190
10.2 Mechanical Units 190
10.3 Rationalized M.K.S. in Electricity 191

### Electromagnetism

10.4 Ampère’s Experiments 191
10.5 Flux Density (Magnetic Induction) 193
10.6 Definition of the Ampere 194
10.7 Definition of the Coulomb 195
10.8 Magnetic Flux 195
10.9 Electromagnetic Induction 196
10.10 Magnetizing Force 197
10.11 Magnetic Materials 198
10.12 Magnetic Circuits 198
10.13 Magnetic Moment 200

### Electrostatics

10.14 Attraction between Point Charges 202
10.15 Intensity of Electric Field and Potential Difference 203
10.16 Electric Flux and Flux Density 203
CONTENTS

10.17 Relation between $D$ and $E$ in a vacuum 204
10.18 Dielectrics 204
10.19 Capacitance 204

11 THE QUANTUM THEORY 208
11.1 Black-body Radiation 208
11.2 The Photo-electric Effect 208
11.3 Waves or Photons? 210
11.4 Matter Waves 210
11.5 Wave Mechanics 211
11.6 Cause and Effect 211
11.7 Heisenberg Uncertainty Relation 212
11.8 The Bound Electron. Energy Levels 214
11.9 The Schroedinger Equation 215

12 ATOMIC SPECTRA AND THE PERIODIC TABLE 217
12.1 Hydrogen Spectrum 217
12.2 The Bohr Atom 218
12.3 Explanation of Atomic Spectra by the Bohr Atom 219
12.4 The Frank–Hertz Experiment 221
12.5 Quantization of Angular Momentum 223
12.6 Radii of the Bohr Orbits 224
12.7 Energy Levels in Hydrogen. Ionization Potential 224
12.8 Application of Wave Mechanics 225
12.9 Elliptic Orbits 226
12.10 Effect of ‘Elliptic Orbits’ on Spectra 228
12.11 The Zeeman Effect. Magnetic Quantum Number 229
12.12 The Spinning Electron 230
12.13 Total Angular Momentum 231
12.14 The Pauli Exclusion Principle and the Periodic Table 232

13 ELECTRICAL AND MAGNETIC PROPERTIES OF MATTER 236

Electrical Properties 236
13.1 Classification 236
13.2 Drude’s Theory of Metallic Conduction 236
13.3 The Hall Effect 239
13.4 Application of the Quantum Theory. Fermi Energy 240
13.5 Electronic Work Function and Contact Potential 243
13.6 Energy Bands 243
13.7 Intrinsic Semi-conductors 245
13.8 Impurity Semi-conductors 246
13.9 The Thermistor 248
13.10 The Junction Diode 248
CONTENTS

13.11 Metal – Semi-conductor Diodes 250
13.12 The Junction Transistor 250
13.13 Applications of Semi-conductor Devices 251
13.14 Dielectrics 252
13.15 Thermo-electric Phenomena 253

Magnetic Properties 255
13.16 Definitions 255
13.17 Classification 256
13.18 Properties of Ferromagnetics 257
13.19 Experimental Determination of the Hysteresis Loop 258
13.20 Explanation of the Magnetic Properties of Matter 258

BIBLIOGRAPHY 260
INDEX 261
ANGULAR MOTION AND GRAVITATION

ANGULAR MOTION

1.1 Angular Velocity

Fig. 1.1 shows a section through a rigid body which is rotating about an axis through O perpendicular to the plane of the diagram. A point A, distance \( r \) from O is moving in a circle of radius \( r \). After time \( \delta t \), A will have moved to A' and the radius OA will have turned through an angle \( \delta \theta \). The ratio \( \delta \theta / \delta t \) is the average angular velocity during the time \( \delta t \). To define the angular velocity at any instant, \( \delta t \) is made to tend to zero so that \( \delta \theta / \delta t \) becomes \( d\theta / dt \).

\[
\text{Angular velocity} = \frac{d\theta}{dt}.
\]

Angular velocity is usually denoted by \( \omega \).

If \( \omega \) is independent of time the body is said to rotate with constant angular velocity.

\( \theta \) is measured in radians, so that \( \omega \) is in radians sec\(^{-1}\). Angular velocities may also be measured in revolutions sec\(^{-1}\). As there are \( 2\pi \) radians in one revolution, an angular velocity of \( \pi \) revolutions sec\(^{-1}\) is \( 2\pi \) radians sec\(^{-1}\).

If the arc length AA' is denoted by \( \delta s \), then

\[
\delta s = r \delta \theta,
\]

and

\[
\frac{\delta s}{\delta t} = r \frac{\delta \theta}{\delta t}.
\]

If \( \delta t \to 0 \), the above equation becomes

\[
\frac{ds}{dt} = r \frac{d\theta}{dt} \quad \quad \quad \quad \quad \quad \quad \quad (1.1)
\]

Now \( ds/dt \) is the linear speed \( (v) \) of A at any instant and \( d\theta/dt \) its angular velocity at that instant.

Thus:

\[
v = r \omega.
\]

1.2 Angular Acceleration

The angular acceleration of A is defined as the rate of change of angular velocity with time.

\[
\text{Angular acceleration} = \frac{d\omega}{dt} = \frac{d^2 \theta}{dt^2}
\]
Angular acceleration is measured in radians sec$^{-2}$.

Differentiating equation (1.1) with respect to time gives, for the relation between angular and tangential accelerations:

$$\frac{d^2\theta}{dt^2} = r \frac{d^2\phi}{dt^2},$$

i.e.

$$\frac{dv}{dt} = r \frac{d\omega}{dt}.$$

The tangential acceleration $dv/dt$ arises only if the speed, $v$, of A is changing. There is also a radial acceleration directed towards O, even if $v$ is not changing. This is given by $v^2/r$ or $\omega^2 r$. (It is assumed that the reader is already familiar with the proof of these formulae.)

_Notation._ It is usually more convenient to adopt the 'dot' notation to indicate differentiations with respect to $t$ so that $d\theta/dt$ becomes $\dot{\theta}$; $d^2\theta/dt^2$ becomes $\ddot{\theta}$; etc.

1.3 Kinetic Energy of Rotation. Moment of Inertia

To form an expression for the kinetic energy of rotation of a rigid body we imagine the body to be made up of a large number of elementary masses of which P (Fig. 1.2) is one. Suppose this has mass $m$ and is at a perpendicular distance $r$ from the axis through O about which the body is rotating with constant angular velocity $\omega$. The linear velocity $v$ of P is given by $v = r\omega$, and its kinetic energy, which is ${1\over 2}mv^2$, is thus ${1\over 2}mr^2\omega^2$.

To compute the kinetic energy of the whole body we must sum the contributions from all the elementary masses such as P, thus:

\[
\text{Kinetic Energy of body} = \Sigma{1\over 2}mr^2\omega^2
= {1\over 2}\omega^2\Sigma mr^2,
\]

as $\omega$ is the same for all parts of the body. The quantity $\Sigma mr^2$ is known as the _moment of inertia_ of the body about the axis through O, and is denoted by $I$.

Thus:

\[
\text{Kinetic Energy} = {1\over 2}I\omega^2.
\]

In the c.g.s. system $I$ is in g cm$^2$ and in the m.k.s. system it is in kg m$^2$.

If a body of mass $M$ has a moment of inertia $I$ about a certain axis it is said to have a _radius of gyration_ $k$ about that axis, defined by the equation

\[
I = Mk^2.
\]

Obviously, $k$ is the distance from the axis at which the whole mass of the body would have to be concentrated to give the same moment of inertia as the body itself.
1.4 Moments of Inertia of Certain Bodies

The moment of inertia of a regular body about a given axis may be calculated from the expression \( I = \Sigma mr^2 \). For the details of these calculations the student is referred to the standard texts on Calculus or Dynamics; it is sufficient for our purpose to quote some of the most important results (Table 1.1).

<table>
<thead>
<tr>
<th>Body of mass ( M )</th>
<th>Axis</th>
<th>Moment of inertia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform thin rod of length ( 2a )</td>
<td>Perpendicular to the rod through the centre of gravity</td>
<td>( \frac{1}{3}Ma^2 )</td>
</tr>
<tr>
<td>Uniform thin rod of length ( 2a )</td>
<td>Perpendicular to the rod through one end</td>
<td>( \frac{2}{3}Ma^2 )</td>
</tr>
<tr>
<td>Uniform disc of radius ( a )</td>
<td>Perpendicular to the plane of the disc through the centre</td>
<td>( \frac{1}{2}Ma^2 )</td>
</tr>
<tr>
<td>Ring of radius ( a )</td>
<td>Perpendicular to the plane of the ring through the centre</td>
<td>( Ma^2 )</td>
</tr>
<tr>
<td>Rectangular lamina with sides of length ( 2a ) and ( 2b )</td>
<td>Perpendicular to the plane of the lamina through the centre</td>
<td>( \frac{M}{3} (a^2 + b^2) )</td>
</tr>
<tr>
<td>Sphere of radius ( a )</td>
<td>Through the centre</td>
<td>( \frac{4}{3}Ma^2 )</td>
</tr>
</tbody>
</table>

Table 1.1 Moments of Inertia of Certain Bodies

1.5 Parallel Axes Theorem

Most of the moments of inertia in Table 1.1 are for axes through the centre of gravity of the body. If it is required to calculate the moment of inertia about any other axis parallel to an axis through the centre of gravity, then this may be done by application of the parallel axes theorem, viz.:

\[ I_O = I_G + Ma^2, \]

where \( I_O \) is the required moment of inertia about a certain axis,

\( I_G \) is the moment of inertia about a parallel axis through the centre of gravity,

\( M \) is the mass of the body,

\( a \) is the perpendicular distance between the two axes.

For example, for a rod of mass \( M \) and length \( 2a \),

\[ I_G = \frac{1}{3}Ma^2 \quad \text{(see Table 1.1)}. \]

The moment of inertia \( I_O \) about an axis perpendicular to the rod through one end is, by the parallel axes theorem, given by

\[ I_O = \frac{1}{3}Ma^2 + Ma^2 = \frac{4}{3}Ma^2, \]

as quoted in Table 1.1.

1.6 Perpendicular Axes Theorem

This theorem applies to a lamina (Fig. 1.3). \( OX, OY \) are perpendicular axes in the plane of the lamina, and \( OZ \) the axis through \( O \) perpendicular to the lamina.
The theorem states that
\[ I_z = I_x + I_y, \]
where \( I_x, I_y, I_z \) are the moments of inertia of the body about the axes \( OX, OY, OZ \) respectively.

![Fig. 1.3](image1)

![Fig. 1.4](image2)

As an example of the application of this theorem consider a disc of mass \( M \), centre \( O \) and radius \( a \) (Fig. 1.4). Its moment of inertia about \( OZ \) is \( \frac{1}{2}Ma^2 \) (see Table 1.1). By symmetry its moments of inertia about \( OX, OY \) are equal, i.e.
\[ I_x = I_y = I \text{ (say)}. \]
Thus, by the perpendicular axes theorem
\[ I_z = I_x + I_y = 2I. \]
\[ \therefore \frac{1}{2}Ma^2 = 2I, \]
\[ \therefore I = \frac{1}{4}Ma^2. \]
That is, the moment of inertia of the disc about a diameter is \( \frac{1}{4}Ma^2 \).

1.7 Kinetic Energy of a Rolling Body

Consider a cylinder of radius \( a \) and mass \( M \) rolling, without slipping, along a horizontal surface (Fig. 1.5). Suppose that the linear velocity of the centre of gravity is \( v \). The motion of the cylinder may be regarded as a combination of two velocities: the linear velocity \( v \) of the centre of gravity plus a rotation about the centre of gravity of angular velocity \( \omega \). The instantaneous velocity of the part of the cylinder in contact with the horizontal surface must be zero so that \( v \) and \( \omega \) must be related by \( v = aw \).

The total kinetic energy of the cylinder is then given by:
\[
\text{K.E.} = \frac{1}{2}Mv^2 + \frac{1}{2}I_\omega \omega^2
\]
\[ = \frac{1}{2}Mv^2 + \frac{1}{2} \cdot \frac{1}{4}Ma^2 \cdot \left( \frac{v}{a} \right)^2 \]
\[ = \frac{3}{4}Mv^2. \]
ANGULAR MOTION AND GRAVITATION

Alternatively, the cylinder may be considered to rotate, instantaneously, about the line of contact with the horizontal surface with angular velocity \( \omega \) \((= v/a)\). The moment of inertia of the cylinder about this axis \((I_O)\) may be found by use of the parallel axes theorem:

\[
I_O = I_G + Ma^2
= \frac{1}{2}Ma^2 + Ma^2
= \frac{3}{2}Ma^2.
\]

Hence,

\[
\text{Kinetic Energy} = \frac{1}{2}I_O\omega^2
= \frac{1}{2} \cdot \frac{3}{2}Ma^2 \cdot \left(\frac{v}{a}\right)^2
= \frac{3}{4}Mv^2.
\]

**Example.** A sphere of mass \(M\) rolls, without slipping, down a plane inclined at angle \(\theta\) to the horizontal. What is its acceleration?

![Fig. 1.6](image)

Suppose that the sphere has radius \(a\) and after travelling a distance \(x\) from rest it has a velocity \(v\) (Fig. 1.6). Its centre has fallen a distance \(x\sin\theta\) so that the loss in potential energy is \(Mgx \sin\theta\).

The gain in kinetic energy is \(\frac{1}{2}Mv^2 + \frac{1}{2}I_G\omega^2\), where \(\omega = v/a\) and \(I_G = \frac{2}{5}Ma^2\) (see Table 1.1). Hence equating gain in kinetic energy to loss in potential energy:

\[
\frac{1}{2}Mv^2 + \frac{1}{2} \cdot \frac{2}{5}Ma^2 \cdot \left(\frac{v}{a}\right)^2 = Mgx \sin\theta,
\]

\[
\therefore \quad Mv^2\left(\frac{1}{2} + \frac{1}{5}\right) = Mgx \sin\theta,
\]

\[
\therefore \quad v^2 = \frac{10}{7}gx \sin\theta.
\]

But, if \(f\) is the acceleration then, \(v^2 = 2fx\),

i.e.

\[
2fx = \frac{10}{7}gx \sin\theta,
\]

\[
\therefore \quad f = \frac{5}{7}g \sin\theta.
\]

The acceleration of the rolling sphere is \(\frac{5}{7}g \sin\theta\).

Note that if the sphere had slipped down a smooth plane without rolling the acceleration would have been \(g \sin\theta\).
1.8 Angular Momentum

The body in Fig. 1.7 is rotating with angular velocity \( \omega \) about an axis through \( O \) perpendicular to the plane of the paper. A small part of the body of mass \( m \) is at \( A \), a perpendicular distance \( r \) from the axis through \( O \). This mass has tangential velocity \( v = r\omega \), and hence has linear momentum \( mv = mr\omega \). The moment of this momentum about \( O \) is \( mr^2\omega \), i.e. \( mr^2\omega \). This moment of momentum is also called the angular momentum, and summing for the whole body gives:

\[
\text{Angular momentum of body} = \Sigma mr^2\omega, \quad = I\omega,
\]

where \( I \) is the moment of inertia of the body about the axis of rotation.

1.9 Equations for Angular Motion

(a) Angular counterpart of \( F = ma \)

Suppose the rigid body in Fig. 1.8 is rotating about an axis through \( O \) with an angular acceleration \( \ddot{\theta} \). The elementary mass \( m \) at \( A \), which is a perpendicular distance \( r \) from the axis, has a linear acceleration \( a = r\ddot{\theta} \), and by Newton’s Second Law of Motion it must therefore be acted on by a force

\[
F = ma = mr\ddot{\theta}.
\]

This force has moment \( Fr \) about the axis. The total moment (or torque) \( T \) of the external forces about the axis is thus given by

\[
T = \Sigma Fr = \Sigma mr\ddot{\theta} \cdot r = \Sigma mr^2 \cdot \ddot{\theta}
\]

\[
\therefore \quad T = I\ddot{\theta} \quad . \quad (1.2)
\]

where \( I \) is the moment of inertia of the body about the axis.

In words, applied torque = (moment of inertia) \( \times \) (angular acceleration).

(b) Impulse

Suppose that a rigid body is free to rotate about a certain axis and that it is acted on by a torque \( T \) (not necessarily constant) about this axis. At any instant \( T \) is related to the angular acceleration by equation (1.2). Multiplying both sides of this equation by \( \delta t \) gives

\[
T \delta t = I\ddot{\theta} \delta t,
\]
and if the torque acts for time \( t \), then

\[
\int_{0}^{t} T \, dt = \int_{0}^{t} I \dot{\theta} \, dt = I \dot{\theta} |_{0}^{t}
\]

\[
\Rightarrow \int_{0}^{t} T \, dt = I \omega - I \omega_{0}
\]

(1.3)

where \( \omega \) and \( \omega_{0} \) are the angular velocities at time \( t \) and zero time respectively. The left-hand side of equation (1.3) is the moment of the impulse produced by \( T \), and the right-hand side is the change in angular momentum, i.e.

\[\text{moment of impulse} = \text{change in angular momentum}.\]

(c) Conservation of angular momentum

If no external torque acts on a rotating body or system of interacting bodies about a fixed axis, then the left-hand side of equation (1.3) is zero, so that there is no change in angular momentum about that axis. This is the principle of conservation of angular momentum which may be stated:

In any system of interacting bodies the total angular momentum about a fixed axis remains constant provided no external torque acts on the system about that axis.

It follows from the principle that if the moment of inertia of a rotating body is made to decrease, then its angular velocity must increase. Use of this fact is made in acrobatics and diving, the performer ‘curling up’ to increase his speed of rotation during a somersault or dive, and by the figure skater when pirouetting.

(d) Work done by a couple

Suppose that a disc of radius \( a \) free to turn about an axle through its centre is acted on by a force \( F \) applied tangentially (Fig. 1.9). There is an equal and opposite force exerted by the axle on the disc at the centre and these two forces constitute a couple of moment \( T = Fa \).

If the disc turns through an angle \( \theta \) radians and \( F \) changes direction so that it always acts tangentially to the edge of the disc, then the point of application of \( F \) has moved through a distance \( a\theta \); the work done is \( F \cdot a\theta \), which is \( T\theta \).

This result has been derived for a particular case, but it is true in general that

\[\text{work done by a couple} = \text{(moment of couple)} \times \text{(angle in radians turned through by couple)}.\]
(c) Conservation of energy

If a couple of moment \( T \) is applied to a rotating body, then its angular velocity will change from, say, \( \omega_0 \) to \( \omega \), while it turns through an angle \( \theta \). Then,

work done by a couple \( = T\theta \)

kinetic energy gained by body \( = \frac{1}{2}I\omega^2 - \frac{1}{2}I\omega_0^2 \),

where \( I \) is the moment of inertia of the body about the axis of rotation. By the principle of conservation of energy,

\[ T\theta = \frac{1}{2}I\omega^2 - \frac{1}{2}I\omega_0^2. \]

Table 1.2 shows a comparison between the formulae for linear and angular motion. It will be noticed that for angular motion mass is replaced by moment of inertia, force by torque; and linear displacement, velocity and acceleration by their angular counterparts.

<table>
<thead>
<tr>
<th>Second law of motion</th>
<th>Linear motion</th>
<th>Angular motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impulse</td>
<td>( F = m\ddot{x} )</td>
<td>( T = I\dot{\theta} )</td>
</tr>
<tr>
<td>Conservation of momentum</td>
<td>( \int F , dt = mv - mu )</td>
<td>( \int T , dt = I\omega - I\omega_0 )</td>
</tr>
<tr>
<td>Work</td>
<td>( m_1v_1 + m_2v_2 = m_1v_1 + m_2w_2 )</td>
<td>( I_1\omega_1 + I_2\omega_2 = I_1\omega_1 + I_2\omega_2 )</td>
</tr>
<tr>
<td>Conservation of energy</td>
<td>( F_s = \frac{1}{2}mv^2 - \frac{1}{2}mu^2 )</td>
<td>( T\theta = \frac{1}{2}I\omega^2 - \frac{1}{2}I\omega_0^2 )</td>
</tr>
</tbody>
</table>

Table 1.2

1.10 Angular Simple Harmonic Motion. Oscillation of a Rigid Body suspended by a Torsion Wire

Fig. 1.10 shows a rigid body suspended by a wire, its centre of gravity being vertically below the point of suspension. Its moment of inertia about the vertical axis through its centre of gravity is \( I \). The torque required to produce one radian twist in the wire is denoted by \( c \). When the body has rotated through an angle \( \theta \) from its equilibrium position the restoring torque is \( c\theta \), and the equation of motion (equation (1.2)) is

\[ -c\theta = I\dot{\theta}, \]

\[ \therefore \dot{\theta} = -\frac{c}{I}\theta. \]

The motion is thus simple harmonic and its period is

\[ T = 2\pi \left( \frac{I}{c} \right)^{\frac{1}{2}} \quad . \quad . \quad . \quad (1.4) \]
1.11 Oscillation of a Rigid Body about a Horizontal Axis

Fig. 1.11 shows a rigid body of mass $M$ pivoted at $O$, its moment of inertia about this axis being $I$. $G$ is the centre of gravity of the body and is distance $h$ from $O$. The body has been displaced through an angle $\theta$ from its equilibrium position, which is with $G$ vertically below $O$.

The moment about $O$ of the restoring force acting on the body is $Mgh \sin \theta$, and so equation (1.2) becomes

$$-Mgh \sin \theta = I\ddot{\theta}.$$  
If $\theta$ is small, then $\sin \theta$ is approximately equal to $\theta$ and

$$Mgh\theta = -I\ddot{\theta},$$  
$$\therefore \dot{\theta} = -\frac{Mgh}{I}\theta,$$

which is the equation of simple harmonic motion, with period

$$T = 2\pi \left(\frac{I}{Mgh}\right)^{\frac{1}{2}}.$$  

If $I_G$ denotes the moment of inertia of the body about a horizontal axis through $G$ parallel to that through $O$, then by the parallel axes theorem

$$I = I_G + Mh^2$$  
$$= Mk^2 + Mh^2,$$

where $k$ is the radius of gyration of the body about the axis through $G$.

Thus

$$T = 2\pi \left(\frac{Mk^2 + Mh^2}{Mgh}\right)^{\frac{1}{2}}$$  
$$\therefore T = 2\pi \left(\frac{k^2 + h^2}{gh}\right)^{\frac{1}{2}} \quad \cdots \quad (1.5)$$

In principle equation (1.5) could be used as the basis of an experimental determination of $g$, but as the quantity $k$ is difficult to determine accurately it is preferable to use a variation which does not require its value to be known. Two such methods are discussed in the following sections.

1.12 Compound Pendulum

A compound pendulum, as used to determine $g$, consists of a wooden or metal bar with several holes drilled through it at various points along its axis. It may be suspended from a knife edge through any of these holes.

Its period of oscillation is given by equation (1.5), in which $h$ can be
varied by changing the axis of oscillation. If the period \( T \) is determined for each of a series of values of \( h \), and the results plotted, a curve of the shape shown in Fig. 1.12 is obtained. This curve is a plot of equation

\[ T^2 = \frac{4\pi^2}{g} \left( \frac{k^2}{h} + h \right), \]

and differentiating with respect to \( h \),

\[ 2T \frac{dT}{dh} = \frac{4\pi^2}{g} \left( \frac{k^2}{h^3} + 1 \right). \]

For a minimum value of \( T \),

\[ \frac{dT}{dh} = 0, \]

\[ \therefore \frac{4\pi^2}{g} \left( \frac{k^2}{h^3} + 1 \right) = 0 \]

\[ \therefore h^2 = k^2. \]

Thus the minimum value of \( T \) occurs when \( h = k \) and is

\[ T_{\text{min}} = 2\pi \left( \frac{k^2 + k^2}{gk} \right)^{\frac{1}{2}} \]

\[ = 2\pi \left( \frac{2k}{g} \right)^{\frac{1}{2}}. \]

\( k \) can be determined by reading off the value of \( h \) corresponding to \( T_{\text{min}} \) on the graph. \( g \) could then be calculated from these results, but as the
position of the minimum is difficult to determine accurately, the following method is used. It will be noticed that for any value of $T$, greater than $T_{\text{min}}$, there are two values of $h$ which will give this period, e.g. the points A and B. Denote these two values by $h_1$ and $h_2$. Then

(i) \[ T^2 = \frac{4\pi^2}{g} \left( \frac{k^2 + h_1^2}{h_1} \right) \]

and (ii) \[ T^2 = \frac{4\pi^2}{g} \left( \frac{k^2 + h_2^2}{h_2} \right) \]

\[ \therefore \frac{k^2}{h_1} + h_1 = \frac{k^2}{h_2} + h_2, \]

\[ \therefore k^2 \left( \frac{h_2 - h_1}{h_1 h_2} \right) = h_2 - h_1, \]

\[ \therefore k^2 = h_1 h_2, \]

and substituting in (ii),

\[ T^2 = \frac{4\pi^2}{g} (h_1 + h_2) \quad \ldots \quad (1.6) \]

If $T$, $h_1$ and $h_2$ are read off from the graph (Fig. 1.12), then $g$ may be calculated.

An alternative way of deriving equation (1.6) uses the concept of the equivalent simple pendulum. This is the simple pendulum which has the same period of oscillation as the compound pendulum oscillating about a given axis.

If the length of the equivalent simple pendulum is $l$, then

\[ T = 2\pi \left( \frac{k^2 + h^2}{hg} \right)^{\frac{1}{2}} = 2\pi (l/g)^{\frac{1}{2}}, \]

so that

\[ l = \frac{k^2 + h^2}{h}, \]

which may be rewritten $h^2 - lh + k^2 = 0$.  

![Fig. 1.13](image)
This is a quadratic in $h$ which has the two roots $h_1$, $h_2$. By the theory of quadratic equations the sum of the roots is given by $-(\text{coefficient of } h)$

i.e.

\[ h_1 + h_2 = l, \]

\[ \therefore T = 2\pi \left( \frac{h_1 + h_2}{g} \right)^{\frac{1}{2}} \]

which is the same as equation (1.6).

When performing the compound pendulum experiment it is usual to call values of $h$ on one side of $G$ positive, and those on the other side negative. The graph obtained is shown in Fig. 1.13. The line ABCD is drawn for some value of $T$, and $h_1 + h_2$ is read off either as AC or BD (or better still the mean of these). Several values for $g$ may be calculated from the same graph by choosing different values of $T$.

### 1.13 The Reversible Pendulum. Kater’s Pendulum

For a compound pendulum there are four axes about which the periods of oscillation are the same (e.g. A, B, C, D in Fig. 1.13). Two of these axes (A and B) are on one side of the centre of gravity and two (C and D) are on the other side. The reversible pendulum is used to locate two of these axes, one on each side of the centre of gravity but asymmetrically placed with respect to it (i.e. A and C or B and D).

A typical reversible pendulum is illustrated in Fig. 1.14. $K_1$ and $K_2$ are two knife edges about which the pendulum may be oscillated. Two cylinders of the same size but made of different materials (wood and brass) are used to load the pendulum so that its centre of gravity ($G$) is much nearer to one knife edge than the other. The position of the weight $E$ is adjusted until the periods of oscillation about $K_1$ and $K_2$ are the same. Then $K_1G = h_1$, $K_2G = h_2$. Thus $K_1K_2 = h_1 + h_2$ and $g$ may be calculated by the application of equation (1.6). The distance $K_1K_2$ may be measured accurately.

The asymmetrical loading ensures that $K_1$, $K_2$ correspond to the required axes and not to the symmetrically placed ones (A and D or B and C in Fig. 1.13).

In practice, it takes a very long time to adjust the periods to equality, but this can be avoided by the following treatment, due to Bessel. Suppose the periods about $K_1$ and $K_2$ are $T_1$ and $T_2$ respectively and these are nearly equal. Denoting $K_1G$ by $h_1$ and $K_2G$ by $h_2$, we have (see equation (1.5)),

\[ T_1^2 = \frac{4\pi^2}{g} \left( \frac{k^2 + h_1^2}{h_1} \right), \]
and
\[ T_2^2 = \frac{4\pi^2}{g} \left( \frac{k^2 + h_2^2}{h_2} \right), \]

so that
\[ T_1^2 h_1 - T_2^2 h_2 = \frac{4\pi^2}{g} (h_1^2 - h_2^2), \]

\[ \therefore \ \frac{4\pi^2}{g} = \frac{T_1^2 h_1 - T_2^2 h_2}{h_1^2 - h_2^2} = \frac{1}{2} \left( \frac{T_1^2}{h_1 + h_2} + \frac{T_2^2}{h_1 - h_2} \right). \]

The first term in the brackets may be determined accurately as \( h_1 + h_2 = K_1 K_2 \) as before. The second term cannot be determined so accurately, because the quantity \( h_1 - h_2 \) is involved, and this necessitates a determination of the position of the centre of gravity; this cannot be done very precisely. However, if \( T_1 \) and \( T_2 \) are nearly equal the numerator is small, so that the second term is much less than the first, and need not be determined as accurately. Thus the allowable difference between \( T_1 \) and \( T_2 \) depends on the accuracy with which \( h_1 - h_2 \) can be determined. Say, for example, that the accuracy of \( h_1 - h_2 \) is only \( \frac{1}{10} \) of the accuracy of \( h_1 + h_2 \); then \( T_1, T_2 \) must be such that the second term is not greater than \( \frac{1}{100} \) of the first.

For the practical details of the experiment the reader is referred to a manual of practical physics and, for the timing methods and corrections applied in very accurate experiments, to Newman and Searle (see bibliography).

**GRAVITATION**

### 1.14 Newton's Law of Gravitation

This law, which was derived from observations on the motions of the planets, states:

'Two point masses attract each other with a force which is proportional to the product of the masses and inversely proportional to the square of their distance apart.'

In symbols
\[ F \propto \frac{M_1 M_2}{r^2}, \]
i.e.
\[ F = G \cdot \frac{M_1 M_2}{r^2}, \]
where \( G \) is a universal constant known as the gravitational constant. It has the dimensions \( M^{-1} L^3 T^{-2} \).

### 1.15 Gravitational Field and Potential

The concepts of gravitational field and potential are analogous to those of electric field and electric potential respectively.

A region in which a mass experiences a gravitational force is said to be a gravitational field. The force experienced by unit mass when placed at some
point in the field is called the strength of the field at that point. One important difference from the electrical case is that there is no possibility of repulsion taking place as between similar electric charges. All gravitational forces are attractions. It follows from the above definition and from Newton’s law that the field at distance \( r \) from a mass \( M \) is \( \frac{GM}{r^2} \).

The strength of a gravitational field is a vector quantity, and hence if two different bodies are producing fields at the same point the resultant must be found by the rules of vector addition. It is useful also to be able to associate a scalar quantity with any point in a gravitational field, and this leads to the concept of potential. The gravitational potential difference between two points in a gravitational field is the work which has to be done against gravitational forces to move unit mass between the two points.

This definition assumes that the work done is independent of the path taken between the two points. In the absence of forces such as friction this must be so, as the following argument shows. Suppose first that the assumption is not true, so that more work is done on a body along path ACB than along ADB (Fig. 1.15).

If a mass now travels the closed circuit ACBDA it will arrive at A with more energy than when it left, because more work is done on it along ACB than it has to do along BDA. There has been no corresponding loss in energy elsewhere, so that the law of conservation of energy has been contravened. Thus the proposition that the work done depends on the path taken must be untrue. A field in which the work done is independent of the path taken is known as a conservative field.

So far we have only defined the gravitational potential difference between two points, not the gravitational potential at a point. To make this latter definition requires us to fix some zero of potential, and for convenience this is usually taken to be at a point remote from all attracting masses (‘infinity’). It must be emphasized that this choice is purely arbitrary, but it is clearly a very convenient one.

With this choice of zero potential, the gravitational potential at a point may now be defined as the work which has to be done to bring unit mass from infinity to that point. As all gravitational forces are attractive, the field will in fact do work on the mass so that all gravitational potentials are negative.

1.16 Potential due to a Point Mass

We require to find an expression for the potential at a point \( P \) at distance \( r \) from a point mass \( M \) (Fig. 1.16). To do this we imagine a unit mass brought up from infinity to \( P \) and calculate the work done.
At some instant the unit mass is at distance $x$ from $M$, and the force on it, towards $M$, is then given by $F = \frac{GM}{x^2}$. Suppose the unit mass now moves a small distance $(-\delta x)$ towards $M$. Work $\delta W$ is done on the unit mass by the field, where

$$\delta W = F(-\delta x) = -GM \frac{\delta x}{x^2}.$$  

Thus the total work done by the field when the unit mass moves from $\infty$ to $P$ is

$$W = -\int_{\infty}^{r} \frac{GM}{x^2} \, dx = \left[ \frac{GM}{x} \right]_{\infty}^{r} = GM \frac{r}{x}.$$  

The gravitational potential $V$ at $P$ is the work done against the field and is thus given by

$$V = -GM \frac{1}{r}.$$  

(1.7)

**1.17 Relation between Field and Potential**

Consider a unit mass at a point $A$ in a gravitational field and suppose it experiences a force $F$ due to the field. Let the mass move a distance $\delta x$ in the direction of the force, to a point $B$, so that work is done on it equal to $F \delta x$. The difference in gravitational potential, $\delta V$ between $A$ and $B$, is equal to the work done against the field, hence

$$\delta V = -F \delta x,$$

and, letting $\delta x \rightarrow 0$,

$$F = -\frac{dV}{dx}.$$  

(1.8)

In words,

At a point in a gravitational field the component of the field in a given direction is equal to minus the gravitational potential gradient in the same direction.

**1.18 Uses of the Concept of Gravitational Potential**

As potential is a scalar quantity it is usually easier to calculate the total potential at a point due to a number of neighbouring masses, than it is to calculate the resultant field. Having calculated the potential, the component of the field in any direction may be found by use of equation (1.8).

This is exemplified in the following sections, which give proofs of important results which are often made use of in problems.
1.19 Gravitational Potential and Field due to a Thin Spherical Shell

The results quoted without proof in most elementary books are:

(i) The field at points outside a spherical shell is the same as that due to a particle of mass equal to that of the shell, placed at the position of the centre of the shell.

(ii) The field inside a spherical shell is zero.

These results are proved via the concept of potential rather than directly in terms of field.

It is required to calculate the potential at a point P, which is at a distance \( r \) from O, the centre of a spherical shell of radius \( a \) and wall thickness \( t \) \((t \ll a)\). Consider the shell to be divided up into narrow rings, perpendicular to OP; a typical ring is shown in Fig. 1.17. The edges of the ring subtend at angle \( \delta \theta \) at the centre of the shell and a line from P to the ring makes an angle \( \theta \) with OP. P is distance \( x \) from the ring.

The width of the ring is \( a \delta \theta \), its circumference \( 2\pi a \sin \theta \) and its thickness \( t \). Hence its mass is \( 2\pi a^2 t \rho \sin \theta \delta \theta \), where \( \rho \) is the density of the material of the shell.

The potential at P due to this ring is therefore given by

\[
\delta V = -G \cdot \frac{2\pi a^2 t \rho \sin \theta \delta \theta}{x}, \text{ from equation (1.7).}
\]

The total potential at P due to the whole shell is found by summing the contributions from all the rings as \( \delta \theta \) tends to zero. This is done by integrating the expression for \( \delta V \).

i.e.

\[
V = -G \int \frac{2\pi a^2 t \rho \sin \theta \, d\theta}{x}.
\]

As this stands it contains the two variables \( \theta \) and \( x \), so that, before integration can be performed, one of these variables must be expressed in terms of the other.

From \( \triangle OPQ \) we have the relation

\[
x^2 = a^2 + r^2 - 2ar \cos \theta.
\]

Hence, differentiating:

\[
2x \, dx = 2ar \sin \theta \, d\theta.
\]
ANGULAR MOTION AND GRAVITATION

Thus, substituting for \( \sin \theta \, d\theta \) in equation (1.9):

\[
V = -G \int \frac{2\pi a^2 \rho}{x} \cdot \frac{2x}{2ar} \, dx \\
= -G \int \frac{2\pi at \rho}{r} \, dx
\]

The range of integration depends on the position of \( P \).

(i) \( P \) outside the shell

The limits are \( x = r - a \) and \( x = r + a \)

\[
\therefore \, V = -G \int_{r-a}^{r+a} \frac{2\pi at \rho}{r} \, dx \\
= -G \frac{2\pi at \rho}{r} [(r + a) - (r - a)] \\
= -G \frac{4\pi a^2 t \rho}{r}.
\]

But the mass, \( M \), of the whole shell is given by

\[
M = 4\pi a^2 t \rho, \\
\therefore \, V = -\frac{GM}{r}.
\]

This is seen to be the same as if the whole mass of the shell were concentrated at its centre. We conclude that the potential at all points outside a shell is the same as that produced by replacing the shell by an equal point mass at its centre. As this is true for all points outside the shell, the same must be true of the potential gradient, and therefore of the field produced by the shell.

(ii) \( P \) on the surface of the shell

The range of integration (equation 1.10) is \( x = 0 \) to \( x = 2r \).

Hence

\[
V = -G \frac{2\pi at \rho}{r} [2r - 0] \\
= -G \cdot 4\pi at \rho \\
= -G \frac{M}{a}.
\]

(iii) \( P \) inside the shell

The range of integration (equation 1.10) is \( x = a - r \) to \( x = a + r \)

\[
\therefore \, V = -G \frac{2\pi at \rho}{r} [(a + r) - (a - r)] \\
= -G 4\pi at \rho \\
= -\frac{GM}{a}.
\]
This is independent of \( r \), which means that the potential is the same at all points inside the shell and has the same value as on the surface. The potential gradient, and hence the field, must therefore be zero. These results are represented graphically in Fig. 1.18.

1.20 Fields Produced by a Thick Spherical Shell and a Solid Sphere

This problem is solved by imagining the body to be divided up into a large number of thin concentric shells and then applying the results of the previous section.

(i) *Outside the shell or sphere*

The point concerned is outside all the thin shells of which the body is made up, so that each of these may be replaced by an equal mass at its centre. Thus the whole thick shell or sphere may be replaced by an equal mass at its centre for the calculation of the field outside the body.

(ii) *Completely inside the thick shell or at the centre of the solid sphere*

Here the point is inside all the thin shells, so that the field for each, and hence the total field, is zero.

(iii) *In the material of the shell or sphere*

In this case the shells fall into two groups: (a) those which enclose the point and may therefore be ignored; (b) those between the point
and the centre which may be considered as concentrated at the centre. Suppose we wish to find the field at a point P, distance b from the centre O of a solid uniform sphere of mass M and radius \( a (b < a) \) (Fig. 1.19). All the shells of radius \( > b \) produce no field and are ignored. The shells of radius \( < b \) are to be replaced by a point mass at the centre. The mass of all these shells is \( \frac{b^3}{a^3} \) (mass of sphere \( \propto (radius)^3 \)). Hence the field at P is that due to a mass \( \frac{b^3}{a^3} \) at O

\[
= G \frac{Mb^3}{a^3} / b^2,
\]

\[
= G \frac{Mb}{a^3} \quad \text{(1.11)}
\]

1.21 Escape Velocity

It has been shown that at a point on the surface of a sphere of mass M and radius R the gravitational potential is \( -G \frac{M}{R} \), and hence the potential energy of a mass \( m \) placed at this point is \( -\frac{GmM}{R} \). If the body is now given kinetic energy equal to this negative potential energy its total energy will be zero, so that the body will just be able to escape from the gravitational attraction of the sphere, i.e. it will arrive at 'infinity' with zero velocity, corresponding to its zero energy. The velocity which must be given to the body to accomplish this is called its escape velocity \( v_e \) and is found by equating kinetic energy plus potential energy to zero,

\[
\frac{1}{2}mv_e^2 - G \frac{Mm}{R} = 0
\]

\[
\therefore \quad v_e = \left( \frac{2GM}{R} \right)^{\frac{1}{2}}.
\]

Note that the escape velocity is independent of the mass of the escaping particle. For the earth:

\[
G = 6.67 \times 10^{-8} \text{ c.g.s. units},
\]

\[
M = 5.98 \times 10^{27} \text{ g}
\]

Mean radius, \( R = 6.37 \times 10^8 \text{ cm} \)

\[
\therefore \quad v_e = \left( \frac{2 \times 6.67 \times 10^{-8} \times 5.98 \times 10^{27}}{6.37 \times 10^8} \right)^{\frac{1}{2}}
\]

\[
= 1.1 \times 10^8 \text{ cm sec}^{-1}.
\]
Kinetic theory calculations show that the root mean square velocity of oxygen molecules at 0°C is $4.6 \times 10^4$ cm sec$^{-1}$. This is only about $\frac{1}{4}$ of the escape velocity, so that there is very little chance of an oxygen molecule acquiring sufficient velocity to escape.

### 1.22 Variation of $g$ with Altitude and Depth

(a) *Altitude*

At the earth's surface the force exerted on a mass $m$ is $mg$.

Hence

$$mg = G\frac{mM}{R^2},$$

where $M$ is the mass of the earth and $R$ its radius.

$$\therefore \ g = \frac{GM}{R^2}.$$

At an altitude $h$, suppose the acceleration due to gravity is $g'$. Then

$$mg' = G\frac{mM}{(R + h)^2}.$$

Hence

$$g' = \frac{GM}{(R + h)^2}.$$

This gives the variation of $g$ with altitude.

If $h$ is small compared with the radius of the earth, then this expression may be approximated as follows:

$$g' = \frac{GM}{R^2 \left(1 + \frac{h}{R}\right)^2}$$

$$= \frac{GM}{R^2} \left(1 - \frac{2h}{R} + \text{terms in } \frac{h^2}{R^2} \text{ and higher} \right)$$

$$= g \left(1 - \frac{2h}{R}\right) \text{ neglecting higher-order terms.}$$

(b) *Depth below the earth's surface*

For this we make use of equation (1.11). If $d$ is the depth below the earth's surface, then

$$a = R,$$

and

$$b = R - d,$$

$$\therefore \ g' = \frac{GM(R - d)}{R^3} = g \left(1 - \frac{d}{R}\right).$$

Thus $g'$ varies linearly, from $g$ at the earth's surface to zero at the centre of the earth.
1.23 Variation of $g$ with Latitude

The actual variation of $g$ with latitude is due to a combination of two factors: (i) the earth is not a true sphere; (ii) the earth rotates about its axis.

(i) At the poles the distance from the centre of the earth is less than at the equator. This results in a value of $g$ greater at the poles than at the equator.

(ii) The effect of rotation is greatest at the equator and results in a reduction in the value of $g$; at the poles rotation has no effect. Thus, rotation also results in a higher value of $g$ at the poles than at the equator.

The treatment of the effect of rotation at the equator is fairly simple, as all the forces involved are acting along the same straight line.

Consider a body of mass $m$ hanging from a plumb line at the equator (Fig. 1.20).

The forces acting on the body are the tension, $T$, in the string and the force of gravity $mg$ directed towards the centre of the earth, where $g$ denotes the acceleration due to gravity in the absence of rotation. Owing to the rotation of the earth, the body is moving with angular velocity $\omega$ in a circle of radius $R$, where $\omega$ and $R$ are respectively the angular velocity and radius of the earth. The acceleration of the body is thus $\omega^2 R$ towards the centre of the earth.

Thus

$$mg - T = m\omega^2 R$$

$$\therefore T = m(g - \omega^2 R).$$

But if $g'$ represents the apparent value of the acceleration due to gravity at the point, then

$$T = mg',$$

$$\therefore g' = g - \omega^2 R.$$

The equatorial radius of the earth is $6.38 \times 10^8$ cm, so that the value of $\omega^2 R$ is

$$\left(\frac{2\pi}{24 \times 3600}\right)^2 \times 6.38 \times 10^8$$

$$= 3.37 \text{ cm sec}^{-2}.$$

It has already been mentioned that the treatment at other latitudes is complicated by the non-spherical shape of the earth; we shall ignore this and consider the effect of rotation only.

Figs. 1.21 shows a mass $m$ hanging from a plumb line at latitude $\lambda$.

The mass $m$ is rotating in a circle of radius $R \cos \lambda$ about the point $A$ so that the difference between $T$ and $mg$ must now be directed towards
A to provide the necessary force to keep the mass moving in its circle. This means that $T$ and $mg$ cannot act along the same straight line, i.e. the plumb line does not point towards the centre of the earth, but makes an angle $\theta$ with the vertical as shown.

Resolving parallel to PA gives the resultant force towards A as $mg \cos \lambda - T \cos (\lambda + \theta)$. This is the force which keeps the mass $m$ moving in a circle of radius $R \cos \lambda$, i.e. it produces an acceleration $\omega^2 R \cos \lambda$.

Hence

$$mg \cos \lambda - T \cos (\lambda + \theta) = m\omega^2 R \cos \lambda.$$  

In the direction perpendicular to PA in the plane of the paper there must be no resultant force.

Hence

$$mg \sin \lambda - T \sin (\lambda + \theta) = 0.$$  

Squaring and adding to eliminate $(\lambda + \theta)$ gives

$$T^2 = (mg \cos \lambda - m\omega^2 R \cos \lambda)^2 + (mg \sin \lambda)^2$$

$$= m^2 g^2 - 2m^2 g \omega^2 R \cos \lambda + m^2 \omega^4 R^2 \cos^2 \lambda.$$  

But

$$T^2 = m^2 g^2,$$

$$\therefore g' = g^2 - 2g \omega^2 R \cos \lambda + \omega^4 R^2 \cos^2 \lambda.$$  

**QUESTIONS**

1. A flywheel of radius $a$ is mounted on an axle of radius $r$, and has mass $M$. Two parallel rails are arranged at an angle $\theta$ to the horizontal so that the axle may rest on the rails with the flywheel clear of the ground. Show that, if there is no slipping, the linear acceleration when the flywheel is released is $\frac{2r^2 g \sin \theta}{a^2 + 2r^2}$. (The moment of inertia of the axle may be neglected.)

2. A flywheel is mounted on an axle of radius $r$ which runs in bearings so that the flywheel can rotate about a horizontal axis. A string wound round the axle carries a mass $M$, which is released from rest. It is found that after time $t$ the mass has fallen a distance $h$. Show that, if friction can be neglected, the moment of inertia, $I$, of the flywheel and axle is given by

$$I = Mr^2(gt^2/2h - 1).$$

If the string comes off the axle after $M$ has fallen the distance $h$ and the flywheel then makes a further $n$ revolutions before coming to rest, show that a correction for friction leads to the equation

$$I = \frac{Mr^2}{1 + h/(2\pi n)} \left( \frac{gt^2}{2h} - 1 \right).$$

Suggest an experiment to determine $I$ for a flywheel, based on the above.
3. A body is suspended from a torsion wire and caused to make rotational oscillations. The moment of inertia of the body about the axis of rotation is \( I_1 \), and the period of the oscillations is \( T_1 \). A second body is attached to the first so that its centre of gravity is vertically below the point of suspension. Show that, if the moment of inertia of the second body about the axis of rotation is \( I_2 \), the period \( T_2 \) of oscillation of the combination is given by

\[
\frac{T_2^2}{T_1^2} = \frac{I_1 + I_2}{I_1}.
\]

If the second body is of a regular shape (so that \( I_2 \) may be calculated), suggest experiments to determine (i) \( I_1 \), (ii) the torque needed to produce one radian twist in the wire.

4. A uniform disc of radius \( a \) has a small hole drilled through it at a point distance \( b \) from the centre. Show that the period of small oscillation about a horizontal axis through this hole is

\[
2\pi \left( a^2 + 2b^2 \right)^{\frac{1}{2}}.
\]

Prove any formulae you use.

5. You are given a metal tube of external diameter \( a \) whose ends are covered by thin plates which may not be removed. Show that it is possible to determine the internal diameter, \( b \), of the tube by rolling it from rest down a rough inclined plane and timing its descent, and that the result may be calculated from the formula

\[
b = a \left( 3 - \frac{gt^2 \sin \theta}{x} \right)^{\frac{1}{2}},
\]

where \( t \) is the time it takes to roll down a slope of length \( x \) inclined at angle \( \theta \) to the horizontal.

Estimate the smallest value of \( b/a \) which you would detect by this means, using ordinary laboratory equipment.

6. A turntable of moment of inertia \( I_1 \) about its centre is rotating freely with angular velocity \( \omega \). A disc of moment of inertia \( I_2 \) about its centre drops centrally on to the turntable. Show that the angular velocity of the turntable is then \( \frac{I_1 \omega}{I_1 + I_2} \) and that the loss of energy is \( \frac{1}{2} \frac{I_1 I_2}{I_1 + I_2} \omega^2 \). How do you account for this loss of energy?

7. Kepler’s third law states that the square of the period of a planet in its orbit is proportional to the cube of the semi-major axis of the orbit. Show, for the particular case of a circular orbit, that this leads to an inverse square law of force.

8. Compare the predicted variations of \( g \) with latitude with those found in practice (see Kaye and Laby’s Tables) and attempt to account for any discrepancies.

**ESSAY TOPICS**

9. Experimental determinations of moments of inertia.
10. Experimental determinations of \( G \).
11. Experimental determinations of \( g \).
SURFACE TENSION AND VISCOSITY

SURFACE TENSION

2.1 Introduction

It is assumed that the reader is already familiar with the phenomena which lead to the concept of surface tension and with their explanation in terms of the attractive forces between molecules. He should also have a knowledge of methods of measuring surface tension and should understand the theory on which the experiments are based. The following sections extend these ideas and show how they may be applied to a wider variety of problems.

2.2 Definition

Consider a line of unit length drawn in the surface of a liquid (Fig. 2.1). There are forces acting on either side of this line at right angles to it and parallel to the surface of the liquid. The magnitude of the force on one side of the line is called the surface tension of the liquid. It is measured in dyn cm⁻¹.

![Fig. 2.1](image)

2.3 Surface Energy

In Fig. 2.2 ABCD is a wire frame which has a soap film of surface tension $\gamma$ formed in it. The side BC (of length $l$) is movable. The film exerts a force $F$ on BC given by $F = 2\gamma l$. (The '2' occurs because the film has two surfaces.)

Suppose now that BC is pulled outwards a distance $x$ to B'C'. This requires work to be supplied, equal to $Fx$, i.e. $2\gamma lx$. But $2lx$ is the increase in surface area of the film (two surfaces), so that the work done per unit area of new surface formed is $\gamma$. If $\gamma$ is in dyn cm⁻¹ then the work done will be in ergs and the work to form unit surface area will be in erg cm⁻². Note that dyn cm⁻¹ and erg cm⁻² have the same dimensions: $MT⁻²$.

The work which has to be done to form unit surface area of a film under isothermal conditions is called the free surface energy.
SURFACE TENSION AND VISCOSITY

If we assume that the only source of energy in the above is the work done in moving BC, then we can say that the total surface energy is also \( \gamma \). However, the assumption that no other supply of energy is involved is not a tenable one. If an attempt is made to satisfy this assumption by preventing any exchange of heat between the film and its surroundings it is found that the temperature of the film falls. Now it is a well-known fact that surface tension varies with temperature, so that we can no longer take \( \gamma \) as a constant.

On the other hand, if the surface is formed isothermally there is a flow of heat into the film so that its surface energy is greater than \( \gamma \) erg cm\(^{-2}\).

Further discussion of this topic involves the use of thermodynamics and is postponed to Chapter 7 (see Section 7.19).

The concept of surface energy is useful in explaining the shape of drops of liquids. In the absence of forces such as gravitational ones, a liquid assumes that shape which gives it the minimum surface energy. It thus forms into that shape which has the minimum surface area for a given volume: i.e. a sphere. Thus drops of any size falling freely under gravity are spherical.

2.4 Contact between Liquid and Solid. Angle of Contact

The shape of the surface of a liquid in contact with a solid depends on the relative magnitudes of: (i) the mutual attractions between molecules of the liquid (force of cohesion); and (ii) the attractions between molecules of the liquid and those of the solid (force of adhesion).

Consider a liquid molecule P in the surface of the liquid close to a vertical wall (Fig. 2.3). This molecule is subject to three forces:

(i) the force of gravity acting vertically downwards along PG;

(ii) the force due to the attraction of molecules in the solid which, by symmetry, must act normal to the surface of the solid, along PS;

(iii) the force of attraction of neighbouring liquid molecules acting along PL.

The resultant force on P is along PR which must be between PL and PS in direction but is not, in general, along PG. The surface of the liquid at P must be such that this resultant force on P is normal to the surface; otherwise P would move along the surface. It follows that the surface of the liquid close to the wall is not generally horizontal (it will be so only if the resultant force on P is vertical).

For a molecule at P' in the surface of the liquid, not close to the wall,
the force of attraction of the molecules in the wall is very small and the resultant force of attraction of the molecules of the liquid is almost vertical. The resultant force on $P'$ is thus almost vertical and the surface practically horizontal.

![Diagram 2.4](image1.png)  
Fig. 2.4

![Diagram 2.5](image2.png)  
Fig. 2.5

The angle between the solid surface and the tangent to the liquid surface at $P$ is called the angle of contact, $\alpha$ (Fig. 2.4). If the resultant force along PL is large enough compared with that along PS the curvature of the surface is the other way and $\alpha$ is greater than 90° (Fig. 2.5).

### 2.5 Pressure Difference across a Surface of any Shape

The reader will already be familiar with the fact that to maintain a curved film of liquid requires that the pressure be greater on the concave side than on the convex side of the surface. For a spherical surface this excess pressure is $2\gamma/r$ for a single surface (e.g. air bubble in water), and $4\gamma/r$ for a double surface (e.g. soap bubble in air), where $r$ is the radius of the bubble, and $\gamma$ the surface tension of the liquid.

For a non-spherical surface the formulae must be modified to take into account that the curvature of the film is different for different sections cut through any point in the surface (Fig. 2.6).

The pressure difference across the surface may be expressed in terms of the radii of curvature of the two principal sections of the surface at any point. The principal sections are the two sections which have the maximum and minimum curvatures of all sections normal to the surface cut at that point. If these have radii $r_1$ and $r_2$, then the excess pressure $p$ is given by:

$$p = \gamma \left( \frac{1}{r_1} + \frac{1}{r_2} \right)$$

for a single-sided surface,

or

$$p = 2\gamma \left( \frac{1}{r_1} + \frac{1}{r_2} \right)$$

for a double-sided surface.

For a convex section $r$ is negative.
SURFACE TENSION AND VISCOSITY

No proof of these formulae will be attempted here; the interested reader should consult a more advanced text (e.g. Newman and Searle — see bibliography). We may, however, note that these expressions reduce to those for a spherical surface on putting \( r_1 = r_2 = r \).

2.6 Applications of the Formula for Pressure Difference across a Curved Surface

(a) Cylindrical surface

For a cylindrical surface \( r_1 = r \), and \( r_2 = \infty \), where \( r \) is the radius of the cylinder.

Thus \( p = \gamma \left( \frac{1}{r} + \frac{1}{\infty} \right) = \frac{\gamma}{r} \) for a single-sided surface.

(b) Force required to separate two wet plates

If two circular glass plates each of radius \( R \) are wetted and then placed in contact a film of water of thickness \( d \) forms between them. The edge of this film forms an approximately semi-circular meniscus, so that one principal section has radius \( d/2 \). The other principal section has radius \( R \). The pressure in the water thus exceeds atmospheric by an amount

\[
p = \gamma \left( \frac{1}{R} + \frac{1}{d/2} \right) = \frac{-2\gamma}{d}, \text{ if } R \gg d.
\]

The pressure in the water is less than atmospheric by an amount \( 2\gamma/d \), and each plate therefore experiences a force \( 2\gamma A/d \) pulling it towards the other plate. For very thin water layers this force can be very large.

2.7 Vapour Pressure above a Curved Surface

If a liquid evaporates into a closed space from a flat surface there comes a stage when the space above the liquid can contain no more vapour under the given conditions. The vapour is then said to be saturated, and the pressure exerted by this vapour is known as the saturated vapour pressure at the flat surface (s.v.p.).

Often, however, one is concerned with evaporation and condensation not at flat surfaces but at the curved surfaces of water droplets. It is necessary to investigate, then, whether the s.v.p. is the same in both cases.

Consider a closed vessel containing a liquid and its vapour only, and imagine a tube inserted in the liquid. If the angle of contact of the liquid and the tube is greater than 90°, then a convex surface of radius \( r \) forms in the tube and the liquid is depressed a distance \( h \) below the rest of the surface (Fig. 2.8).

Suppose that the s.v.p. at the flat surface of the liquid is \( p \), that the
densities of the liquid and vapour are $\rho$ and $\sigma$ respectively, and that the surface tension of the liquid is $\gamma$.

We require to find an expression for the vapour pressure just above the curved surface in the tube. This may be obtained as follows.

Pressure at $A = \text{pressure at } A' = p$.
Pressure at $B, p' = \text{pressure at } A + \text{pressure due to column of vapour of height } h$,

\[
(i) \quad p' = p + h\sigma g
\]

Pressure at $C = \text{pressure at } A' + \text{pressure due to column of liquid of height } h$,

\[
= p + h\rho g.
\]

Pressure at $C = \text{pressure at } C'$.

Pressure at $B, p' = \text{pressure at } C + \text{pressure difference across the curved surface},$

\[
(ii) \quad p' = p + h\rho g - \frac{2\gamma}{r}
\]

(the last term is negative because the surface is crossed from the concave side). $h$ may be eliminated between equations (i) and (ii) to give

\[
p' - p = \frac{2\gamma}{r} \left( \frac{\sigma}{\rho - \sigma} \right)
\]

(2.1)

The vapour pressure above a convex surface thus exceeds that above a flat surface by an amount $\frac{2\gamma}{r} \left( \frac{\sigma}{\rho - \sigma} \right)$. Table 2.1 shows the magnitude of this effect for drops of water of various diameters at 100°C. The first three values in the table have been calculated from equation (2.1) using the data at 100°C: $\gamma = 59 \text{ dyn cm}^{-1}$, $\sigma = 0.61 \times 10^{-3} \text{ g cm}^{-3}$, $\rho = 0.96 \text{ g cm}^{-3}$. The last value in the table has been calculated from the formula given at the end of this section.

<table>
<thead>
<tr>
<th>Diameter of drop (cm)</th>
<th>$p' - p$ (atmospheres)</th>
<th>$p'$ (atmospheres)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-3}$</td>
<td>0.00015</td>
<td>1.00015</td>
</tr>
<tr>
<td>$10^{-6}$</td>
<td>0.15</td>
<td>1.15</td>
</tr>
<tr>
<td>$5 \times 10^{-7}$</td>
<td>0.30</td>
<td>1.30</td>
</tr>
<tr>
<td>$10^{-7}$</td>
<td>$\ln \frac{p'}{p} = 1.43$</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Table 2.1

It is seen that the effect is appreciable only if the diameter of the drops is very small.

This treatment has assumed that the density of the vapour $\sigma$ is constant over the distance $h$, whereas in fact it varies slightly with height.
SURFACE TENSION AND VISCOSITY

This may be taken into account (see, e.g., Newman and Searle) and leads to the formula \( \ln \frac{p'}{p} = \frac{2\gamma M}{R \rho T} \), where \( M \) is the molecular weight of the vapour, \( R \) the gas constant per gram molecule and \( T \) the temperature in °K.

2.8 Effects of Surface Tension on Condensation and Evaporation of Drops

(a) Evaporation

Equation (2.1) shows that drops of a liquid cannot be in equilibrium with a flat liquid surface at the same temperature. The s.v.p. above the drop is greater than that above the flat surface. If the s.v.p. above the flat surface is slightly exceeded condensation occurs at the flat surface, but evaporation still takes place from the drop because its s.v.p. has not been exceeded. The drop evaporates and liquid condenses at the flat surface.

(b) Condensation

The vapour pressure necessary for condensation to take place on a drop increases as the radius of the drop decreases (equation (2.1) and Table 2.1). It seems then that drops cannot easily form because the vapour pressure necessary to start them off would be too great. It is a well-known fact that a very clean vapour can be cooled well below the temperature at which condensation would normally take place without any drops forming – the vapour is said to be supersaturated. The degree of supersaturation is expressed as the percentage by which the s.v.p. has been exceeded.

In a clean supersaturated vapour drops less than a certain size evaporate and those greater than this size continue to grow. From Table 2.1 if water vapour at 100°C is at a pressure of 1.15 atmospheres, drops less than \( 10^{-6} \) cm in diameter evaporate and those larger than this grow.

The mechanism of normal condensation into drops is that the vapour contains small particles (e.g. dust) which act as nuclei for the formation of drops. It will be seen from Table 2.1 that even if these particles are very small (e.g. \( 10^{-3} \) cm in diameter) the degree of supersaturation need only be very slight before condensation can begin. Once drops have been formed they continue to grow and the degree of supersaturation decreases.

2.9 Effect of Surface Tension on Boiling

When a bubble of vapour forms in a liquid the pressure inside it must exceed the pressure in the liquid by \( 2\gamma/r \). Thus to form a small bubble requires a very high vapour pressure, i.e. the temperature of the liquid needs to be well above its normal boiling point. Once a bubble has
formed it can grow rapidly because the increasing radius means that the $2\gamma/r$ term decreases. This gives rise to the phenomenon of 'bumping', in which the temperature of the liquid rises above its boiling point before any bubbles form, and then when a bubble does form it grows very quickly and may shoot the entire contents of the flask out of the neck.

For normal boiling, nuclei are required on which the bubbles may form, and in this case they must be gaseous. Pieces of porous pot are usually placed in a flask to promote steady boiling, as they can provide the necessary gaseous nuclei.

2.10 Effect of Electrically Charged Particles on Condensation

An electrically charged particle results in an effective decrease in the surface tension effect for a liquid forming on it, and thus enables drops to form on much smaller nuclei.

The reason for this is that the charges spread over the surface of the drop and due to their mutual repulsion produce an outward force on its surface, in opposition to the surface tension effect.

The Wilson cloud chamber for the detection of ionizing radiations and particles makes use of this effect. Essentially the instrument consists of a chamber containing a clean vapour which has been made supersaturated by the cooling produced in an adiabatic expansion. When the ionizing radiation or particles pass through the chamber ions are formed which then act as nuclei for the formation of water droplets. These therefore form along the path of the radiation, so producing visible tracks which can be photographed.

VISCOSITY

2.11 Introduction

It is a matter of common observation that some liquids flow more easily than others — it requires more effort to stir oil than it does to stir water. We say that oil is more viscous, or has a higher viscosity than water. Viscosity in a liquid may be thought of as an internal friction between parts of the liquid, opposing their relative motion. This friction is pictured as arising because of the forces of attraction between neighbouring molecules.

In a gas viscous effects are much smaller, but nevertheless are present and often important. The mechanism of viscosity in a gas is explained in terms of the kinetic theory (see Section 6.9 for this treatment).

2.12 Coefficient of Viscosity

Consider a wide stream of liquid flowing over a plane horizontal surface (Fig. 2.9). Provided the flow is not too fast any portion of the liquid is always moving in a straight line parallel to the bottom of the stream. It is as if there were a large number of planes or layers of liquid
all moving parallel to each other. This is known as laminar flow and for the present we confine ourselves to this type of motion.

It is found experimentally that the water in contact with the bottom of the stream is at rest and that the surface layer is moving with maximum velocity. In Fig. 2.9 the magnitudes of the velocities are indicated by the length of the arrows.

![Figure 2.9](image)

Suppose that at a point P, distance $x$ from the bottom, the velocity is $v$ and that at a point distance $\delta x$ above P the velocity is $v + \delta v$. Then the rate of change of velocity with $x$ is $\lim_{\delta x \to 0} \frac{\delta v}{\delta x}$, i.e. $\frac{dv}{dx}$. This quantity is called the velocity gradient at P.

Each layer of liquid exerts frictional forces on the adjacent layers in a direction parallel to the direction of flow. If one of these frictional forces is $F$ and acts on a layer of area $A$ it causes tangential or shearing stress of magnitude $F/A$ on the neighbouring layer.

Newton assumed that for laminar flow the velocity gradient at any point is proportional to the shearing stress at that point,

\[ \frac{dv}{dx} \propto \frac{F}{A}, \]

which may be written

\[ \frac{F}{A} = \eta \frac{dv}{dx} \quad \cdots \cdots \quad (2.2) \]

The constant $\eta$ in equation (2.2) is called the coefficient of viscosity of the liquid.

The dimensions of $\eta$ are

\[ \frac{MLT^{-2}}{L^2} \left/ \frac{LT^{-1}}{L} \right. = ML^{-1}T^{-1}. \]

The unit of $\eta$ on the c.g.s. system could be given as g cm$^{-1}$ sec$^{-1}$ but is more usually referred to as the poise.

The value of $\eta$ for a liquid decreases rapidly as the temperature rises.
2.13 Streamline and Turbulent Flow

In the previous section it has been stipulated that the flow is to be laminar. In cases where flow takes place past curved surfaces (e.g. liquid flowing along a tube), the term 'laminar flow' is not appropriate and *streamline flow* is used instead. This means that the liquid flows in 'stream-lines', i.e. any part of the liquid flows along a definite smooth path (often straight). The discussion in the previous section applies to all cases of streamline flow. In many practical examples (e.g. a swift-moving stream), the flow of the liquid deviates from streamlines to form eddies. These are regions of disturbed flow taking the form of small 'whirlpools' (Fig. 2.10).

![Streamline flow and Turbulent motion](image)

Fig. 2.10

Non-streamline flow is called *turbulent motion*, and the concepts of viscosity outlined above do not apply in this case.

2.14 Critical Velocity. Reynolds' Number

If the velocity of flow is sufficiently small, then the flow is streamline, but as the velocity increases there comes a point at which turbulent motion begins. The velocity at which this occurs is called the *critical velocity*. This cannot be given a very precise value because there is a range of velocities over which the motion is unstable and may be turbulent or streamline, but in any experiments designed to measure coefficients of viscosity it would not be wise even to approach this region of instability. One is primarily concerned then with the maximum velocity at which the flow will definitely be streamline. It is this value of the critical velocity which is normally quoted.

The way in which the critical velocity depends on various factors may be deduced by the method of dimensional analysis, but as always with this method we must first decide what these factors are.

First, it is obvious that the coefficient of viscosity \( \eta \) must affect the value of the critical velocity; a very viscous liquid is less likely to form eddies than one of low viscosity.

Secondly, if we consider the formation of eddies they involve giving to the liquid a component of momentum perpendicular to direction in which it was originally travelling, so that the mass of a portion of the fluid must be taken into account. This is done by including the density \( \rho \) as one of the factors.

Thirdly, the lateral dimension \( r \) of the space through which the liquid
SURFACE TENSION AND VISCOSITY

flows will affect the onset of turbulence – turbulent flow is less likely in a narrow tube.

If we assume that these are the only factors involved, then we may write the critical velocity $v_c$ as

$$v_c = k\eta^x \rho^y r^z,$$

where $k$ is a dimensionless constant.

The dimensions of the quantities are:

$$v: LT^{-1},$$
$$\eta: ML^{-1} T^{-1},$$
$$\rho: ML^{-3},$$
$$r: L.$$

The condition for the dimensions of the two sides of the equation to be the same is:

$$LT^{-1} = M^x L^{-x} T^{-x} M^y L^{-3y} L^z$$

Equating the powers of $T$: $-1 = -x,$

$$\therefore x = 1.$$ 

"""" M: \[ 0 = x + y, \]
$$\therefore y = -1.$$ 

"""" L: \[ 1 = -x - 3y + z, \]
$$\therefore z = -1.$$ 

$$\therefore v_c = \frac{k\eta}{\rho} \quad \quad \quad \quad \quad \quad \quad (2.3)$$

It was shown experimentally by Reynolds that for flow through a narrow tube $k$ is about 1,100.

If, in a given experiment, liquid is flowing through a tube with a maximum velocity $v$, then we can calculate the dimensionless ratio $v \rho / \eta$ for the flow. This ratio is known as Reynolds’ Number, $R$, and if it is less than 1,100, then $v < v_c$ and the flow is streamline. It is then permissible to use the results of such an experiment to calculate the coefficient of viscosity in the way indicated in the following sections.

2.15 Flow of Liquid through a Tube. Poiseuille’s Formula

Fig. 2.11 shows the flow of a liquid of coefficient of viscosity $\eta$ along a horizontal tube of circular cross-section of radius $a$; streamline flow is assumed. The lengths of the arrows represent the velocities at various points in the tube, the velocity being zero at the walls and a maximum along the axis. The liquid does not now flow in planes, but in concentric cylinders. Fig. 2.12 is an end-on view of the tube
showing rings, each drawn through points at which the flow velocity is the same.

The velocity gradient is not constant and we proceed to find an expression for the velocity at a distance \( r \) from the centre of the tube. Suppose that the tube is of length \( l \) and that the flow is maintained by a pressure difference \( \Delta p \) between the ends of the tube. Imagine a cylinder

of radius \( r \) in the liquid (Fig. 2.13). The liquid in this cylinder is not being accelerated, so that the resultant force on it must be zero. There are two forces acting on it:

(i) the force due to the pressure difference between the ends of the tube = pressure \times area = \( \Delta p \cdot \pi r^2 \);

(ii) the viscous drag, \( F \), at the curved surface of the cylinder, which is given by

\[
\frac{F}{A} = \frac{dv}{dr} = \eta \frac{dv}{dr},
\]

i.e.

\[
F = \eta \left( \frac{dv}{dr} \right) 2\pi rl.
\]

Equating the sum of these two forces to zero:

\[
\Delta p \cdot \pi r^2 + \eta \left( \frac{dv}{dr} \right) 2\pi rl = 0
\]

and integrating:

\[
\Delta p \int r \, dr = -2\eta l \int dv.
\]

The velocity \( v_r \) at distance \( r \) from the axis is found by integrating over the range of values \( r = a \) to \( a; v = v_r \) to 0, which gives

\[
\left. \frac{\Delta p}{2} \right|_{\frac{a^2}{r}}^{a} = -2\eta l \left. v \right|_{v_r}^{0}
\]

\[
\therefore \frac{\Delta p}{2} (a^2 - r^2) = 2\eta l v_r
\]

\[
\therefore v_r = \frac{\Delta p (a^2 - r^2)}{4\eta l} \quad \therefore \quad (2.4)
\]
The velocity $v_{\text{max.}}$ along the axis is found by putting $r = 0$;

$$v_{\text{max.}} = \frac{p a^2}{4\eta l} \quad \ldots \quad (2.5)$$

In assessing whether the flow will be streamline under given conditions the maximum velocity, $v_{\text{max.}}$, is substituted in the formula for Reynolds’ number;

$$R = \frac{v_{\text{max.}} a p}{\eta} \quad \ldots \quad (2.6)$$

To find the volume of liquid flowing through the tube per second we calculate first the volume flowing between two cylinders of radii $r$ and $r + \delta r$, and then integrate to find the total volume (Fig. 2.14).

The velocity of flow at a distance $r$ from the axis is $v_r = p(a^2 - r^2)/4\eta l$ (see equation (2.4)). Thus the volume per second, $\delta V$, flowing between the cylinders of radii $r$ and $r + \delta r$ is that contained in a cylindrical shell of length $v_r$ having internal and external radii $r$ and $r + \delta r$.

$$\therefore \delta V = 2\pi r \cdot \delta r \cdot v_r$$

$$= 2\pi r \cdot \frac{p(a^2 - r^2)}{4\eta l} \cdot \delta r.$$

The total volume, $V$, flowing per second is found by integrating this expression between the limits $r = 0$ and $r = a$.

$$\therefore V = \frac{\pi p}{2\eta l} \int_0^a (a^2 - r^3) \, dr$$

$$= \frac{\pi p}{2\eta l} \left[ \frac{a^2 r^2}{2} - \frac{r^4}{4} \right]_0^a$$

$$= \frac{\pi p}{2\eta l} \left( \frac{a^4}{2} - \frac{a^4}{4} \right).$$

$$\therefore V = \frac{\pi a^4 p}{8\eta l} \quad \ldots \quad (2.7)$$

Equation (2.7) is known as Poiseuille’s formula.

2.16 Application of Poiseuille’s Formula to the Determination of the Viscosity of Water

The principle of the experiment is that if water flows through a narrow tube with a constant pressure between the ends, then a measure-
ment of the volume flowing per second enables the coefficient of viscosity to be calculated by the application of equation (2.7).

A simple experimental arrangement is shown in Fig. 2.15.

Fig. 2.15

The constant-head apparatus maintains a constant pressure \( p = hg \). The volume of water flowing per second along the capillary tube of length \( l \) is determined by collecting the water for a known time and weighing it. The radius of the tube may be determined by introducing a thread of mercury into it, measuring the length of the thread, and then tipping it out and weighing it. A knowledge of the density of mercury enables its volume, and hence the area of cross-section of the tube, to be calculated.

It is very important that the tube be of uniform internal diameter, and before a tube is selected for use the uniformity of its bore should be checked by measuring the length of a short thread of mercury in various parts of the tube.

Care must also be taken to ensure that the flow is streamline, i.e. that the critical velocity is not reached. An estimate of the greatest pressure which may be used is made by using equations (2.5) and (2.6) to calculate Reynolds' number, \( R \),

\[
R = \frac{v_{\text{max}} \cdot a \cdot \rho}{\eta} \\
= \frac{pa^3 \cdot \rho}{4 \eta^2 l}
\]

Hence \( \frac{pa^3 \cdot \rho}{4 \eta^2 l} < 1,100 \) for streamline flow;

\[
\therefore p < \frac{4,400 \eta^2 l}{a^2 \rho}.
\]

If an approximate value of \( \eta \) is known, then an estimate of the maximum value of \( p \) for a given tube may be found. For example, suppose
the tube is of length 50 cm, and radius 1 mm; taking \( \eta \simeq 0.01 \) poise gives,

\[
p < \frac{4400 \cdot 10^{-4} \cdot 50}{(0.1)^3},
\]

\[
\therefore p < 22,000 \text{ dyn},
\]

\[
\therefore h \text{ must be less than } \frac{22,000}{g} \text{ cm},
\]

i.e. less than 22 cm.

It is advisable to have the pressure much less than the theoretical limit, e.g. less than 15 cm of water in the above example. As a further check the experiment may be repeated with successive increasing values of \( h \) and a graph plotted of \( V \) against \( h \). This graph is linear over the region of streamline flow, but curves when turbulence begins (Fig. 2.16).

![Graph showing rate of flow vs. pressure head](image)

Fig. 2.16

For full experimental details the student should consult a practical manual.

In accurate determinations two corrections must be applied:

(i) for the acceleration of the liquid in the inlet end of the tube;

(ii) for the kinetic energy of the issuing liquid.

For a discussion of these corrections and for a refinement of the apparatus see Champion and Davy.

### 2.17 Cylinder in a Rotating Liquid

Suppose that a hollow cylinder of internal radius \( b \) contains a liquid of viscosity \( \eta \) and that in the liquid is fixed a smaller, solid cylinder of radius \( a \), coaxial with the first (Fig. 2.17).

If the larger cylinder is rotated with constant angular velocity \( \Omega \), then the liquid also rotates and exerts a viscous drag on the inner cylinder, tending to rotate it in the same direction. The magnitude of the torque exerted on the inner cylinder is calculated as follows.

The liquid in contact with the outer cylinder is assumed to be
rotating with angular velocity $\Omega$ and that in contact with the inner to be at rest, so that there is a velocity gradient in the liquid, from which the viscous torque may be calculated. At distance $r$ from the centre let the angular velocity of the liquid be $\omega$, then the tangential velocity is $\omega r$ and the velocity gradient is the rate of change of this velocity with $r$,

$$\frac{d}{dr} (\omega r) = \omega + r \frac{d\omega}{dr}.$$

The first term in this expression still exists even if $d\omega/dr = 0$, i.e. if the angular velocity is the same at all points. This would be the case if there were no slipping, when there would be no viscous effects at all, as in the case of a rotating solid cylinder. It is obvious then that this term does not contribute to the viscous forces, so that only the second term is to be used in determining the torque.

The tangential force, $F$, on the liquid cylinder of radius $r$ is thus given by substituting

$$\frac{d\omega}{dx} = r \frac{d\omega}{dr}$$

and

$$A = 2\pi rl$$

in equation (2.2), where $l$ is the length of the inner cylinder immersed in the liquid.

Thus

$$\frac{F}{2\pi rl} = \eta r \frac{d\omega}{dr}.$$ 

This force, $F$, is acting tangentially to the surface of the cylinder of radius $r$, and thus produces a torque $T = Fr$ about the axis.

$$\therefore T = Fr = 2\pi l \eta r^3 \frac{d\omega}{dr},$$

$$\therefore \frac{T}{r^3} = 2\pi l \eta \frac{d\omega}{dr} \quad \therefore \quad (2.8)$$

When a steady state has been reached there is no angular acceleration, so that the value of $T$ must be the same at all points in the liquid, and in particular is the torque experienced by the inner cylinder.

Integrating equation (2.8) gives

$$T \int \frac{dr}{r^3} = 2\pi l \eta \int d\omega.$$
SURFACE TENSION AND VISCOSITY

When \( r = a, \omega = 0 \)
and when \( r = b, \omega = \Omega \); so that:

\[
T \left| \frac{-1}{2r^2} \right|_a^b = 2\pi \cdot b \eta \left| \omega \right|_0^\Omega,
\]

\[
\therefore \frac{T}{2 \left( \frac{1}{a^2} - \frac{1}{b^2} \right)} = 2\pi \eta \Omega
\]

\[
\therefore G = \frac{4\pi \eta \Omega}{\left( \frac{1}{a^2} - \frac{1}{b^2} \right)} = \frac{4\pi \eta a^2 b^2}{(b^2 - a^2)} \quad (2.9)
\]

If the torque on the inner cylinder is measured (e.g. by suspending it from a torsion wire), then the coefficient of viscosity of the liquid may be determined. Allowance must also be made for the torque on the bottom surface of the cylinder, and this is usually done by performing the experiment twice with different values of \( l \), but with the same distance between the bases of the two cylinders. The effect of the torque on the base of the inner cylinder may then be eliminated. For experimental details see Newman and Searle or Champion and Davy.

2.18 Determination of the Coefficient of Viscosity of a Very Viscous Liquid

The method is based on a formula due to Stokes, for the viscous force \( F \) which opposes the motion of a sphere of radius \( r \) moving with velocity \( v \) through a fluid of viscosity \( \eta \). \( F \) is given by

\[
F = 6\pi \eta rv.
\]

This is known as Stokes’ Law and assumes streamline flow.

As the velocity of the sphere increases, the viscous drag also increases, so that for a sphere falling vertically through the fluid there is a maximum or terminal velocity, when the net force on the sphere is zero.

The forces acting are:

(i) downwards: the weight of the sphere = \( \frac{4}{3} \pi r^3 \rho g \) where \( \rho \) is its density;
(ii) upwards:
   (a) the upthrust of the liquid = \( \frac{4}{3} \pi r^3 \sigma g \) where \( \sigma \) is the density of the liquid;
   (b) the viscous drag = \( 6\pi \eta rv \).

When the terminal velocity \( v_t \) has been reached,

\[
\frac{4}{3} \pi r^3 \rho g = \frac{4}{3} \pi r^3 \sigma g + 6\pi \eta rv_t
\]

\[
\therefore \eta = \frac{2r^2 g (\rho - \sigma)}{9v_t} \quad (2.10)
\]
The coefficient of viscosity of a very viscous liquid is determined by measuring the terminal velocity of a sphere falling through it and then calculating \( \eta \) from equation (2.10).

Stokes' formula involves two assumptions which must be remembered in designing the experiment:

(i) the flow of liquid past the sphere is streamline;
(ii) the sphere is falling in an 'infinite sea' of liquid.

The conditions necessary for streamline flow may be found from equation (2.3), which is applicable also to this case if \( r \) is interpreted as the radius of the sphere. The maximum permissible terminal velocity is thus \( 1,100\eta/\sigma \).

Thus

\[
\frac{2r^2g(\rho - \sigma)}{9\eta} < \frac{1,100\eta}{\sigma r},
\]

i.e.

\[
r^3 < \frac{4,950\eta^2}{\sigma(\rho - \sigma)g}.
\]

A maximum value for the radius of the sphere is thus set by the value of \( \eta^2/\sigma(\rho - \sigma) \), and the method becomes impractical if this fraction is too small.

The onset of turbulence may be discovered by performing the experiment with spheres of different radii and plotting \( v_t \) against \( r^2 \). This graph is a straight line for streamline flow, but begins to curve at the onset of turbulence.

**2.19 Viscosity of Gases**

The physical explanation of the viscous behaviour of gases is in terms of the Kinetic Theory and is dealt with in Section 6.9.

Some of the formulae derived in the previous sections need modification to take into account the compressibility of gases. For example, the derivation of Poiseuille's formula for the flow of liquid through a narrow tube assumes that the volume flowing per second is the same throughout all cross-sections of the tube. For a liquid this is a very good approximation, as the liquid is almost incompressible, but for a gas this assumption is certainly not valid, and equation (2.7) is not applicable.

The compressibility of a gas is taken into account as follows. Suppose that the pressures at the ends of the tube are \( p_1 \) and \( p_2 (p_1 > p_2) \) and that at a distance \( x \) from the inlet end the pressure is \( \rho \). Consider a further small length of the tube \( \delta x \) in which the pressure changes from \( \rho \) to \( \rho + \delta \rho \), i.e. there is a pressure drop of \( -\delta \rho \). Then the volume \( V \) flowing per second through the tube at this point is, by equation (2.7),

\[
V = \frac{\pi a^4(-\delta \rho)}{8\eta \delta x}.
\]

The mass of gas flowing per second across all cross-sections of the tube must be the same.
SURFACE TENSION AND VISCOSITY

If we assume that the gas obeys Boyle’s law the pressure and volume of this mass of gas at any point along the tube are related by:

\[ pV = K \text{ (a constant),} \]

\[ \frac{K}{p} = \frac{\pi a^4 (-\delta p)}{8 \eta \delta x}, \]

\[ \therefore 8 \eta K \int_0^t dx = -\pi a^4 \int_{p_1}^{p_2} p \, dp, \]

\[ \therefore 8 \eta Kl = \frac{\pi a^4 (p_1^2 - p_2^2)}{2}. \]

But

\[ K = p_1 V_1 = p_2 V_2, \]

where \( V_1, V_2 \) are the volumes of gas per second measured at pressures \( p_1, p_2 \) respectively.

\[ \therefore V_1 = \frac{\pi a^4 (p_1^2 - p_2^2)}{16 \eta p_1} \quad . \quad . \quad . \quad (2.11) \]

2.20 Measurement of the Coefficient of Viscosity of a Gas

Equation (2.11) may be applied directly, provided that the inlet and outlet pressures of the gas are kept the same throughout the experiment. The simplest experiment of this type uses the apparatus shown in Fig. 2.18.

The gas is forced through the capillary tube by a falling pellet of mercury in the wider right-hand tube. The time, \( t \), for the bottom of the pellet to pass between two fixed points \( A \) and \( B \) is measured. If the volume between \( A \) and \( B \) is \( V \), then the volume per second \( V_1 = V/t \).

The pressures are \( p_2 = \) atmospheric pressure, \( P \); and \( p_1 = P + p \), where \( p \) is the pressure produced by the pellet of mercury, i.e. \( p = mg/\pi r^2 \), where \( m \) is the mass of the pellet and \( r \) the radius of the tube \( AB \).

The value of \( \eta \) is found by substitution in equation (2.11). As usual, precautions must be taken to ensure that the flow is streamline. A correction has to be applied because the mercury sticks to the glass.

The rotating cylinder method is applicable also to the measurement of \( \eta \) for a gas and the formula (equation (2.9)) needs no modification, as incompressibility has not been assumed.

Stokes' formula is applicable also to an object falling in a gas. If \( \eta \) for the gas is known a measurement of the terminal velocity of a small drop of liquid in the gas enables the size of the drop to be calculated (e.g. in
QUESTIONS

1. In Section 2.4 it was stated that the angle of contact of a liquid with a solid is greater than 90° if the force along PL is large enough compared with that along PS (Fig. 2.5). Show that the condition is that the former force should be greater than \(\sqrt{2}\) times the latter.

2. A piece of wire is bent to form three sides of a rectangle. Between the ends of the wire is fixed a light thread of such a length that it is slack. Show that, if a soap film of surface tension \(\gamma\) is formed in the frame, the thread takes the shape of an arc of a circle and that the radius, \(r\), of the arc is related to the tension, \(T\), in the thread by \(T = 2\gamma r\).

3. Derive the formula, \(\ln \frac{\rho'}{\rho} = \frac{2\gamma M}{R\rho T}\), quoted at the end of Section 2.7. Calculate the first three rows in Table 2.1 using this formula; calculate the last row using the approximate formula \(\rho' - \rho = \frac{2\gamma}{r} \left( \frac{\sigma}{\rho} - \sigma \right)\).

4. Two discs each of radius \(r\) are mounted coaxially a distance \(d\) apart, the space between them being filled with a fluid of viscosity \(\eta\). Show that if one disc is rotated at constant angular velocity \(\Omega\), the other experiences a torque \(\pi\eta \Omega r^4/2d\). This experiment may be used to determine \(\eta\) for liquids or gases.

5. Use the method of dimensions to verify
   (i) Stokes' formula
   (ii) Poiseuille's formula

   (apart from the numerical constants).

6. Why is the falling-sphere method of determining \(\eta\) used only for very viscous liquids?

ESSAY TOPICS

7. Determination of surface tension.
8. Determination of the variation of surface tension with temperature.
9. Determination of viscosity (i) of liquids, (ii) of gases.
10. Spreading of films on liquids.
11. Lubrication.
OSCILLATIONS AND WAVES

OSCILLATIONS

3.1 Simple Harmonic Motion

A body is said to execute simple harmonic motion if it moves such that its acceleration is always proportional to its distance from a fixed point and directed towards that point. For a body moving along the x-axis this may be expressed mathematically as

$$\frac{d^2x}{dt^2} = -kx,$$

the fixed point in this case being the origin.

The constant $k$ is usually written as $\omega^2$ and the equation then becomes

$$\frac{d^2x}{dt^2} + \omega^2 x = 0 \quad . \quad . \quad . \quad (3.1)$$

This equation has a solution of the form

$$x = a \sin (\omega t + \epsilon) \quad . \quad . \quad . \quad (3.2)$$

which may be checked as follows

$$\frac{dx}{dt} = a\omega \cos (\omega t + \epsilon),$$

$$\frac{d^2x}{dt^2} = -a\omega^2 \sin (\omega t + \epsilon) = -\omega^2 x.$$

The graph of equation (3.2) is shown in Fig. 3.1.

It is seen that the maximum value of $x$ is $a$. This is called the amplitude of the motion.
The angle \((\omega t + \epsilon)\) is called the phase of the motion; \(\epsilon\) is known as the initial phase. The angle \(\epsilon\) may be made equal to zero by choosing the zero of time so that \(x = 0\) when \(t = 0\).

The period, \(T\), of the motion is the time taken for one complete oscillation.

Thus

\[
T = \frac{2\pi}{\omega}.
\]

The frequency, \(f\), is the number of oscillations per second and is thus \(1/\text{period}\), i.e.

\[
f = \frac{1}{T} = \frac{\omega}{2\pi}.
\]

The velocity, \(v\), of the body at any instant is found by differentiating equation (3.2) with respect to \(t\):

\[
v = \frac{dx}{dt} = a\omega \cos (\omega t + \epsilon).
\]

But

\[
\cos^2 (\omega t + \epsilon) = 1 - \sin^2 (\omega t + \epsilon) = 1 - \frac{x^2}{a^2},
\]

\[
\therefore v = \pm a\omega (1 - \frac{x^2}{a^2})^{\frac{1}{2}}
\]

\[
= \pm \omega (a^2 - x^2)^{\frac{1}{2}}.
\]

The maximum velocity, \(v_{\text{max}}\), occurs when \(x = 0\), and is thus

\[
v_{\text{max.}} = \omega a.
\]

The total energy of the system at any instant is the sum of its potential and kinetic energies at that instant. If there are no damping forces this total energy remains constant throughout the motion. In general, we may fix the zero of potential energy as we please, but if we wish to calculate the energy of the oscillations we must take this zero to be when the body is in its equilibrium position \((x = 0)\). The total energy of the oscillations is thus equal to the kinetic energy at \(x = 0\) and is given by \(\frac{1}{2}ma^2v_{\text{max.}}\), which is equal to \(\frac{1}{2}ma^2a^2\).

We see, then, that for a given system executing simple harmonic motion, the energy of the oscillations is proportional to the square of the amplitude.

### 3.2 Damped Simple Harmonic Motion

In all practical cases the motion of an oscillating body is affected by some type of damping force, opposing the motion, and hence altering the type of oscillations. For example, the motion of a body on the end of a spiral spring is damped by the viscous drag of the air. The damping is only slight in this case, but the effects of heavy damping may be demonstrated by immersing the bob in a liquid of suitable viscosity.

In many important cases the damping force is proportional to the
velocity of the body at any instant. This force opposes the motion so that it has the opposite sign to \( \frac{dx}{dt} \). If the damping force per unit mass is \( k \frac{dx}{dt} \) the acceleration of the body is now \(-k \frac{dx}{dt} - \omega^2 x\), so that the equation of motion becomes

\[
\frac{d^2x}{dt^2} = -k \frac{dx}{dt} - \omega^2 x,
\]
i.e.

\[
\frac{d^2x}{dt^2} + k \frac{dx}{dt} + \omega^2 x = 0 .
\] (3.3)

This equation has a solution of the form

\[
x = Ae^{mt}.
\]

Differentiating this with respect to \( t \) gives

\[
\frac{dx}{dt} = Am e^{mt},
\]
and

\[
\frac{d^2x}{dt^2} = Am^2 e^{mt}.
\]

Substituting in equation (3.3) leads to the equation for \( m \):

\[
m^2 + km + \omega^2 = 0,
\]
and therefore

\[
m = \pm \left( \frac{k^2 - 4\omega^2}{2} \right)^{\frac{1}{2}}.
\]

The complete solution for \( x \) includes both values of \( m \), and is therefore,

\[
x = e^{-\frac{k}{2}t} \left( Ae^{\frac{(k^2 - 4\omega^2)^{\frac{1}{2}}}{2}t} + Be^{-\frac{(k^2 - 4\omega^2)^{\frac{1}{2}}}{2}t} \right),
\]

\( A \) and \( B \) being constants which depend on the initial conditions.

Three separate cases arise according to whether \( k^2 \) is less than, equal to or greater than \( 4\omega^2 \).

(i) \( k^2 < 4\omega^2 \)

The square root involved is now that of a negative quantity and is re-written as \( j(4\omega^2 - k^2)^{\frac{1}{2}} \) where \( j = ( -1)^{\frac{1}{2}} \). If we put \( \theta = (4\omega^2 - k^2)^{\frac{1}{2}}/2 \) to simplify the writing of the equation, it becomes

\[
x = e^{-\frac{k}{2}t} \{ Ae^{j\theta t} + Be^{-j\theta t} \}.
\]

If we arrange the measurement of time so that \( x = 0 \) when \( t = 0 \), then

\[
A + B = 0,
\]

\[
\therefore \ x = Ae^{-\frac{k}{2}t} \{ e^{j\theta t} - e^{-j\theta t} \}.
\]
To simplify this result we use the facts that
\[ e^{j\theta t} = \cos \theta t + j \sin \theta t, \]
and \[ e^{-j\theta t} = \cos \theta t - j \sin \theta t, \]
so that \[ e^{j\theta t} - e^{-j\theta t} = 2j \sin \theta t, \]
and \[ x = 2jAe^{-\frac{k}{2}t} \sin \theta t. \]

Now, both \( x \) and \( t \) are real quantities, so that the constant \( 2jA \) must also be real (\( A \) itself must be imaginary) and we put \( 2jA = a \), so that
\[ x = ae^{-\frac{k}{2}t} \sin \theta t. \quad (3.4) \]

If there is no damping (i.e. \( k = 0 \)), then \( \theta = \omega \) and \( e^{-\frac{k}{2}t} = 1 \), so that equation (3.4) reduces to
\[ x = a \sin \omega t. \]

Returning now to an examination of equation (3.4), this solution is seen to be an oscillatory one because of the presence of the sine term. The period of the oscillations is \( 2\pi/\theta \), which is \( 2\pi/(\omega^2 - k^2/4)^{\frac{1}{2}} \). If \( k \) is very small compared with \( \omega \), then the period is approximately \( 2\pi/\omega \), the same as for the undamped oscillations.

The exponential term \( e^{-\frac{k}{2}t} \) means that the amplitude of the motion decreases with time, tending to zero as \( t \) tends to infinity.

Fig. 3.2 shows a graph of displacement against time for the oscillations.

\[ (ii) \quad k^2 > 4\omega^2 \]

\((k^2 - 4\omega^2)^{\frac{1}{2}}\) is now real, and the solution of equation (3.3) is
\[ x = Ae^{-\frac{k}{2}t}\left(e^{\frac{(k^2-4\omega^2)^{\frac{1}{2}}}{2}t} - e^{-\frac{(k^2-4\omega^2)^{\frac{1}{2}}}{2}t}\right) \quad (3.5) \]
where we have again taken \( x = 0, t = 0 \), as initial conditions.
Oscillations and Waves

There is now no oscillation and the value of $x$ increases to a maximum and then decreases steadily to zero as $t$ tends to infinity. This type of motion is known as dead beat.

The variation of $x$ with $t$ in this case is depicted in Fig. 3.3.

(iii) $k^2 = 4\omega^2$

The roots of the equation for $m$ are now both equal to $-k/2$ and the solution of equation (3.3) is

$$x = e^{-\frac{k}{2}t}[A + Bt]$$

(3.6)

This again is a motion in which there is no oscillation. This case is known as critical damping. The motion dies away most rapidly with this type of motion.

Suppose, again, that when $t = 0, x = 0$.

Substituting in equation (3.6) gives

$$A = 0.$$  

$$\therefore x = Bte^{-\frac{k}{2}t}.$$  

The curve for critical damping is shown in Fig. 3.3.

![Fig. 3.3](image)

The equations for damped simple harmonic motion derived above are applicable not only to mechanical systems but also have a much wider significance. They may be used to determine the flow of current in an electrical circuit containing inductance, capacitance and resistance (see Section 9.22) and to interpret the motion of the coil of a ballistic galvanometer (see Section 9.12).

3.3 Forced Simple Harmonic Motion – No Resistance

Suppose that a simple pendulum is making small oscillations and that the upper end of the string is then made to vibrate in a horizontal direction, with simple harmonic motion of small amplitude by some driving force. The motion of the bob is then altered and it is said to be
executing forced simple harmonic motion. It is well known that, if the frequency of the driving force approaches the natural frequency of vibration of the pendulum in the absence of a driving force, the amplitude of the motion becomes large; in this case the conditions for simple harmonic motion no longer apply.

Let us consider the ideal case of a system which executes simple harmonic motion at all amplitudes and is subject to no damping forces. If it has a natural frequency of vibration $\omega/2\pi$, then

$$\frac{d^2x}{dt^2} + \omega^2 x = 0.$$  

The application of a forcing motion of frequency $p/2\pi$ giving rise to an additional acceleration $A \sin pt$ leads to the equation

$$\frac{d^2x}{dt^2} + \omega^2 x = A \sin pt \quad . \quad . \quad (3.7)$$

The complete solution of a differential equation of this type has two parts: the complementary function and the particular integral. The former is the solution obtained by setting the left-hand side of equation (3.7) equal to zero, and is thus: $x = a \sin (\omega t + \epsilon)$ (see equation (3.2)).

The particular integral is any particular solution of the complete equation (3.7). There is a solution of the form,

$$x = B \sin pt + C \cos pt.$$  

Differentiating with respect to $t$:

$$\frac{dx}{dt} = Bp \cos pt - Cp \sin pt,$$

and

$$\frac{d^2x}{dt^2} = -Bp^2 \sin pt - Cp^2 \cos pt.$$  

Substituting in equation (3.7) gives

$$B(\omega^2 - p^2) \sin pt + C(\omega^2 - p^2) \cos pt = A \sin pt.$$  

Equating coefficients, first of $\sin pt$, and then of $\cos pt$, gives

$$B = \frac{A}{\omega^2 - p^2}, \text{ and } C = 0;$$

and the particular integral is,

$$x = \frac{A}{\omega^2 - p^2} \sin pt.$$  

The complete solution is the sum of the complementary function and the particular integral, i.e.

$$x = a \sin \omega t + \frac{A}{\omega^2 - p^2} \sin pt \quad . \quad . \quad (3.8)$$  

where $x = 0, t = 0$ have been taken as initial conditions.
OSCILLATIONS AND WAVES

The resultant motion is thus a combination of two simple harmonic motions. If the natural frequency of vibration and the frequency of the driving force are nearly the same, then \( \omega^2 - \omega^2 \) is very small, and \( A/(\omega^2 - \omega^2) \) very large. The second term in equation (3.8) then becomes the dominant one, and the resultant motion has very large amplitude, and frequency nearly \( \omega/2\pi \); the system is almost at a resonance position. At resonance, i.e. when \( \omega = \omega \), equation (3.8) is not a solution of equation (3.7).

* In practice, equation (3.8) does not even hold near the resonance position for two reasons. First, when the amplitude, and hence the velocity of the motion, become high effects of friction become increasingly important. Secondly, most practical vibrating systems execute simple harmonic motion for small amplitudes of oscillation only.

The effect of frictional or damping forces may be taken into account if one makes certain assumptions about the nature of such forces; a particular (electrical) example of this problem will be considered in Section 9.22.

WAVES

3.4 Wave Motion

The description ‘wave’ is applied to many different types of phenomena in physics. We talk of water waves, sound waves, electromagnetic waves, waves on strings, etc., but despite their apparent diversity, these must have some characteristics in common which enable them to be grouped together under the same general description – waves. These characteristics are: (i) there is a transfer of energy from one point to another; (ii) the medium through which the wave is propagated does not, as a whole, suffer any permanent displacement. It is useful to distinguish between two types of waves: transverse, in which the vibration of the particles in the medium is in a direction at right angles to the direction of travel of the wave; and longitudinal, in which the direction of vibration is parallel to the direction of propagation. Much of the mathematics of wave motion may, however, be carried out without reference to any particular type of wave, and the results obtained may then be modified to apply to particular cases. The sections which follow give such a general treatment with reference to actual examples where appropriate.

3.5 Harmonic Waves

We shall consider a wave of any type which is travelling along the \( x \)-axis with the velocity \( v \), and denote the disturbance which the wave produces by \( \psi \). Thus, for a wave on a string, \( \psi \) would denote the displacement of the particles of the string from their equilibrium positions;
for a sound wave, the displacement of 'particles' of air; and for an electromagnetic wave the value of the electric field. \( \psi \) is a function both of position (value of \( x \)) and of time.

If at a certain time, e.g. \( t = 0 \), we plot a graph of \( \psi \) against \( x \), we obtain a curve which is called the profile of the wave. If this wave profile has the shape of a sine curve the wave is called a harmonic wave. It is produced by a source which is vibrating with simple harmonic motion, as, for example, the prongs of a tuning fork. It can be shown that any periodic wave (i.e. one with a profile that is repeated at regular intervals of \( x \)) may be regarded as the sum of a number of harmonic waves. The study of harmonic waves is thus of fundamental importance in wave theory, and we shall confine ourselves to this class of waves, further restricting the discussion to waves travelling in one dimension only, i.e. along the \( x \)-axis.

3.6 The Equation of a Harmonic Wave

Fig. 3.4 depicts the profile of a harmonic wave at a time such that \( \psi = 0 \) at \( x = 0 \). The maximum value attained by \( \psi \) is called the amplitude of the wave and is denoted by \( a \). The distance between successive similar points on the profile is called the wavelength, and is denoted by \( \lambda \).

![Fig. 3.4](image)

The equation of the profile is determined as follows. Since \( x = 0 \) when \( \psi = 0 \), it is a sine curve which must be of the form \( \psi = A \sin Bx \). The curve first begins to repeat when \( x = \lambda \) so that \( Bx = 2\pi \) when \( x = \lambda \), i.e. \( B = \frac{2\pi}{\lambda} \). \( A \sin Bx \) has a maximum value \( A \), which is the amplitude \( a \). The profile is therefore

\[
 \psi = a \sin \left( \frac{2\pi x}{\lambda} \right) . \tag{3.9}
\]

After a time \( t \) has elapsed the profile of the curve will still have the same shape (if we assume that the wave is propagated without distortion), but will have been displaced a distance \( vt \) (\( v \) being the velocity of
the wave). If the origin of the $x$-axis is also displaced a distance $vt$ in the same direction the profile will be given by

$$\psi = a \sin \left( \frac{2\pi x'}{\lambda} \right)$$  \hspace{1cm} (3.10)

where $x'$ is the distance measured from the new origin (see Fig. 3.5).

![Diagram](image)

**Fig. 3.5**

It is seen that when $x' = 0$, $x = vt$, so that

$$x' = x - vt.$$

Substituting this value of $x'$ in equation (3.10) yields:

$$\psi = a \sin \frac{2\pi}{\lambda} (x - vt)$$  \hspace{1cm} (3.11)

which is one form of the equation of a harmonic wave.

The *period*, $T$, of the wave is the time taken for one complete wave to pass any point. This is the same as the time taken for the wave-front to move forward a distance $\lambda$, and hence $T = \lambda/v$.

Substituting $v = \lambda/T$ in equation (3.11) leads to

$$\psi = a \sin 2\pi \left( \frac{x}{\lambda} - \frac{t}{T} \right)$$  \hspace{1cm} (3.12)

which is a second form of the equation of a harmonic wave.

A third form of the equation is obtained in terms of two other properties of the wave: its *frequency*, $f$, which is the number of waves passing any point in unit time; and the *wave number*, $k$, which is the number of waves contained in unit distance.

The frequency is the reciprocal of the period, so that $f = 1/T$ and therefore $v = f\lambda$. 
The wave number is the reciprocal of the wavelength, i.e. \( k = 1/\lambda \). Substituting in equation (3.12) gives

\[
\psi = a \sin 2\pi(kx - ft) .
\] (3.13)

For a wave travelling towards the left, i.e. in the direction \(-x\), equations (3.11), (3.12), and (3.13) become:

\[
\psi = a \sin \frac{2\pi}{\lambda} (x + vt);
\]

\[
\psi = a \sin 2\pi \left( \frac{x}{\lambda} + \frac{t}{T} \right);
\]

\[
\psi = a \sin 2\pi(kx + ft).
\]

If two waves of the same frequency and velocity are both travelling along the \(x\)-axis they will not, in general, coincide, so that it is impossible to impose the condition \( \dot{\psi} = 0 \), at \( x = 0, t = 0 \) for both waves. Fig. 3.6 shows the profiles of two such waves at some instant. The equations of the two waves are:

\[
\psi_1 = a_1 \sin \left\{ \frac{2\pi}{\lambda} (x - vt) + \epsilon_1 \right\}
\]

and

\[
\psi_2 = a_2 \sin \left\{ \frac{2\pi}{\lambda} (x - vt) + \epsilon_2 \right\}.
\]

\( \epsilon_1 - \epsilon_2 \) is the phase difference between (1) and (2); it will be denoted by \( \phi \). The profile of wave (1) is displaced by a distance \( \phi\lambda/2\pi \) relative to the profile of wave (2).

3.7 The Equation of Wave Motion

The equation of a harmonic wave (equation (3.11)) may be shown to be a particular solution of a certain differential equation, known as the equation of wave motion, as follows:
Differentiating equation (3.11) partially with respect to \( x \) gives
\[
\frac{\partial \psi}{\partial x} = \frac{2\pi a}{\lambda} \cos \frac{2\pi}{\lambda} (x - vt),
\]
and
\[
\frac{\partial^2 \psi}{\partial x^2} = -\frac{4\pi^2 a}{\lambda^2} \sin \frac{2\pi}{\lambda} (x - vt).
\]

The partial derivatives with respect to \( t \) are given by:
\[
\frac{\partial \psi}{\partial t} = -\frac{2\pi va}{\lambda} \cos \frac{2\pi}{\lambda} (x - vt),
\]
and
\[
\frac{\partial^2 \psi}{\partial t^2} = -\frac{4\pi^2 v^2 a}{\lambda^2} \sin \frac{2\pi}{\lambda} (x - vt).
\]

\[
\therefore \frac{\partial^2 \psi}{\partial x^2} = \frac{1}{v^2} \cdot \frac{\partial^2 \psi}{\partial t^2}
\]  \hspace{1cm} (3.14)

This partial differential equation is the general equation governing any wave motion propagated with velocity \( v \) along the \( x \)-axis. A harmonic wave is only one particular solution of this equation.

For a wave in three dimensions the equation becomes
\[
\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = \frac{1}{v^2} \cdot \frac{\partial^2 \psi}{\partial t^2}.
\]

3.8 Stationary Waves

Stationary waves are formed when a progressive wave incident on a boundary is reflected back along the same path, so that this reflected wave interferes with the forward wave. The resultant disturbance at any point on the \( x \)-axis may be found by adding together the separate disturbances due to the two waves. This is known as the principle of superposition.

The two waves to be added together are of the same frequency and approximately the same amplitude, but are travelling in opposite directions. They may therefore be represented by the two equations:
\[
\psi_1 = a \sin 2\pi(kx - ft),
\]
and
\[
\psi_2 = a \sin 2\pi(kx + ft).
\]

The resultant disturbance \( \psi \) is the sum of \( \psi_1 \) and \( \psi_2 \):
\[
\psi = a\{\sin 2\pi(kx - ft) + \sin 2\pi(kx + ft)\}.
\]

Using the trigonometrical relation
\[
\sin A + \sin B = 2 \sin \frac{A + B}{2} \cdot \cos \frac{A - B}{2},
\]
we obtain
\[
\psi = 2a \sin 2\pi kx \cos 2\pi ft \quad . \quad \quad (3.15)
\]
This no longer represents a progressive wave because $x$ and $t$ do not occur together in the equation. This type of wave is called a stationary wave, and its characteristics may be deduced from equation (3.15). A 'particle' at distance $x$ from the origin executes simple harmonic motion of frequency $f$, and amplitude $2a \sin 2\pi kx$. The amplitude is zero for values of $x$ for which $\sin 2\pi kx = 0$;

i.e. \[ x = 0, \frac{1}{2k}, \frac{1}{k}, \frac{3}{2k}, \ldots \] \[ \text{i.e. } x = 0, \frac{\lambda}{2}, \frac{3\lambda}{2}, \ldots \]

These positions are called nodes.

The amplitude is a maximum wherever $\sin 2\pi kx = \pm 1$;

i.e. \[ x = \frac{1}{4k}, \frac{3}{4k}, \frac{5}{4k}, \ldots \] \[ \text{i.e. } x = \frac{\lambda}{4}, \frac{3\lambda}{4}, \frac{5\lambda}{4}, \ldots \]

These positions are antinodes.

Fig. 3.7 shows the profiles of the wave at $t = 0, t = 1/4f$ and $t = 1/2f$; i.e. $t = 0, T/4, T/2$.

3.9 Beats

Beats are produced when two notes of slightly different frequencies are sounded together. The situation is thus that of two waves of nearly the same frequency travelling in the same direction with the same velocity. For the sake of simplicity we shall assume that the amplitudes of the two waves are the same, so that they may be represented by the equations

\[ \psi_1 = a \sin \frac{2\pi}{\lambda_1} (x - vt) = a \sin 2\pi f_1 \left( \frac{x}{v} - t \right), \]
and \( \psi_2 = a \sin \frac{2\pi}{\lambda_2} (x - vt) = a \sin 2\pi f_2 \left( \frac{x}{v} - t \right) \).

Applying the principle of superposition to determine the resultant wave, we obtain
\[
\psi = \psi_1 + \psi_2 = a \left\{ \sin 2\pi f_1 \left( \frac{x}{v} - t \right) + \sin 2\pi f_2 \left( \frac{x}{v} - t \right) \right\}
= 2a \sin 2\pi \left( \frac{f_1 + f_2}{2} \right) \left( \frac{x}{v} - t \right) \cos 2\pi \left( \frac{f_1 - f_2}{2} \right) \left( \frac{x}{v} - t \right).
\]

This resultant is seen to be the product of two progressive waves, one of frequency \((f_1 + f_2)/2\), the other of frequency \((f_1 - f_2)/2\). Since \(f_1, f_2\) are only slightly different, \((f_1 + f_2)/2\) is almost the same as \(f_1\) or \(f_2\), and \((f_1 - f_2)/2\) is small.

The profile of the wave at some instant is shown in Fig. 3.8. The wave of the higher frequency, \((f_1 + f_2)/2\), has an amplitude which varies with \(x\) and \(t\) in accordance with the cosine factor in the above equation. The amplitude is zero twice for each complete cycle of the wave of frequency \((f_1 - f_2)/2\), so that there are \(f_1 - f_2\) zeros per second.

\[\text{Fig. 3.8}\]

If the two waves are sound waves one hears a note of frequency \((f_1 + f_2)/2\), which has a varying loudness, there being \(f_1 - f_2\) beats per second.

3.10 Interference

Interference takes place between two waves of the same frequency but different phases. If the amplitudes are the same, then the two waves may be represented by
\[
\psi_1 = a \sin \{2\pi (kx - ft) + \epsilon_1\},
\]
and
\[
\psi_2 = a \sin \{2\pi (kx - ft) + \epsilon_2\}.
\]
The resultant is given by
\[
\psi = \psi_1 + \psi_2 = 2a \cos \frac{\epsilon_1 - \epsilon_2}{2} \sin \left\{ 2\pi (kx - ft) + \frac{\epsilon_1 + \epsilon_2}{2} \right\}.
\]

This is a wave of the same frequency as the original waves but with amplitude \(2a \cos [(\epsilon_1 - \epsilon_2)/2]\). The amplitude thus depends on the
phase difference $\epsilon_1 - \epsilon_2$, ($= \phi$), between the two waves. If $\phi = (2m+1)\pi$, where $m$ is an integer, then $\cos (\phi/2) = 0$ and the amplitude is zero. If $\phi = 2m\pi$, then $\cos (\phi/2) = \pm 1$ and the amplitude is $2a$.

In general, the amplitude is $2a \cos (\phi/2)$. If the interfering waves are light waves, then one is concerned, not with the amplitude of the resultant, but with its intensity. The intensity, $I$, is proportional to the square of the amplitude, so that

$$I \propto \cos^2 \frac{\phi}{2} \quad . \quad . \quad . \quad (3.16)$$

In experiments such as Young’s double-slit experiment the phase difference is brought about by splitting the beam of light into two parts, which then travel different distances. The difference between the two distances is known as the path difference, and if this is $s$, then $\phi = 2\pi s/\lambda$. In terms of the path difference $s$ we have

$$I \propto \cos^2 \left( \frac{\pi s}{\lambda} \right).$$

3.11 Vector Method for the Addition of Waves

The discussion in the previous section was limited to the addition of two waves of the same amplitude and frequency. If it is desired to add together more than two waves of the same frequency and without the restriction of equal amplitudes, then the above treatment becomes difficult. In this case it is easier to use a vector method, in which the waves are represented in both amplitude and phase by suitable vectors. The amplitude and phase of the resultant wave are then found by the usual rules of vector addition.

Fig. 3.9 illustrates the addition of three waves of the same frequency, with amplitudes $a_1$, $a_2$, $a_3$, the first two differing in phase by $\phi_1$, and the second two by $\phi_2$.

The amplitude of the resultant is $a$, and the difference in phase between it and wave (1) is $\phi$.

This method is very useful in the study of interference and diffraction (see Chapter 4).

3.12 Dispersion

In the preceding sections we have assumed the velocity of propagation of the waves to be independent of the frequency. In many practical cases this is not so, and the wave velocity is dependent on the frequency of the wave. This phenomenon is known as dispersion.
OSCILLATIONS AND WAVES

Light waves undergo dispersion when they travel through glass, or indeed through any transparent medium. In a vacuum, however, they travel without dispersion.

If a harmonic wave travels through a dispersive medium it is still propagated without change in shape because it has only a single frequency, so that the change of velocity with frequency can have no effect. If, however, a wave which is a combination of two or more harmonic waves is transmitted through the medium it suffers a change in shape as it travels, because the harmonic waves of which it is composed travel with different velocities.

3.13 Group Velocity

We have previously seen that if two waves of slightly different frequencies are transmitted through a non-dispersive medium beats are produced, and the resultant wave has the profile shown in Fig. 3.8. This wave profile travels at the same velocity as the wave velocity of each of the two waves producing the beats.

In a dispersive medium the wave velocities of the two waves are slightly different and the profile moves at a velocity which is different from either of these two wave velocities. This velocity is known as the group velocity, and it is the velocity at which energy is transferred by the wave.

![Fig. 3.10](image)

To obtain an expression for the group velocity, consider the case of two waves of wavelengths \( \lambda \) and \( \lambda + \delta \lambda \) which are travelling through a dispersive medium with velocities \( v \) and \( v + \delta v \) respectively, and whose profiles are shown in Fig. 3.10. The resultant disturbance is found by adding together the two waves according to the principle of superposition, and the profile obtained is also shown in Fig. 3.10.

At the instant shown the two crests A and B coincide and produce a maximum in the resultant wave at C. The lower wave is moving faster,
so that a short time later the crests $A'$ and $B'$ will coincide and the maximum of the resultant will have moved back a distance $\lambda$ relative to the upper wave. The time, $\delta t$, taken for this to occur is the time taken for the two waves to move a distance $\delta \lambda$ relative to each other, at a relative velocity $\delta v$,

$$\delta t = \frac{\delta \lambda}{\delta v}.$$

Although the maximum of the resultant has moved back a distance $\lambda$ relative to the lower wave, this wave has itself moved forward a distance $v \delta t$ during the same time. The maximum of the resultant has therefore moved forward a distance $v \delta t - \lambda$ in time $\delta t$.

The group velocity is thus $(v \delta t - \lambda)/\delta t$

$$= v - \frac{\lambda}{\delta t},$$

$$= v - \frac{\delta v}{\delta \lambda}.$$

This expression has been derived on the assumption of two waves of slightly different wavelengths being the only ones present, but it may be extended to several waves, provided they all have wavelengths close to $\lambda$. The expression then becomes:

$$\text{Group velocity} = v - \lambda \frac{d\nu}{d\lambda}.$$

### 3.14 Modulation

A modulated wave is one whose waveform has been altered in some regular way by combination with a wave of different frequency. One of the most common examples of the use of modulated waves is the transmission of information by radio waves. A carrier wave of radio frequency (i.e. above 100 kc/s) is modulated by the wave of audio frequency (between 50 c/s and 10 kc/s) which it is required to transmit. The two most common ways in which this modulation is carried out are amplitude modulation and frequency modulation.

(a) Amplitude modulation

The amplitude of the carrier wave is caused to vary at the frequency of the audio-frequency wave. Suppose that the carrier wave has frequency $p/2\pi$ and the audio-frequency wave has frequency $\omega/2\pi$. The equation of the modulated wave is then

$$\phi = a(1 + m \sin \omega t) \sin pt,$$

where $a$ is the amplitude of the unmodulated carrier wave and $m$ is a constant dependent on the amplitude of the wave of audio frequency.
This waveform is depicted in Fig. 3.11; in practice, $p$ is much greater than $\omega$, but for the sake of clarity the figure has been drawn with a much smaller ratio between $p$ and $\omega$.

![Fig. 3.11](image)

(b) Frequency modulation

The frequency of the carrier wave is caused to vary at the audio frequency to be transmitted. The extent of the fluctuation in carrier frequency is determined by the amplitude of the modulating wave. The carrier waves used in sound broadcasting are of very high frequency (e.g. 90 Mc/s).

**QUESTIONS**

1. Show that the solution of equation (3.7) when $p = \omega$ (resonance) is $x = a \sin \omega t - \frac{At}{2\omega} \cos \omega t$. Sketch this curve.

2. For a forced S.H.M. with resistance proportional to velocity the equation of motion is $\ddot{x} + k\dot{x} + \omega^2 x = A \sin pt$. Show that the solution of this equation consists of two parts: terms with an amplitude that decreases with time (transients), and terms which persist. Show that the transients are represented by the equations in Section 3.2; and that the other terms are given by $x = D \sin (pt - \phi)$

where $D = \frac{A}{[(\omega^2 - p^2)^2 + k^2 p^2]^\frac{1}{2}}$, and $\tan \phi = \frac{k p}{\omega^2 - p^2}$.

3. A sphere of radius $r$ and mass $m$ is suspended from a light spiral spring. Show that the period of oscillation is given by $T = 2\pi (m/\mu)^{\frac{1}{2}}$, where $\mu$ is the force required to produce unit extension of the spring.

If the mass is immersed in a fluid of coefficient of viscosity $\eta$, show that the equation of motion is

$$\ddot{x} + k\dot{x} + \omega^2 x = 0,$$

where $k = (6\pi \eta r)/m$ and $\omega^2 = \mu/m$. (Assume Stokes' law for the viscous drag.) Hence show that the motion is oscillatory only if $9\pi^2 \eta^2 r^2 < \mu m$. 
4. Show from equation (3.4) that the amplitudes of successive oscillations are related by \( \ln \left( \frac{\alpha_1}{\alpha_2} \right) = \ln \left( \frac{\alpha_3}{\alpha_4} \right) = \ldots = \frac{\pi k}{\theta} = \text{constant} \). This constant is known as the logarithmic decrement (8). Hence show that \( \alpha \) may be determined from \( \alpha = \alpha_1 e^{\delta^4} \) (see also Chapter 9, Question 4).
WAVE PROPERTIES OF LIGHT

INTRODUCTION

4.1 Light as a Wave-motion

Interference, diffraction, and polarization are all phenomena which can be explained on the assumption that light is some form of wave-motion. It is well known that these waves are a combination of oscillating electric and magnetic fields (electromagnetic waves), but their nature need not concern us here, as all the results we require apply equally well to any type of transverse harmonic wave. Use will be made of several formulae deduced in Chapter 3, and the student is strongly advised to read that chapter before commencing the present one.

4.2 Huygens’ Construction

This construction is developed from Huygens’ principle, which states that every point on a wave-front may be considered as a source of secondary spherical wavelets spreading out from that point. The wave-front at time \( t \) later is then the envelope of the secondary wavelets at that time, in the forward direction only.

The simplest example of the construction is for a plane wave-front (AB in Fig. 4.1), travelling with velocity \( v \). To find the position of this wave-front at time \( t \) later by Huygens’ construction, several points are taken along AB and circles of radius \( vt \) are drawn with these points as centres. The new wave-front \( A'B' \) is the envelope of these circles and is a straight line parallel to AB a distance \( vt \) away from it.

In this simple case the position of \( A'B' \) could, of course, have been predicted without the aid of Huygens’ construction, but in more complicated cases the construction is very helpful. The student will already have come across its use in the treatment of reflection and refraction.

One justification for Huygens’ construction is that it leads to results verified by experiment; theoretically it is more difficult to justify. Why, for example, must the envelope of the wavelets in only the forward direction be drawn; why not that in the backward direction as well, thus predicting a wave travelling backwards? Further, why should only the points of contact of the wavelets with their envelope produce a disturbance, and all other parts of the wavelets cancel each other out?
The answer to these questions requires a mathematical treatment far beyond the scope of this book, and the reader will have to be content with the knowledge that such a treatment has been carried out and is the theoretical justification for Huygens’ principle. One important factor that emerges from the full treatment is that the amplitude of the secondary wavelet varies as \((1 + \cos \theta)\), where \(\theta\) is the angle the radius makes with the forward direction (Fig. 4.2).

The amplitude at C is \(1 + \cos 0 = 2\),
\[
\text{at } D \text{ is } 1 + \cos \pi/2 = 1,
\]
\[
\text{at } E \text{ is } 1 + \cos \pi = 0.
\]

This accounts for the absence of a wave in the backward direction. \((1 + \cos \theta)\) is called the obliquity factor.

### 4.3 Optical Path

The velocity of light in any medium is less than its velocity in a vacuum, so that in a certain time the distance it travels in the medium is less than the distance it travels in a vacuum. The refractive index \(\mu\) of the medium is the ratio of the velocity of light, of specified wavelength, in a vacuum to its velocity in the medium. It follows that the time taken to traverse a distance \(l\) in the medium is the same as that to travel a distance \(\mu l\) in a vacuum. The quantity \(\mu l\) is called the optical path length in the medium.

### 4.4 Derivation of the Thin Lens Formula

In Fig. 4.3 O is an object and I its image formed by a thin lens whose faces have radii of curvature \(r\) and \(s\). The incident wave-front ABC is
part of a spherical surface with centre $O$, and the spherical wave-front $DEF$ emerging from the lens has centre of curvature $I$. The dotted curves $MB'N$ and $ME'N$ are drawn with centres $O$ and $I$ respectively such that $AM = BB'$ and $MD = EE'$.

Use is made of the rectangular property of the chords of a circle (Fig. 4.4):

$$h^2 = x(2R - x),$$
$$= 2Rx \text{ if } x \ll R,$$
$$\therefore x = \frac{h^2}{2R}.$$

As $ABC$ and $DEF$ are wave-fronts, the optical path length from $ABC$ to $DEF$ in air must be the same as that in the glass.

$$\therefore AM + MD = \mu BE.$$ 

But $AM = BB' = BP + PB' = \frac{h^2}{2r} + \frac{h^2}{2OB'} = \frac{h^2}{2r} + \frac{h^2}{2u}$,

as $u = OP; OB' \approx$

and $MD = EE' = EP + PE' = \frac{h^2}{2s} + \frac{h^2}{2IE'} = \frac{h^2}{2s} + \frac{h^2}{2v}$,

as $v = IP \approx IE'$.

Also $BE = BP + PE = \frac{h^2}{2r} + \frac{h^2}{2s}$.

$$\therefore h^2 \left( \frac{1}{2r} + \frac{1}{2u} + \frac{1}{2s} + \frac{1}{2v} \right) = \mu \left( \frac{h^2}{2r} + \frac{h^2}{2s} \right),$$

$$\therefore \frac{1}{v} + \frac{1}{u} = (\mu - 1) \left( \frac{1}{r} + \frac{1}{s} \right). \quad (4.1)$$

It has been assumed in the derivation that all distances are positive. The signs of equation (4.1) are adjusted to suit the two major sign conventions thus:

Real is Positive. \hspace{1cm} New Cartesian.

$$\frac{1}{v} + \frac{1}{u} = (\mu - 1) \left( \frac{1}{r} + \frac{1}{s} \right). \hspace{1cm} \frac{1}{v} - \frac{1}{u} = (\mu - 1) \left( \frac{1}{r} - \frac{1}{s} \right).$$

4.5 Fermat’s Principle

Fermat’s original statement of the principle was incomplete. This was that a ray of light travels between two points along that path which takes the least time.

In fact, the time along the ray may be at any stationary value. The
principle is usually stated in terms of the optical path rather than the
time, thus:

*A ray of light travels along that path whose optical path length is a stationary
value, i.e. is to a first approximation equal to neighbouring paths.*

Reflection and refraction at plane surfaces are examples of a mini-
mum optical path, reflection at a spherical surface of a maximum and
refraction through a lens of a point of inflexion (Fig. 4.5).

![Fig. 4.5](image)

**4.6 Fermat’s Principle Applied to Refraction and Reflection at a Plane Surface**

(a) *Refraction*

O and I (Fig. 4.6) are two fixed points in media of refractive indices
$\mu_1$ and $\mu_2$, respectively, separated by a plane boundary. The distances
$h_1$, $h_2$ and $d$ are therefore fixed and the variable point is B, that at which
the path of a ray from O to I crosses the boundary. Let this be at

![Fig. 4.6](image)
distance $x$ from C. The length, $p$, of the optical path from O to I is given by:

$$p = \mu_1 OB + \mu_2 BI$$

$$= \mu_1 (OA^2 + AB^2)^{\frac{1}{2}} + \mu_2 (IC^2 + CB^2)^{\frac{1}{2}}$$

$$= \mu_1 (h_1^2 + (d - x)^2)^{\frac{1}{2}} + \mu_2 (h_2^2 + x^2)^{\frac{1}{2}}.$$

This has a stationary value when $dp/dx = 0$.

$$\frac{dp}{dx} = \frac{1}{2} \mu_1 (h_1^2 + (d - x)^2)^{-\frac{1}{2}}(-2(d - x)) + \frac{1}{2} \mu_2 (h_2^2 + x^2)^{-\frac{1}{2}} 2x$$

$$\therefore 0 = -\frac{AB}{OB} + \frac{CB}{BI},$$

$$\therefore \mu_1 \sin i = \mu_2 \sin r,$$

which is Snell’s law.

(b) Reflection

In this case O and I are on the same side of the boundary so that the above leads to

$$\mu_1 \sin i = \mu_1 \sin r,$$

$$\therefore i = r.$$

INTERFERENCE

4.7 Double-slit Fringes

The student should already be familiar with the basic principles of Young’s double-slit experiment, and the details of this will not be repeated here. The important results are quoted for reference.

In Fig. 4.7, S is the single slit; $S_1$, $S_2$ the double slits distance $d$ apart; and LMON the screen, distance $D$ from the double slits.

Fringes are formed on the screen, the maxima occurring when the path difference $S_2M - S_1M$ is an integral number of wavelengths, and minima when $S_2M - S_1M$ is an odd number of half-wavelengths.
Path difference \( S_2M - S_1M = \frac{x'd}{D} \).

For minima \( \frac{x'd}{D} = (n + \frac{1}{2})\lambda; \) i.e. \( x = (n + \frac{1}{2})\frac{\lambda D}{d} \).

For maxima \( \frac{x'd}{D} = n\lambda; \) i.e. \( x = n\frac{\lambda D}{d} \).

The separation of the fringes is \( (n + 1)\frac{\lambda D}{d} - n\frac{\lambda D}{d} = \frac{\lambda D}{d} \).

We shall now consider some further points in connection with the formation of the fringes.

### 4.8 Intensity Distribution in Young's Fringes

In Fig. 4.7 the two waves arriving at \( M \) have a path difference of \( x'd/D \), if we take \( \mu = 1 \). The phase difference \( \phi \) between the two waves will be given by \( \phi = 2\pi x'd/\lambda D \).

(Phase difference of \( 2\pi \approx \) path difference of \( \lambda \).)

If the two waves have equal amplitudes, \( a \), their resultant may be found by the vector method of addition (see Section 3.11).

The resultant \( R \) (Fig. 4.8) is given by:

\[
R^2 = a^2 + a^2 - 2a \cdot a \cdot \cos(\pi - \phi) = 2a^2(1 + \cos \phi) = 4a^2 \cos^2 \frac{\phi}{2}.
\]

But the square of the resultant amplitude is proportional to the intensity*, \( I \),

\[
\therefore I = 4ka^2 \cos^2 \frac{\phi}{2},
\]

where \( k \) is a constant of proportionality.

\[
\therefore I = 4ka^2 \cos^2 \left( \frac{\pi dx}{\lambda D} \right) \quad \ldots \quad (4.2)
\]

A plot of this curve is shown in Fig. 4.9.

Notice that this gives maxima when \( \cos^2 \left( \frac{\pi dx}{\lambda D} \right) = 1 \).

* The intensity of a wave is the energy crossing unit area, perpendicular to the direction of travel of the wave, per second; it was shown in Section 3.1 that the energy of a S.H.M. is proportional to the square of the amplitude.
WAVE PROPERTIES OF LIGHT

i.e. \[ \frac{\pi dx}{\lambda D} = n\pi, \therefore x = \frac{n\lambda D}{d}; \]

and minima when \[ \cos^2 \left( \frac{\pi dx}{\lambda D} \right) = 0, \]
i.e. \[ \frac{\pi dx}{\lambda D} = (n + \frac{1}{2})\pi, \]
\[ \therefore x = \frac{\left(n + \frac{1}{2}\right)\lambda D}{d}; \]

in accordance with the simple theory. Between the centres of the fringes the intensity varies as \( \cos^2 \left( \frac{\pi dx}{\lambda D} \right) \).

Equation (4.2) may be obtained also by the algebraic addition of two harmonic waves (see Section 3.10).

4.9 Effect of Finite Width of Source Slit

It has so far been assumed that the width of the source slit S is infinitesimal, but in practice the effect of the finite width of this slit must be taken into account.
The slit S may be considered to be made up of a large number of parallel narrow slits, each one acting as the source slit of an interference pattern on the screen. These patterns are all displaced relative to one another corresponding to the different positions of these source slits. Fig. 4.10 shows the effect of considering just two such slits $S', S''$.

The paths of the rays forming the central maxima are drawn, and also the intensity distribution curves. If $S', S''$ are close together, then the fringes are only slightly displaced, but as $S', S''$ are moved apart the fringes separate until the maxima of one set fall in the same places on the screen as the minima of the other, and the resultant illumination of the screen by these two slits is uniform.

To find the width of S necessary for the fringes to disappear we must take into account the whole width of the slit and not just two portions of it as above. This is done by dividing the slit into a large number of narrow sources and treating these in pairs as shown in Fig. 4.11.

If the conditions are such that the sources A, $A'$ produce sets of fringes that exactly cancel each other, then the same will be true of B, $B'$; C, $C'$; ..., and there will be no fringes on the screen.

![Diagram](image)

The conditions for this to occur will now be deduced. In Fig. 4.12 the distance $AA'$ is represented by $\delta y$ (i.e. the source slit is of width $2\delta y$). If the source slit is placed with its centre on the perpendicular bisector of a line joining the 'double slits, then the central maximum of the fringe system produced by $A'$ will be at O. We require, then, to find the value of $\delta y$ for which the first minimum on one side of the centre of the fringe system due to A is also at O. This is when $AS_2 - AS_1 = \lambda/2$.

Now

$$(AS_1)^2 = \left(\frac{d}{2} - \delta y\right)^2 + l^2$$
WAVE PROPERTIES OF LIGHT

where \( l \) is the distance between the source slit and the double slits,

and

\[
(AS_2)^2 = \left(\frac{d}{2} + \delta y\right)^2 + l^2
\]

\[
\therefore (AS_2)^2 - (AS_1)^2 = 2d\delta y,
\]

\[
\therefore (AS_2 - AS_1)(AS_2 + AS_1) = 2d\delta y.
\]

If \( d \ll l \), then \( AS_2 + AS_1 = 2l \)

\[
\therefore AS_2 - AS_1 = \frac{2d\delta y}{2l}.
\]

The required condition is

\[
AS_2 - AS_1 = \frac{\lambda}{2},
\]

i.e.

\[
\frac{d\delta y}{l} = \frac{\lambda}{2}.
\]

\[
\therefore \text{width of source slit } (= 2\delta y) = \frac{l\lambda}{d}.
\]

This is the condition for the first complete disappearance of the fringes. As the source slit is widened the fringes reappear (but are less distinct than before), and disappear again when \( 2\delta y = 2l\lambda/d \), and so on.

The conditions for disappearance may also be expressed in terms of the angle \( \alpha \) subtended by the edges of the single slit at a point midway between the double slits (Fig. 4.13).

\[
\alpha \simeq \frac{2\delta y}{l}
\]

\[
\therefore \text{for disappearance of the fringes, } \alpha = \frac{\lambda}{d}, \frac{2\lambda}{d}, \frac{3\lambda}{d}, \ldots
\]

4.10 Conservation of Energy

Although there is no light energy at the position of the interference minima, this energy has not been destroyed but has been redistributed so that the places where there are maxima are brighter than they would be in the absence of interference. The total energy is unchanged.

If the two incident beams have equal amplitudes \( a \), then their total intensity is proportional to \( 2a^2 \), i.e. \( I = 2ka^2 \), where \( k \) is a constant.

We have seen (equation (4.2)) that when these two beams interfere, the intensity, \( I \), of the resultant fringes is given by

\[
I = 4ka^2 \cos^2 \frac{\phi}{2}.
\]
The average intensity of these fringes is

\[ I_{av} = 4ka^2 \cos^2 \frac{\phi}{2} \]

\[ = 4ka^2 \cdot \frac{1}{2}(1 + \cos \phi) \]

\[ = 2ka^2 \text{ as } \cos \phi = 0, \]

where the average is taken over \( \phi = 0 \) to \( 2\pi \).

Thus the average intensity is the same as the sum of the intensities of the two incident beams, i.e. energy is conserved (see Fig. 4.14).

**Fig. 4.14**

### 4.11 Interference Effects Produced by Thin Films

When a beam of light is incident on the surface of a film part of it passes into the film and is refracted and part is reflected from the surface (Fig. 4.15). A series of partial reflections and refractions at both surfaces of the film (assumed parallel) produce the emergent rays \( R_1, R_2, R_3, \ldots \) and \( T_1, T_2, T_3, \ldots \). Successive rays are of diminishing intensity.

**Fig. 4.15**
Interference effects may be observed in either the reflected or the transmitted rays. For interference to take place the parallel rays (e.g. \( R_1, R_2, R_3, \ldots \)) must be brought together by a lens. This is often done by the eye lens, as, for example, when colours are seen in films of oil floating on water.

### 4.12 Thin Films – Reflected Light

The path difference between rays \( R_1 \) and \( R_2 \) is found as follows (Fig. 4.16).

\[
\text{We require to find the optical path difference between rays travelling either in air along AD or in the film along ABC. DC is perpendicular to the emergent rays. Suppose the film is of thickness \( t \) and refractive index \( \mu \) and the incident light of wavelength \( \lambda \).}
\]

\( i \) and \( r \) are the angles of incidence and refraction at A, respectively.

\[
\begin{align*}
\text{Optical path in film} & = \mu(AB + BC) = 2\mu AB. \\
\text{Optical path in air} & = AD = AC \sin i. \\
\text{But} & \\
AC & = 2AE = 2AB \sin r, \\
\therefore AD & = (2AB \sin r) \sin i \\
& = 2\mu AB \sin^2 r, \text{ as } \sin i = \mu \sin r.
\end{align*}
\]

\[
\text{Optical path difference} = 2\mu AB(1 - \sin^2 r) = 2\mu AB \cos^2 r.
\]

\[
\begin{align*}
\text{But} & \\
AB & = t/\cos r. \\
\therefore \text{optical path difference} & = 2\mu t \cos r.
\end{align*}
\]

There is a phase change of \( \pi \) in the light reflected at A, so that the condition for constructive interference is that the optical path difference shall be an odd number of half-wavelengths, and for destructive interference an even number of half-wavelengths. Thus:

\[
\begin{align*}
\text{for maxima } & 2\mu t \cos r = (n + \frac{1}{2})\lambda & (4.3) \\
\text{for minima } & 2\mu t \cos r = n\lambda & (4.4)
\end{align*}
\]

This applies only as far as rays \( R_1, R_2 \) are concerned. Between rays \( R_2, R_3 \) there is the same path difference as between \( R_1 \) and \( R_2 \), but now
there are no phase changes as both reflections are for light incident on
the boundary from the medium of greater refractive index. It follows
that if equation (4.4) is fulfilled, $R_1$ and $R_2$ are out of phase, but $R_2,
R_3, R_4, \ldots$ are all in phase with each other. It can be shown that if the
total amplitude of $R_2, R_3, R_4, \ldots$ all in phase is found, this is just
equal to (and $\pi$ out of phase with) the amplitude of $R_1$. The condition
$2\mu t \cos r = n\lambda$ thus gives a position of complete darkness in the re-
lected beam.

If equation (4.3) is fulfilled, $R_1$ and $R_2$ are in phase with each
other and with $R_4, R_6, \ldots$ etc. $R_3, R_5, \ldots$ are in phase with each
other but $\pi$ out of phase with the other group. As the first group con-
tains the two most intense rays ($R_1$ and $R_2$), it will have a much greater
intensity than the second, so that constructive interference occurs.

![Fig. 4.17](image)

An interference pattern can be observed only if an extended source
of light is used. If a point source of light is used, then the eye, when at
any point, sees, not an interference pattern, but only a reflection of the
source in the film (Fig. 4.17). This image is bright or dark, depending
on the angle at which the light is reflected from the film into the eye,
and for certain angles no image is seen at all (i.e. when equation (4.4)

![Fig. 4.18](image)
is fulfilled). The complete pattern can be seen only by moving the eye from place to place to pick up the rays reflected at various angles.

If an extended source is used rays reflected at a range of angles can all enter the eye at the same time (Fig. 4.18). A large area of the film appears illuminated and an interference pattern is seen. A fringe given by a particular value of \( n \) in equation (4.3) or (4.4) occurs at all points where \( r \) is the same, i.e. the fringes are circular.

### 4.13 Thin Films – Transmitted Light

For a film of refractive index \( \mu \) and thickness \( t \), in air, the path difference between successive transmitted rays (e.g. \( T_1 \) and \( T_2 \) in Fig. 4.15) is the same as that between successive reflected rays, i.e. \( 2\mu t \cos r \), where \( r \) is the angle of refraction. There are no phase changes due to reflection, so that the condition for constructive interference is

\[
2\mu t \cos r = n\lambda.
\]

If \( 2\mu t \cos r = (n + \frac{1}{2})\lambda \), then \( T_1 \) and \( T_2 \) are \( \pi \) out of phase, and so are \( T_2 \) and \( T_3 \). This means, however, that \( T_1 \) and \( T_3 \) are in phase.

We have then \( T_1, T_3, T_5, \ldots \) etc., in phase with each other, and \( T_2, T_4, T_6, \ldots \) etc., in phase with each other, but \( \pi \) out of phase with the first group.

![Fig. 4.19](image)

Now, the amplitude of \( T_1 \) is greater than that of \( T_2, T_3 \) greater than \( T_4 \) and so on. It follows that the amplitude of the resultant of the first group is greater than that of the second, so that the two do not completely annul each other and the minima are not of zero intensity. Fig. 4.19 shows a typical variation of intensity with angle in the reflected and transmitted light.
4.14 White Light Fringes

If a thin film is illuminated by an extended source of white light, then coloured interference fringes are seen, as, for example, in soap bubbles or oil floating on water. For light incident on the film at a given angle the condition $2\mu t \cos r = n\lambda$ may be fulfilled for one or more values of $\lambda$ in the visible spectrum. These colours are therefore missing from the reflected beam, whose colour is thus a mixture of the remaining wavelengths.

If the film is several wavelengths thick, then the condition is fulfilled for many values of $\lambda$ in the visible spectrum, corresponding to different values of $n$. Suppose, for example, that $t = 10^{-4}$ cm, $\mu = \frac{3}{5}$ (oil), and that the light is incident normally. Then $\cos r = 1$ and $2 \cdot \frac{3}{5} \cdot 10^{-4} = n\lambda$.

$$\therefore \lambda = \frac{3 \cdot 10^{-4} \text{ cm}}{n} = \frac{3 \cdot 10^4}{n} \text{ Å}.$$  

The visible spectrum stretches from $\lambda = 4,000$ Å to $\lambda = 7,000$ Å.

For $n = 1$, $\lambda = 30,000$ Å.  
$n = 5$, $\lambda = 6,000$ Å.

$n = 2$, $\lambda = 15,000$ Å.  
$n = 6$, $\lambda = 5,000$ Å [visible.]

$n = 3$, $\lambda = 10,000$ Å.  
$n = 7$, $\lambda = 4,290$ Å

$n = 4$, $\lambda = 7,500$ Å; infra-red. $n = 8$, $\lambda = 3,750$ Å; ultra-violet.

At wavelengths close to the above the intensity is reduced, as the conditions are those for partially destructive interference. Interference maxima occur where

$$2\mu t \cos r = \left(n + \frac{1}{2}\right)\lambda,$$

i.e.

$$\lambda = \frac{3 \cdot 10^4}{(n + \frac{1}{2})} \text{ Å}.$$  

For $n = 4$, $\lambda = 6,700$ Å.  
$n = 6$, $\lambda = 4,620$ Å.

$n = 5$, $\lambda = 5,560$ Å.  
$n = 7$, $\lambda = 4,000$ Å.

At these wavelengths in the visible spectrum constructive interference takes place, and these colours are stronger. The resultant colour as seen by the eye is difficult to predict in this case, but the light will be only slightly coloured because of the large number of wavelengths involved.

However, if the film is thinner a much smaller number of wavelengths is involved and the reflected light will be a very definite colour. The student should work out the appearance at normal incidence, of a film of thickness $3 \times 10^{-5}$ cm.

When the film is very thin (i.e. $t \ll \lambda$) the actual optical path difference of the two rays is very small, but the phase change at the first reflection means that the two rays are practically $\pi$ out of phase, for all wavelengths. Under these conditions the film appears black. It is sometimes possible to obtain soap films sufficiently thin to demonstrate this effect.
4.15 Non-reflecting Coatings — ‘Blooming’

If a glass surface is coated with a thin film whose refractive index, \( \mu \), is less than that of the glass, then light reflected either at the top or bottom surfaces of the coating undergoes a phase change of \( \pi \). If light of wavelength \( \lambda \) is incident normally on the film there will be destructive interference in the reflected light if the thickness \( t \) of the film is such that

\[
2\mu t = \frac{\lambda}{2}.
\]

This process is often used to reduce reflection from lens surfaces. These are coated with a thin film of thickness \( \lambda/4\mu \), and reflection of light of wavelength \( \lambda \) is almost entirely eliminated. \( \lambda \) is chosen in the middle of the visible spectrum, and there is some reflection at the red and blue ends of the spectrum. For this reason coated or ‘bloomed’ surfaces appear purple when viewed by reflected light.

The reduction in reflected light brought about by this process results, of course, in a corresponding increase in the intensity of the transmitted light.

4.16 Wedge Fringes

If a wedge-shaped film of narrow angle is viewed by reflection of monochromatic light incident normally, a series of alternate dark and bright bands parallel to the edge of the wedge is observed. A dark band occurs at points where the thickness \( t \) of the wedge is such that

\[
2\mu t = n\lambda,
\]

and the centre of a bright band where \( 2\mu t = (n + \frac{1}{2})\lambda \).

It is assumed that the refractive indices of the media on each side of the film are either both greater, or both less, than \( \mu \).

4.17 Newton’s Rings

The arrangement for the observation of Newton’s rings is shown in Fig. 4.20. A source of monochromatic light, \( S \), is placed at the focus of
the lens $L_1$ and produces a horizontal parallel beam. This beam is partially reflected downwards by a glass plate at an angle of 45° to the horizontal and is then incident on the plane surface of a plano-convex lens, $L_2$, of long focal length, resting on a flat glass plate. If the microscope is focused on the space between the curved surface of $L_2$ and the plate on which it rests, a series of dark and bright circles is seen, centred on the point of contact of the lens and plate. These fringes are formed by interference between light reflected from the curved surface of the lens and from the upper surface of the plate (Fig. 4.21).

![Fig. 4.21](image)

The path difference at any point is $2\mu t$ as before (almost normal incidence), where $t$ is the thickness of the air film ($\mu = 1$) between lens and plate at that point. There is one phase change, on reflection at the plate, so that the interference conditions are:

\[
2t = n\lambda \quad \text{dark fringe} \quad \quad \quad (4.5)
\]

\[
2t = (n + \frac{1}{2})\lambda \quad \text{bright fringe}. \quad \quad \quad (4.6)
\]

The fringes thus follow lines of equal $t$, i.e. circles centred on the point of contact of lens and plate.

The radii of the rings may be deduced as follows (Fig. 4.22).

$R$ is the radius of curvature of the bottom surface of the lens, and $r$ the radius of one of the rings.

By the rectangular property of chords of a circle

\[
r^2 = (2R - t)t
\]

\[
= 2Rt \text{ because } R \gg t.
\]

Substituting for $t$ in equation (4.5) gives

\[
\frac{2r^2}{2R} = n\lambda
\]
i.e. 

\[ r = (nR\lambda)^\frac{1}{2}, \text{ dark ring}, \]

and in equation (4.6), 

\[ r = [(n + \frac{1}{2})R\lambda]^\frac{1}{2}, \text{ bright ring}. \]

There is a dark area in the centre of the system, because here the path difference is zero, and the phase change of \( \pi \) on reflection causes destructive interference.

The wavelength of the light may be calculated by measuring the radii of the rings (see a practical manual for details).

Newton's rings may also be observed in the transmitted light.

4.18 Classification of Fringes

(a) Localized and non-localized fringes

It was mentioned that to see Newton's rings the microscope has to be focused on the region of the air film. This is because the rays which interfere to produce a fringe at a certain point (Fig. 4.23) diverge from the point A which is in the vicinity of the air gap. Fig. 4.23 shows an exaggerated picture.

![Fig. 4.23](image)

The fringes are said to be localized at A. The same is true of the fringes produced by reflection from a wedge-shaped film.

In the case of a parallel-sided film the rays entering the eye or microscope are parallel (Fig. 4.24), so that the eye is focused on infinity; the fringes are localized at infinity.

Young's double-slit fringes, on the other hand, may be observed on a screen placed at almost any distance from the slits, and no focusing is required. These fringes are said to be non-localized. Other examples are Fresnel's biprism and Lloyd's mirror.

(b) Division of wave-front and division of amplitude

Young's double-slit fringes are produced by dividing the wave-front from the single slit into two parts by means of the double slits, and then allowing these two wave-fronts to come together to produce interference effects. The fringes are said to have been produced by division of wave-front. This is also true of Fresnel's biprism and Lloyd's mirror.
In the case of the parallel-sided film the whole area of the wave-front forms both reflected and refracted rays, but its amplitude is divided between the two. The fringes are said to be produced by division of amplitude. Wedge fringes and Newton's rings are produced in a similar way.

4.19 Interferometry

Interference effects may be used in many ways, and it is possible to give only the briefest outline here. It is hoped that the interested student will be encouraged to read about some of these interferometers in more detail (see bibliography).

(a) Rayleigh interferometer (refractometer)

This was developed to measure the refractive indices of gases. It is based on double-slit interference, the two rays travelling through different tubes into which gases may be introduced (Fig. 4.25).

![Diagram of Rayleigh interferometer](image)

Fig. 4.25

The two beams are brought together by the microscope, and interference fringes are observed. If the pressure of the gas in one of the tubes is altered its refractive index alters, and so the optical path length \((\mu l)\) in that tube also changes. The path difference between the two rays alters, and so the fringe system moves. If the pressure is changed very slowly the number of fringes moving across the cross-wire of the microscope may be counted. Every time one bright fringe has been replaced by another the optical path difference will have changed by one wavelength. Thus if \(n\) bright fringes cross the field of view, the change in refractive index \(\delta \mu\) is given by

\[ l \delta \mu = n \lambda \]  \hspace{1cm} (4.7)

If one tube is initially evacuated \((\mu = 1)\) and the gas then slowly introduced into it the refractive index of the gas at any required pressure may be found by counting the number of fringes and using equation (4.7).

(b) Michelson interferometer

The principle of the design of this instrument is shown in Fig. 4.26. Monochromatic light from the extended source S is divided at the
lower (silvered) surface of the glass plate P. Part of the beam is reflected back through the plate and then on to the mirror M₁, which it strikes normally, so that it goes back through P and into the telescope T. The rest of the beam is refracted at the lower surface of P, travels to the mirror M₂ and back along its own path to P, where it is reflected into the telescope.

A compensating plate P’ is usually introduced so that both beams have travelled the same distance in glass. This is necessary only if white light is used, its purpose being to cancel the effect of dispersion in the glass.

Fig. 4.26

The two beams entering T have travelled along different paths and interfere to produce fringes, which are circular if M₁, M₂ are accurately perpendicular and P is at 45° to them. These fringes are equivalent to those produced by a film of thickness equal to the distance between M₁ and M₂’ (Fig. 4.26).

If M₂ is moved a distance λ/2 (where λ is the wavelength of the light from S), the distance travelled by the beam reflected from it is altered by λ so that the fringe system is displaced by just one fringe. Thus by observing through T, very small movements of M₂ may be detected and measured.

The length of the standard metre in terms of the wavelength of red cadmium light has been determined by the use of a modified Michelson interferometer. It was not possible to move M₂ a whole metre and count the number of fringes (3,106,327 of them), but instead a series of intermediate sub-standards, known as etalons, was constructed. Each etalon consisted of a pair of mirrors at a fixed distance apart. These distances
ranged from 0.39 to 10 cm, each etalon being approximately twice the length of the previous one. The length of the smallest etalon was determined directly by counting fringes; the length of the next was found in terms of a multiple of the length of the first plus a few fringes, and so on up to the longest, which was compared with the standard metre.

The interferometer was also used in the famous Michelson–Morley experiment to determine whether the velocity of light was different in two directions at right angles to each other on the earth's surface. The negative results of this experiment gave support to the development of the theory of relativity.

(c) Measurement of small displacements – e.g. expansions or strains

For this wedge fringes are used, two glass plates being held apart to form a wedge, in such a way that the displacement to be measured causes the angle of the wedge to change. Fringes are formed in the wedge-shaped air film in the usual way. If the rod in Fig. 4.27 changes its diameter the angle of the wedge is altered and the fringes move. Very small changes may be measured by this means.

DIFFRACTION

4.20 Fresnel and Fraunhofer Diffraction

Diffraction effects may be divided into two groups:

(i) Fresnel diffraction. In this group the source of light and the screen on which the diffraction effects are observed are both at finite distances from the object at which the light is diffracted.

(ii) Fraunhofer diffraction. The source and screen are effectively at infinite distances from the obstacle – i.e. the light is parallel when it meets the object. In practice, this is accomplished by the use of lenses.

We deal with some cases of Fraunhofer diffraction first.

4.21 Fraunhofer Diffraction at a Single Slit

The experimental arrangement used to observe this effect is shown in Fig. 4.28. The source slit $S_1$ is to limit the width of the beam from the source and is placed at the focus of the convex lens $L_1$, so that a parallel beam of light falls on the slit $S_2$. The light passing through $S_2$ is then brought to a focus by $L_2$ on a screen placed in its focal plane.

The diffraction pattern formed on the screen is indicated in Fig. 4.29. If $S_2$ is removed a sharp image of $S_1$ is formed on the screen.

The positions of the maxima and minima in Fig. 4.29 may be de-
duced by using Huygens' principle, regarding every point in the slit $S_2$ as a source of secondary wavelets, and then adding these together to find the resultant intensity at any point on the screen. Fig. 4.30 shows

Fig. 4.28

![Diffraction pattern](image)

Geometrical image of $S_1$ in absence of $S_2$

Fig. 4.29

Fig. 4.30

the slit $S_2$ of width $d$ and the lens $L_2$ which focuses the light on to the screen.

Three points are chosen as secondary sources, one at each edge of the slit (A and C) and one in the middle (B). Wavelets spread out from these points, pass through $L_2$ and fall on the screen.

Consider first the direction perpendicular to the plane of $S_2$. In this
direction wavelets from A, B and C all travel the same distance before reaching the lens, and so arrive in phase at $P_1$, and produce a maximum.

Now consider parts of the wavelets travelling at an angle $\theta_1$ to $BP_1$, i.e. along AF, BG and CH. When these wavelets reach the lens the ones from B and C have travelled farther than the one from A by extra distances BD and CE respectively. Thus they are not, in general, in phase when they reach the lens. The intensity at $Q_1$, the point on the screen at which waves travelling in this direction are brought to a focus, therefore depends on the path differences BD and CE (which is twice BD).

Suppose that $BD = \lambda/2$, where $\lambda$ is the wavelength of the light used. Then the wavelets arriving at F and G are $\pi$ out of phase and interfere destructively at $Q_1$. Similarly, a wavelet from a point just below A cancels out one from a point just below B, and so on for pairs of points equidistant below A and B respectively. The whole slit may be covered in this way so that there is zero intensity at $Q_1$. $Q_1$ is thus the position of the first minimum on one side of the central maximum. It occurs when $BD = \lambda/2$,

i.e.

$$\sin \theta_1 = \frac{\lambda/2}{d/2} = \frac{\lambda}{d}.$$  

The second minimum occurs when $BD = \lambda$,

i.e.

$$\sin \theta_2 = \frac{2\lambda}{d}.$$  

This follows because BD is now the length that CE was for the first minimum. We may therefore apply the above argument to the top half of the slit and then similarly to the bottom half. Thus $\sin \theta = \lambda/d, 2\lambda/d, \ldots n\lambda/d, \ldots$ for minima. For the first minimum, at least, $\theta$ is usually small, so we may write $\theta_1 \simeq \lambda/d$ for first minimum.

To find the position of the first maximum to one side of the central maximum we divide the slit into three equal parts (Fig. 4.31). If $\theta$ is such that $LR = \lambda/2$ then $MS = \lambda$ and the secondary sources between A and M interfere to produce zero intensity. This leaves the sources between M and C, which produce a resultant intensity on the screen. In fact, this is not exactly the position of the maximum but is very close to it.

If the slit is divided into five equal parts and $\theta$ is such that there is again a path difference of $\lambda/2$ between waves from adjacent points, then the wavelets from the first four parts cancel, leaving $\frac{1}{5}$ of the slit to produce a resultant intensity on the screen. This value of $\theta$ gives a position close to that of the next maximum. It is less bright than the preceding one.
which has $\frac{1}{2}$ of the slit. A plot of intensity against $\sin \theta$ is shown in Fig. 4.32.

The angular distance between successive minima on the same side of the central maximum is $\lambda/d$ (if $\theta$ is small). The angular width of the central fringe is $2\lambda/d$, twice the width of the other fringes.

![Graph showing intensity versus $\sin \theta$](image)

Fig. 4.32

The linear width of the fringes on the screen depends on the focal length of $L_2$. If this is denoted by $f$, then the distance between successive minima $= f \frac{\lambda}{d}$.

The narrower the slit the wider the pattern. The pattern is also wider for light of longer wavelength, so that if white light is used the edges of the central fringe are reddish (long wavelength).

4.22 Diffraction by a Rectangular Aperture

In Section 4.21 it was assumed that the slit extended a relatively large distance perpendicular to the plane of the paper, so that diffraction effects in that direction did not need to be considered. If the length of the slit is reduced so that it is not much greater than its width (i.e. the slit becomes a rectangular aperture), then it is found that the bright fringes in the diffraction pattern are largely in the two directions parallel to the edges of the aperture; they consist mainly of two narrow patterns (each similar to that in Fig. 4.29) in the form of a cross.

4.23 Resolving Power

We have seen that to have a slit between the lenses $L_1$, $L_2$ in Fig. 4.28 changes the sharp image of $S_1$ into a diffraction pattern with ‘blurred edges’. Suppose that this lens system is used to form the image of two line sources $S_1$, $S_1'$ close together (Fig. 4.33).

On the screen two diffraction patterns appear, one of which is the
image of $S_1$ and the other of $S_1'$. In each case the centre of the pattern is on a line drawn through the appropriate source and the centre of $S_2$. If the angle $\alpha$ is large enough the central fringes of the two patterns do not overlap and it is possible to distinguish the two diffraction patterns, and

![Diagram of light sources and lens system](image)

Fig. 4.33

hence observe that there are two sources. If $S_1$, $S_1'$ are moved together the angle $\alpha$ decreases and the central fringes overlap. There is a value of $\alpha$ below which it is not possible to distinguish the two patterns. $S_1$, $S_1'$ are not then identifiable as separate sources and are said not to be resolved by the lens system. Fig. 4.34 shows the effect on the diffraction pattern on the screen as $\alpha$ is decreased.

![Graphs showing intensity vs position](image)

(a) $S_1$, $S_1'$ well resolved  
(b) $S_1$, $S_1'$ just resolved  
(c) $S_1$, $S_1'$ not resolved

Fig. 4.34

It was stated by Rayleigh that if the central maximum of one pattern falls in the same place as the first minimum of the other pattern (as in Fig. 4.34 (b)), then it is just possible to resolve the two sources; he suggested that this be adopted as a test of whether two objects would be resolved by a given instrument. This test is known as Rayleigh's criterion.

It has been shown that the position of the first minimum for a slit or
rectangular aperture is at an angle \( \theta \) to the forward direction, where \( \theta = \frac{\lambda}{d} \) (\( d \) is the dimension of the aperture parallel to the line joining \( S_1, S_1' \)). Rayleigh's criterion gives \( \alpha = \theta \) when \( S_1, S_1' \) are just resolved. Thus the minimum angular separation of two objects which are to be resolved by the lens system is

\[
\alpha_{\text{min.}} = \frac{\lambda}{d} \tag{4.8}
\]

\( \alpha_{\text{min.}} \) is a measure of the resolving power of the lens system.


Diffraction at a circular aperture is more difficult to treat theoretically, but it can be shown that the diffraction pattern consists of concentric circles and that a radius of the smallest dark ring subtends an angle \( \theta \) at the centre of the aperture, given by \( \sin \theta = 1.22 \frac{\lambda}{a} \), where \( a \) is the diameter of the aperture.

The lens \( L_2 \) in Fig. 4.33 may be regarded as the objective of a telescope, the slit \( S_2 \) being replaced by a circular aperture of diameter \( a \). The minimum angle which two distant objects (e.g. stars) may subtend if they are to be resolved is then given by

\[
\alpha_{\text{min.}} = 1.22 \frac{\lambda}{a} \tag{4.9}
\]

this being derived in a similar way to equation (4.8).

The only way, then, to improve the resolving power of a telescope is to increase the aperture, and this is one of the advantages of making telescope objectives of very large diameter. It must be emphasized that equation (4.9) gives the theoretical best performance, and this will be achieved only if the lens system is optically perfect.

4.25 Resolving Power of a Microscope

Here we are dealing with objects at finite distances, so that it is not a case of Fraunhofer diffraction, as is the telescope, where the objects are very distant. Fig. 4.35 shows the objective \( AB \) of a microscope.

\( O \) is an object on the axis of the lens and \( O' \) an object a small distance

![Fig. 4.35](image-url)
off the axis; the angle subtended by AB at O is denoted by $2i$. The
centre of the diffraction pattern due to O is at P and the centre of that
due to O' is at $P'$; P and $P'$ are
the respective image points. If
we apply Rayleigh's criterion,
then for O and O' just to be re-
solved the first minimum of the
pattern due to O' must fall at P.
If the path difference between
waves from O' diffracted along
O'AP and O'BP is $\lambda$, then a
wave which travels along O'CP
arrives at P $\pi$ out of phase with
one that goes via A, and they
cancel out. All waves from O' can be paired off in this way, so that P is
the position of the first minimum in the diffraction pattern due to O' if

$$O'AP - O'BP = \lambda,$$

i.e. if

$$O'A - O'B = \lambda, \text{ as } AP = BP.$$

If $s$ is the distance OO' the geometry of Fig. 4.36 gives:

$$O'A = OA + s \sin i,$$

and

$$O'B = OB - s \sin i.$$

But

$$OA = OB$$

.$$O'A - O'B = 2s \sin i,$$

.$$s = \frac{\lambda}{2 \sin i}.$$

$s$ is the minimum value of the separation of O and O' for which they
are resolved. If the space between O and the objective is filled with a
liquid of refractive index $\mu$ the actual paths must be multiplied by $\mu$
to give the optical paths, and the condition becomes

$$s = \frac{\lambda}{2\mu \sin i} \quad \ldots \quad \ldots \quad (4.10)$$

The above argument requires two assumptions that are not true in
practice:

(i) that the aperture is rectangular;
(ii) that O, O' are independent sources of light.

In practice, the aperture is circular, and O, O' are illuminated by a
condenser lens so that there is a phase relation between light from the
two points.

A treatment taking into account these two factors was carried out by
Abbe, and he showed that equation (4.10) still holds, i.e. the two
factors tend to cancel out.
The quantity $\mu \sin i$ is often called the numerical aperture of the microscope and needs to be as large as possible for good resolving power. One way of increasing it is to have oil in the space between the object and the objective so that $\mu$ is increased.

The resolving power may also be improved by decreasing $\lambda$. This has been done by using ultra-violet light and X-rays, the image being recorded photographically.

Even higher resolution is obtained in the electron microscope. It is shown in Chapter 10 that electrons may be diffracted and behave as if they are waves of very short wavelength (e.g. 1 Å).

### 4.26 Diffraction Grating

The student will already be familiar with the action of a diffraction grating and how it is used in a spectrometer to measure the wavelength of light. We shall now investigate the nature of the diffraction pattern in more detail. We consider only the case in which the incident light is perpendicular to the plane of the grating.

![Diffraction Grating Diagram](image)

Suppose that the grating consists of very narrow slits, the centres of successive ones being a distance $d$ apart. Suppose that there are $N$ spaces, i.e. $N + 1$ slits. In practice, $N$ is so large that the difference between $N$ and $N + 1$ is negligible. It will be clear as the argument proceeds why we choose $N$ as the number of spaces and not as the number of slits.

Each slit may be regarded as a source of secondary wavelets spreading out in all directions. In the direction perpendicular to the plane of the grating all the wavelets arrive at the lens in phase so that a bright image is formed. This is known as the zero order spectrum.

Now consider the direction at an angle $\theta$ with the normal to the grating such that there is a path difference of $\lambda$ between the wavelets from the first and last slits. The wavelet from the first slit is now $\pi$ out of phase with that from the centre one when they reach the lens, and these two interfere to produce zero amplitude. All the slits may be paired off
in this way, and so the total amplitude in this direction is zero. (Note that this is similar to the argument used for the single slit.) The angle \( \theta \) thus gives zero intensity in the diffraction pattern, and is therefore the angular half-width of the central fringe. From the geometry of Fig. 4.37,

\[
\sin \theta = \frac{\lambda}{Nd'}
\]

\[
= \frac{\lambda}{b'}
\]

where \( b \) is the total width of the grating. This is a very small angle, so the fringe is very sharp; e.g. for a grating of width 2 cm, and for \( \lambda = 5,000 \text{ Å} \):

\[
\sin \theta = 2.5 \times 10^{-5}
\]

\[
\theta = 2.5 \times 10^{-5} \text{ radians}
\]

\[
= 0.086 \text{ minutes of arc}.
\]

Notice that the width of the central fringe does not depend on the number of lines on the grating, but only on its total width. It would be the same for two slits separated by a distance \( b \) (i.e. \( N = 1 \)) or for a single slit of width \( b \) (i.e. \( N = \infty \)).

Suppose now that \( \theta \) is increased until the distance AB (Fig. 4.37) becomes \( 3\lambda/2 \). By dividing the slits into three equal groups, it will be seen that wavelets from the first two of these groups combine to give zero intensity (because the path difference between the first and last is \( \lambda \)) and those from the third group have a non-zero resultant. This, then, is near a maximum in the diffraction pattern. However, the intensity of this maximum is very low compared with that of the central one, because even the wavelets from the one-third of the grating that is operative are not in phase, and their resultant is nowhere near as large as their algebraic sum. As \( \theta \) is increased there is a series of minima and subsidiary maxima until AB becomes equal to \( N\lambda \). When this occurs wavelets from all slits arrive in phase at the lens and a very bright maximum is produced. This is known as the first-order spectrum.

The condition for the formation of this fringe is

\[
\sin \theta_1 = \frac{N\lambda}{Nd'} = \frac{\lambda}{d'}.
\]

This fringe also is very narrow, because when AB becomes \( (N\lambda + \lambda) \) another minimum occurs.

Fig. 4.38 shows a plot of the intensity against \( \sin \theta \) for a grating of 11 slits (\( N = 10 \)). As the number of lines on the grating is increased the number of subsidiary maxima \( (N - 2) \) between successive orders increases, and the intensity of each one relative to the principal maxima decreases. For the gratings used in spectroscopy (e.g. \( N = 20,000 \)) the subsidiary maxima are too faint to be seen.
The condition for the formation of the \( m \)th order spectrum is

\[
AB = mN\lambda,
\]
i.e.

\[
\sin \theta_m = \frac{m\lambda}{d}.
\]

The next minimum occurs when \( AB = mN\lambda + \lambda \).

\[\begin{array}{cc}
-\frac{10\lambda}{b} & \frac{-\lambda}{b} \\
\text{first order} & \text{zero order} \\
\frac{\lambda}{b} & \frac{10\lambda}{b} \\
& \text{first order}
\end{array}\]

Fig. 4.38

The angular half-width of the \( m \)th order is given by

\[
\delta \theta_m = \frac{\lambda}{Nd \cos \theta_m} = \frac{\lambda}{b \cos \theta_m}.
\]

This may be verified by reference to Fig. 4.39.

**Overlapping of orders**

The \( m \)th order spectral line of wavelength \( \lambda_1 \) is formed at the same angle as the \((m + 1)\)th order spectral line of wavelength \( \lambda_2 \) if

\[
m\lambda_1 = (m + 1)\lambda_2.
\]

If the orders are sufficiently high, then both \( \lambda_1 \) and \( \lambda_2 \) may be in the visible spectrum. Suppose that \( \lambda_1 \) is at the red end of the spectrum (i.e. say \( \lambda_1 = 7,000 \text{ Å} \)), then for overlapping:

<table>
<thead>
<tr>
<th>( m )</th>
<th>( \lambda_1 ) (Å)</th>
<th>( m + 1 )</th>
<th>( \lambda_2 ) (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7,000</td>
<td>2</td>
<td>3,500 not visible</td>
</tr>
<tr>
<td>2</td>
<td>7,000</td>
<td>3</td>
<td>4,700 blue</td>
</tr>
<tr>
<td>3</td>
<td>7,000</td>
<td>4</td>
<td>5,250 green</td>
</tr>
</tbody>
</table>

There is no overlapping in the first order, but the red end of the second order overlaps with the blue end of the third order. Overlapping between the third and fourth orders is more pronounced, as the third order red coincides with the fourth order green. If it is desired to examine spectra at higher orders it is usually necessary to use filters to cut out the unwanted wavelengths from adjacent orders.

*
4.27 The Resolving Power of a Diffraction Grating

In Fig. 4.39 the solid lines are in the direction in which the mth order spectrum for light of wavelength \( \lambda \) occurs, so that \( AB = mN\lambda \).

The dotted lines are in the direction of the mth order spectrum for light of wavelength \( \lambda + \delta \lambda \), so that \( AC = mN(\lambda + \delta \lambda) \).

Now if the spectral lines formed by \( \lambda \) and \( \lambda + \delta \lambda \) are just to be resolved the direction \( AC \) must be that of the first minimum after the mth order spectrum for light of wavelength \( \lambda \) (Rayleigh’s criterion). Thus \( AC = mN\lambda + \lambda \).

Therefore

\[
mN(\lambda + \delta \lambda) = mN\lambda + \lambda,
\]

\[
\therefore \frac{\lambda}{\delta \lambda} = mN. \quad \quad (4.11)
\]

The quantity on the left of this equation is called the resolving power of the grating.

Equation (4.11) shows that the resolving power of the grating may be increased either by increasing the number of slits or by working at a higher-order spectrum. However, for a grating of fixed total width the quantities \( m \) and \( N \) cannot be varied independently.

The angle \( \theta_m \) of the mth order spectrum for light of wavelength \( \lambda \) is given by \( \sin \theta_m = m\lambda/d \).

Thus the mth order cannot be observed unless \( m\lambda/d < 1 \), as \( \sin \theta_m \) must be less than one. Now if \( N \) is increased \( d \) is decreased, so that the highest value of \( m \) obtainable is also reduced.

Putting \( Nd = b \) (the width of the grating) gives

\[
\frac{mN\lambda}{b} < 1,
\]

i.e.

\[
mN < \frac{b}{\lambda},
\]

which shows that the maximum value of \( mN \) (resolving power) obtainable at a given wavelength is determined solely by the width of the grating and not by the number of lines on it. The maximum resolving power is obtained for \( \theta \) near 90°; if the number of lines on the grating is small the order being observed is high and if the number of lines is large the order is low, but in each case the product \( mN \approx b/\lambda \).

One advantage of having a large number of lines is that one obtains the maximum resolving power at low orders where the problem of overlapping of orders is not serious.
4.28 Echelon Gratings

To obtain very high resolving powers it is necessary to devise some means of increasing either \( m \) or \( N \) without decreasing the other. This has been accomplished in the reflection echelon, which is designed to increase the value of \( m \). Fig. 4.40 shows the design of the grating.

The path difference between successive rays is accomplished by 'stepping' the grating. It can be arranged that the path difference between rays reflected from successive steps is, say, 10,000 wavelengths at a small angle of reflection. This means that the spectrum formed at this angle is the 10,000th order. \( m \) is large, and therefore the resolving power is large.

The problem of overlapping of orders is, of course, very acute when working at such a high order. For example the spectral line of wavelength 5,000 Å in the 10,001th order overlaps with that of 5,000·5 Å in the 10,000th order. It is therefore necessary to remove all wavelengths except those between, say, 4,999·6 and 5,000·4 Å before the light reaches the echelon. This is usually done by first forming a spectrum of the source with a high-resolution diffraction grating and then selecting the portion required by means of a narrow slit placed in this spectrum. The light coming through this slit is then reflected from the echelon and the high-resolution spectrum obtained.

4.29 Fresnel Diffraction. Half-period Zones

The theory of Fresnel diffraction is much more difficult than that of Fraunhofer diffraction; it is not intended to enter into this in any detail, but simply to indicate the lines of approach and to quote some of the results obtained.
S (Fig. 4.41) represents part of a spherical wave-front; it is desired to find the effect of this wave at the point P. To tackle this problem the wave-front is first divided up into annular zones centred on O such that the outer edge of each zone is $\lambda/2$ farther from P than the outer edge of the previous zone. Each of these zones is then regarded as a source of secondary wavelets spreading out in all forward directions. We are concerned with the wavelets arriving at P; because of the way in which the zones have been constructed, the resultant wavelet at P from any zone is $\pi$ out of phase with those from each of the two adjacent zones—hence the name half-period zones.

The resultant amplitude at P is found by adding together the amplitudes of the wavelets from all the zones. If the magnitudes of the resultsants from successive zones are denoted by $R_1, R_2, R_3, \ldots$, then the total amplitude at P is $R = R_1 - R_2 + R_3 - R_4 \ldots$. The alternate positive and negative signs take into account the successive changes of $\pi$ in phase.

To perform the addition it is necessary to know the relative values of $R_1, R_2, R_3, \ldots$, and each of these depends on three factors:

(i) the area of the zone;
(ii) the distance of the zone from P;
(iii) the obliquity factor of the secondary wavelets (see Section 4.2).

We will consider each of these factors separately.

(i) A rough value for the area of the $n$th zone may be found by making certain approximations. Suppose that the radius of curvature of the wave-front is very large, so that we may take ON (Fig. 4.41) as straight and perpendicular to OP. Then the radius, $r_n$, of the outer edge of the $n$th zone is given by

$$r_n^2 = ON^2 = NP^2 - OP^2$$
$$= \left( d + \frac{n\lambda}{2} \right)^2 - d^2$$
$$= n\lambda d + \frac{n^2\lambda}{4}.$$

If $n$ is not large then, because $\lambda$ is very small, the second term may be neglected in comparison with the first.

Hence

$$r_n^2 = n\lambda d \quad \ldots \ldots \quad (4.12)$$

Similarly,

$$r_{n-1}^2 = (n - 1)\lambda d.$$

Now the area of the $n$th zone is given by,

$$A_n \approx \pi(r_n^2 - r_{n-1}^2),$$

as the zone is approximately planar,

$$\approx \pi[n\lambda d - (n - 1)\lambda d],$$

$$\approx \pi n\lambda d.$$

To a first approximation, then, the area of the zones is independent of $n$. 
A full treatment shows that \( A_n \) increases slowly with \( n \).

(ii) The amplitude of a spherical wavelet at distance \( d \) from its source is proportional to \( 1/d \) (the intensity is proportional to \( 1/d^2 \)), so that for the \( n \)th zone:

\[
R_n \propto \frac{1}{d + n\lambda/2}.
\]

It may be shown that the decrease in \( R_n \) as \( n \) increases, due to this effect, is exactly cancelled by the increase in the size of the zones mentioned in (i).

(iii) As factors (i) and (ii) have cancelled out, any difference between the amplitudes of the wavelets from successive zones is due entirely to the obliquity factor. For the wavelets from the \( n \)th zone this factor is \( 1 + \cos \theta_n \), where \( \theta_n \) is the angle shown in Fig. 4.41. As \( n \) increases, \( \theta_n \) increases so that \( 1 + \cos \theta_n \) decreases, becoming zero when \( \theta_n = \pi \).

We have, then, that \( R_n \) decreases slowly to zero as \( n \) increases.

To find the resultant at \( P \) the vector method of addition may be used.

Consider the first zone split up into several sub-zones. Suppose, for example, it is split up into ten such sub-zones, so that the outer edges of

![Fig. 4.42](image)

successive sub-zones are \( \lambda/20 \) farther from \( P \). All sub-zones have approximately the same area, and the wavelet from each is \( \pi/10 \) out of phase with that from the adjacent one. The vector diagram for the amplitude of the disturbance at \( P \) then takes the form shown in Fig. 4.42 (a). The resultant is \( R_1 \). If the number of sub-zones is made to tend to infinity, then Fig. 4.42 (b) is obtained. When the second zone is included the vector diagram is as in Fig. 4.42 (c). The vector diagram does not quite close because of the obliquity factor.

The resultant amplitude due to the second zone is \( R_2 \), and the resultant of the first two zones is \( R_1 - R_2 \), which is almost zero. Fig. 4.42 (d) shows the effect of including several zones.
When the whole wave-front is included the end of the spiral reaches the centre C and the resultant, \( R = R/2 \). This means that the amplitude at P due to the whole wave-front is one-half that due to the first zone.

### 4.30 Fresnel Diffraction at a Circular Aperture

If P is chosen such that the aperture has the same radius as the first half-period zone the amplitude of the resultant at P is \( R_1 \), which is twice that due to the unobstructed wave (see previous section). The intensity at P is thus four times that due to the unobstructed wave. It is interesting to find a rough value of the radius of the aperture which will cause this effect.

Suppose \( d = 100 \text{ cm} \), and \( \lambda = 5,000 \text{ Å} \).

Then from Fig. 4.43 and equation (4.12),
\[
R_1^2 = \lambda d.
\]

Substituting numerical values:
\[
R_1^2 = 5 \cdot 10^{-5} \cdot 100
= 5 \times 10^{-3},
\]
\[
\therefore R_1 \simeq 7 \times 10^{-2} \text{ cm, i.e. approximately 0.7 mm.}
\]

If the aperture is now widened to include the first two half-period zones, then \( R = R_1 - r_2 \), which is almost zero as \( R_1 \simeq R_2 \) (see Fig. 4.42). The resultant intensity at P is thus almost zero. P becomes alternately dark and bright as the aperture is widened to include successive zones; for an odd number of zones P is bright and for an even number P is dark. It will be seen from Fig. 4.42 that this does not apply when \( n \) becomes large.

Another way of gradually including more zones is to keep the size of the aperture fixed and move P. As P moves towards the aperture the radii of the zones decrease (equation (4.12)) so that more are included, and P is alternately bright and dark.
4.31 Fresnel Diffraction at a Circular Obstacle

A circular obstacle placed in the path of a spherical wave cuts out some of the central zones, i.e. those which are represented by the outer parts of the spiral in Fig. 4.42 (d). The resultant obtained when these are removed is indicated in Fig. 4.44. This shows that, provided the number of zones covered by the obstacle is not large, there is always an appreciable resultant amplitude at P; in other words, there is a bright spot at the centre of the shadow of the circular obstacle.

![Diagram of Fresnel Diffraction](image)

Fig. 4.44

It is fairly easy to observe this central spot in the laboratory, provided an intense point source of light is used. Its presence is one of the most conclusive pieces of evidence in favour of the wave theory in general, and Huygens’ principle in particular.

4.32 Zone Plate

A zone plate (Fig. 4.45) is designed to remove alternate half-period zones from an incident wave-front. It is usually made by photographing the negative of the required pattern drawn on a white card in black ink. The photographic plate then consists of alternate transparent and opaque zones.

It is necessary that the radii of successive rings should be in the ratio of the squareroots of the integers, i.e. $1 : \sqrt{2} : \sqrt{3}$. That this is so may be seen by reference to equation (4.12), which shows that

$$r_n^2 \propto n,$$

i.e.

$$r_n \propto n^{1/2}.$$

At a point P on the axis of the zone plate, such that the rings do cover successive half-period zones, the amplitude of the wave is

$$R = R_2 + R_4 + R_6 + \ldots ,$$

assuming that the central ring is dark. This value of $R$ is very large, as all the terms are positive, and a very bright spot of light is formed at P.
The plate is, under these conditions, acting in a similar way to a convex lens. The ‘focal length’ of the plate is the distance of $P$ from the plate, which is given by equation (4.12);

$$r_1^2 = d\lambda,$$

$$\therefore d = \frac{r_1^2}{\lambda}.$$

It is possible to use a zone plate to form an image of an illuminated object.

There are also subsidiary foci at points such that the rings each cover 3 or 5 or 7 . . . half-period zones,

i.e. for

$$d = \frac{r_1^2}{3\lambda}, \frac{r_1^2}{5\lambda}, \text{etc.}$$

### 4.33 Cylindrical Wave-fronts

If a line source of light is used, then the wave-fronts are cylindrical and the method of division is into half-period strips (Fig. 4.46).

![Diagram of cylindrical wavefronts](image)

The resultant amplitude at $P$ due to a number of strips is found by the method of vector addition as for circular zones, but two fundamental differences result in a rather different vector diagram in this case. First, the area of the zones decreases rapidly away from the axis, so that the shape of the spiral is altered. Secondly, there are two zones at a given distance from $P$, one above and one below the axis, which results in two spirals joined end to end. The complete diagram is known as a Cornu Spiral (Fig. 4.47). The lettering in this figure corresponds to that in Fig. 4.46, so that, for example, the part of the spiral from $A$ to $B_1$ shows the effect of the first strip above the axis.
The resultant amplitude at P due to a number of strips may easily be found from the Cornu Spiral. Suppose one wished to find the resultant effect of all the strips between $C_1$ and $B_2$. The resultant amplitude is the line joining $C_1$ and $B_2$ on the Cornu Spiral in Fig. 4.47. Similarly, the resultant of all the wave-front above the axis is the line $AZ_1$.

4.34 Fresnel Diffraction at a Straight Edge

In Fig. 4.48 $S$ is a line source of light perpendicular to the plane the paper. One edge of the obstacle $DA$ is parallel to $S$. A shadow of the obstacle is formed on the screen. Treated geometrically, $O$ is the edge of the shadow, $P$ is fully illuminated and $Q$ is in shadow. To find the effect of diffraction we divide the wave-front into strips as in the
previous section, and find the resultant effect on the screen by the method of vector addition, using the Cornu Spiral.

To find the effect at P, the strips are drawn on either side of A₁, the point where SP intersects the wave-front. If P is a long way from O, then almost all the strips contribute to the resultant amplitude, which is thus Z₁Z₂ in Fig. 4.49. As P moves towards O the obstacle gradually obscures some of the half-period strips from one extreme of the wave-front.

These are the ones represented by the centre of the bottom spiral so that the lower end of the resultant in Fig. 4.49 moves out from Z₂ round the spiral. As it does so its length goes through successive maxima and minima so that the amplitude (and hence the intensity) of the disturbance at P fluctuates in a similar way. For example, the tail of the resultant goes successively through D₂, C₂, B₂, and Z₁D₂ > Z₁C₂ < Z₁B₂. Fringes are therefore observed in the region outside the geometrical shadow.

At O half the wave-front is obscured, so that the resultant amplitude is AZ₁, which is one-half the amplitude due to the whole wave-front (Z₂Z₁). Thus the intensity at O is one-quarter that due to the unobstructed wave.

As the point of observation moves towards Q strips in the other half of the wave-front are obscured, so that the tail of the resultant vector moves round the upper spiral. This results in a steady decrease in amplitude (Z₁A > Z₁B₁ > Z₁C₁), so that the illumination falls off steadily inside the geometrical shadow and there are no fringes.
Fig. 4.50 shows a graph of the intensity observed on the screen. Fresnel diffraction at a single slit and a thin wire may also be investigated theoretically by the use of the Cornu Spiral. For details of the method used the reader should consult a more advanced text.

![Graph of intensity](image)

**Fig. 4.50**

### 4.35 Rectilinear Propagation

The question arises how we are to reconcile the wave theory of light and diffraction with the common observation that 'light travels in straight lines'. There is no real difficulty here when one realizes that diffraction effects are confined to a region very close to the edge of the shadow. In a typical case all the observable parts of the diffraction pattern due to a straight edge, discussed in the previous section, are within 1 mm of the edge of the shadow. It is not surprising that, unless one sets out deliberately to observe diffraction effects, they are not normally noticed, and rectilinear propagation appears to take place.

### QUESTIONS

1. What conditions must be fulfilled if interference effects with visible light are to be observed?
2. Summarize evidence for the wave nature of light.
3. What evidence is there that light incident on the boundary between two media from the side of smaller refractive index, undergoes a phase change of $\pi$ on reflection?
4. Describe and account for the appearance of thin films viewed by reflection, using an extended source of white light. Calculate the thicknesses of film which produce the various effects you describe.
5. Classify all the interference effects you know of, under the headings given in Section 4.18.
6. It is shown in Section 4.27 that the resolving power of a diffraction grating does not depend on the number of lines per centimetre on the grating. In view of this, discuss the reasons for making gratings with very large numbers of lines per centimetre.

7. What effect does the finite width of the grating slits have on the spectra produced?

8. Use the Cornu Spiral to predict the Fresnel diffraction patterns due to: (i) a thin wire; (ii) a narrow slit.

ESSAY TOPICS

9. Interferometry.

10. High-resolution spectroscopy.
HEAT TRANSFER

CONDUCTION

5.1 Coefficient of Thermal Conductivity

Suppose that a thin slab of material with parallel faces each of area $A$, distance $\delta x$ apart, has these faces maintained at a temperature difference $\delta \theta$. It has been shown experimentally that when a steady state has been reached the rate of flow of heat, $dQ/dt$, between these two faces is given by

$$\frac{dQ}{dt} \propto A \frac{\delta \theta}{\delta x}.$$

Introducing a constant of proportionality into this equation and letting $\delta x \to 0$ gives

$$\frac{dQ}{dt} = -kA \frac{d\theta}{dx} \quad \quad \quad \quad (5.1)$$

the minus sign being introduced so that $Q$, $\theta$ and $x$ are all measured in the same direction; if the heat is to flow in the direction of $x$, then $\theta$ must be decreasing in that direction, i.e. $d\theta/dx$ is negative. $d\theta/dx$ is known as the temperature gradient. $k$ is the coefficient of thermal conductivity, and may be defined in words thus:

The coefficient of thermal conductivity of a substance is the rate at which heat flows across unit area, drawn perpendicular to the direction of flow when there is unit temperature gradient in the substance.

Equation (5.1) may be rewritten

$$k = \frac{-dQ/dt}{A \frac{d\theta}{dx}},$$

from which it is seen that the units of $k$ on the c.g.s. system are cal sec$^{-1}$ cm$^{-1}$ deg C$^{-1}$ or watt cm$^{-1}$ deg C$^{-1}$.

It is not intended to discuss here any of the methods of determining $k$; these are described in the 'A' level theory and practical books. For refinements of these methods the reader should consult one of the books mentioned in the bibliography.

5.2 The Wiedemann–Franz Law

This law was based on experimental results and relates the thermal and electrical conductivities of metals. It states that the ratio of the coefficient of thermal conductivity ($k$) to the electrical conductivity ($\sigma$) is approximately the same for all metals at the same temperature.

It was later shown by Lorentz that the value of this ratio alters as the
temperature changes and is approximately proportional to the absolute temperature $T^\circ K$,

i.e.

$$\frac{k}{\sigma T} = \text{constant.}$$

This law may be shown to be a consequence of the theory that a metal contains a number of 'free' electrons which are responsible both for electrical and thermal conduction (Drude's theory of conduction – see Section 13.2).

### 5.3 Flow of Heat along an Unlagged Bar

The flow of heat along a lagged bar of uniform cross-section is easy to treat theoretically and is dealt with in more elementary books. If it is assumed that no heat leaves the sides of the bar, then in the steady state the rate of flow of heat across any section must be the same, so that the temperature gradient is the same at all points in the bar. This means that the temperature varies linearly along the bar (Fig. 5.1).

![Fig. 5.1](image1)

For an unlagged bar heat is lost from its sides, so that less heat flows through successive sections (Fig. 5.2).

As the rate of flow of heat through the bar is least at the cool end, it follows from equation (5.1) that the temperature gradient is less at this end than at the hot end. The graph of $\theta$ against $x$ must therefore be such that $- \frac{d\theta}{dx}$ decreases as $x$ increases; it has the general shape indicated in Fig. 5.3.

![Fig. 5.2](image2)

The exact nature of the curve depends on the relation between the amount of heat lost from the sides of the bar and that conducted along the bar. Suppose that the rate of loss of
heat from the surface is proportional to the excess temperature of the surface over its surroundings which are at temperature \( \theta_0 \). Denote by \( \varepsilon \) the amount of heat lost from unit area of surface in 1 sec when the excess temperature is 1°C.

Consider a bar of circular cross-section of radius \( r \), with one end maintained at temperature \( \theta_1 \) (Fig. 5.4). Suppose the bar is long, so that the unheated end is at the same temperature, \( \theta_0 \), as the surroundings. Let the temperature at distance \( x \) from the hot end be \( \theta \). Consider a slab of the bar of thickness \( \delta x \) at this point, and let the temperature difference across its faces be \( \delta \theta \).

The amount of heat entering the slab per sec is, from equation (5.1),

\[
\left( \frac{dQ}{dt} \right)_x = -k\pi r^2 \left( \frac{d\theta}{dx} \right)_x.
\]

where \( \left( \frac{d\theta}{dx} \right)_x \) denotes the value of \( d\theta/dx \) at distance \( x \). The rate at which heat is conducted out through the other face is

\[
\left( \frac{dQ}{dt} \right)_{x+\delta x} = -k\pi r^2 \left( \frac{d\theta}{dx} \right)_{x+\delta x},
\]

and the rate at which heat is lost from the sides is

\[
2\pi r \delta x \varepsilon (\theta - \theta_0),
\]

so that in the steady state

\[
-k\pi r^2 \left( \frac{d\theta}{dx} \right)_x = -k\pi r^2 \left( \frac{d\theta}{dx} \right)_{x+\delta x} + 2\pi r \delta x \varepsilon (\theta - \theta_0).
\]

Now

\[
\left( \frac{d\theta}{dx} \right)_{x+\delta x} \approx \left( \frac{d\theta}{dx} \right)_x + \left( \frac{d^2\theta}{dx^2} \right) \delta x.
\]

Therefore

\[
\frac{d^2\theta}{dx^2} - \frac{2\varepsilon}{kr} (\theta - \theta_0) = 0.
\]

To solve this equation put \( (\theta - \theta_0) = \phi \),

\[
\therefore \frac{d^2\phi}{dx^2} - \frac{2\varepsilon}{kr} \phi = 0.
\]

This has a solution of the form

\[
\phi = Ae^{mx} + Be^{mx}.
\]
where $m_1$, $m_2$ are the roots of the equation

$$m^2 - \frac{2\varepsilon}{kr} = 0$$

$$\therefore m = \pm \left(\frac{2\varepsilon}{kr}\right)^{\frac{1}{2}}$$

$$\therefore \phi = Ae^{\left(\frac{2\varepsilon}{kr}\right)^{\frac{1}{2}} \cdot x} + Be^{-\left(\frac{2\varepsilon}{kr}\right)^{\frac{1}{2}} \cdot x}.$$

At $x = 0$, $\theta = \theta_1$; i.e. $\phi = \theta_1 - \theta_0$.

$$\therefore A + B = \theta_1 - \theta_0$$

as $x \to \infty$, $\theta \to \theta_0$; i.e. $\phi \to 0$.

$$\therefore A = 0.$$

$$\therefore \theta - \theta_0 = (\theta_1 - \theta_0)e^{-\left(\frac{2\varepsilon}{kr}\right)^{\frac{1}{2}} \cdot x},$$

$$\therefore \ln \left(\frac{\theta_1 - \theta_0}{\theta - \theta_0}\right) = \left(\frac{2\varepsilon}{kr}\right)^{\frac{1}{2}} \cdot x \quad \ldots \quad (5.2)$$

If two long rods of different materials, but similar in all other respects, including the nature of the surfaces, are both heated to the same temperature $\theta_1$ at one end, then the distance along each at which the temperature has some value $\theta$ may be determined from equation (5.2). Let the thermal conductivities of the rods be $k_1$ and $k_2$, and the distances from the heated end at which the temperature is $\theta$ be $l_1$ and $l_2$ respectively. From equation (5.2)

$$\ln \left(\frac{\theta_1 - \theta_0}{\theta - \theta_0}\right) = \left(\frac{2\varepsilon}{k_1r}\right)^{\frac{1}{2}} \cdot l_1,$$

and

$$\ln \left(\frac{\theta_1 - \theta_0}{\theta - \theta_0}\right) = \left(\frac{2\varepsilon}{k_2r}\right)^{\frac{1}{2}} \cdot l_2,$$

$$\therefore \frac{k_1}{k_2} = \frac{l_1^2}{l_2^2}.$$

This result is the basis of an experiment sometimes used to obtain a rough comparison of the coefficients of thermal conductivity of different materials in the form of rods. One method of determining the lengths $l_1$ and $l_2$ is to coat the rods with wax of suitable melting point and find the boundary between the liquid and solid wax. This boundary is at the same temperature on each of the rods, and hence the above equation may be applied. The ends of the rods may be heated in a water bath, and the extent of melting may be made apparent by sticking metal spheres in the wax; these drop off when the wax melts.

5.4 Conduction through Composite Slabs

Fig. 5.5 shows two slabs X and Y of different materials and thicknesses, but of the same area of cross-section $A$, in perfect thermal con-
tact with one another. The outer faces are maintained at temperatures \( \theta_1 \) and \( \theta_2 \) respectively (\( \theta_1 > \theta_2 \)). The problem is to find out at what rate heat is conducted through the slabs.

The guiding principle in problems of this type is that the same rate of flow of heat takes place in both X and Y in the steady state. This must be so, otherwise there would be an accumulation or diminution of heat energy at the boundary with a consequent change in temperature.

Let the temperature of the interface be \( \theta \). Denote the thicknesses of X and Y by \( d_x \) and \( d_y \); and their thermal conductivities by \( k_x \) and \( k_y \) respectively.

Then
\[
\frac{dQ}{dt} = \frac{k_x A(\theta_1 - \theta)}{d_x} = \frac{k_y A(\theta - \theta_2)}{d_y}.
\]

The value of \( \theta \) may be found from the second equality, and substitution of this value of \( \theta \) in the first will give the required value of \( dQ/dt \).

This method may be extended to any number of slabs ‘in series’.

A neater solution may often be obtained, particularly in the case of several slabs, by the use of an electrical analogy. The equation for heat conduction is
\[
\frac{dQ}{dt} = \frac{kA}{d}(\theta_1 - \theta_2),
\]
and for electrical conduction is
\[
I = \frac{V}{R}.
\]

But \( I = dq/dt \) where \( q \) denotes electric charge.

If the conductor is in the form of a wire of length \( l \) and area of cross-section \( A \), then
\[
R = \frac{\rho l}{A} = \frac{l}{\sigma A}
\]
where \( \rho \) is the resistivity and \( \sigma \) the electrical conductivity of the material.

Thus
\[
\frac{dq}{dt} = \frac{\sigma A}{l} V.
\]

The analogy is now obvious. The ‘heat current’ \( (dQ/dt) \) is analogous to the electric current \( (dq/dt) \), the ‘thermal resistance’ \( (d/kA) \) to the electrical resistance \( (l/\sigma A) \) and the temperature difference \( (\theta_1 - \theta_2) \) to the potential difference \( (V) \). ‘Thermal resistances’ in series or parallel
are then treated in the same manner as electrical resistance and ‘Ohm’s law’ is applied.

In the previous example one would proceed as follows:

thermal resistance of \[ X = \frac{d_x}{k_x A} \]

thermal resistance of \[ Y = \frac{d_y}{k_y A} \]

\[ \therefore \text{ total thermal resistance of } X \text{ and } Y \text{ in series } = \frac{d_x}{k_x A} + \frac{d_y}{k_y A} \]

\[ \therefore \text{ heat current } \frac{dQ}{dt} = \frac{\theta_1 - \theta_2}{\frac{d_x}{k_x A} + \frac{d_y}{k_y A}} \]

The reader should check that this equation is the same as that obtained by the elimination of \( \theta \) from equation (5.3).

If the temperature of the interface is required it may be found by applying ‘Ohm’s law’ to the first slab, and hence finding the temperature difference across it.

In a simple example, such as the one above, there is only a slight saving in labour obtained by using this method, but in more complicated examples considerable saving is accomplished.

5.5 Conduction through the Walls of Pipes

An approximate expression for the rate of flow of heat through a thin-walled pipe may easily be found. Suppose that the pipe has internal and external radii \( r_1 \) and \( r_2 \) respectively, where \( (r_2 - r_1) \ll r_1 \). Let the pipe be of length \( l \), and the material of which it is made have a coefficient of thermal conductivity \( k \). We require to find the rate of flow of heat \( (dQ/dt) \) through the walls of the pipe when the inner and outer surfaces are maintained at temperatures \( \theta_1 \) and \( \theta_2 \) respectively \( (\theta_1 > \theta_2) \).

As the walls of the pipe are thin we may take the area normal to the flow of heat as that of the surface of a cylinder of radius equal to the mean of the radii of the inner and outer surfaces of the pipe, i.e. \( (r_1 + r_2)/2 \). We also assume that the temperature gradient is uniform. The values to be substituted in equation (5.1) are thus:

\[ A = 2\pi \left( \frac{r_1 + r_2}{2} \right) l \]

\[ \frac{d\theta}{dx} = \frac{(\theta_2 - \theta_1)}{(r_2 - r_1)} \]

Hence

\[ \frac{dQ}{dt} = \frac{k\pi(r_1 + r_2)l(\theta_1 - \theta_2)}{r_2 - r_1} \]
This equation may be used as the basis of experiments to determine the value of $k$ for bad conductors available in the form of thin-walled tubes (e.g. glass or rubber). One surface is usually maintained at 100°C by the passage of steam, and the rate of conduction of heat through the wall determined by having the other surface of the tube in cold water, and observing the rate of rise of temperature of the water. Details of these experiments may be found in practical manuals.

If the pipe is not thin-walled the above approximation is not valid; the treatment which follows applies to all cases. Fig. 5.6 shows the tube described above, except that there is now no restriction on the thickness of the wall.

![Fig. 5.6](image)

Consider a cylinder or radius $r$, whose surface is in the wall of the tube. The area of the curved surface of this cylinder is $2\pi rl$ and the temperature gradient at its surface is $d\theta/dr$. Substituting in equation (5.1) gives for the rate of flow of heat across the surface of the cylinder:

$$\frac{dQ}{dt} = -k \cdot 2\pi rl \cdot \frac{d\theta}{dr},$$

$$\therefore \int \frac{dQ}{dt} \cdot \frac{dr}{r} = -2\pi kl \int d\theta.$$  

Now $dQ/dt$ is independent of $r$, so that integrating the equation over the range $r = r_1$ to $r = r_2$ gives

$$\frac{dQ}{dt} \int_{r_1}^{r_2} \frac{dr}{r} = -2\pi kl \int_{\theta_1}^{\theta_2} d\theta,$$

$$\therefore \frac{dQ}{dt} \cdot \ln \frac{r_2}{r_1} = -2\pi kl (\theta_2 - \theta_1),$$

$$\therefore \frac{dQ}{dt} = \frac{2\pi kl (\theta_1 - \theta_2)}{\ln \frac{r_2}{r_1}}.$$
5.6 Conduction through the Wall of a Hollow Sphere

Consider the hollow sphere shown in cross-section in Fig. 5.7. The area of a spherical surface of radius \( r \) in the wall of the sphere is \( 4\pi r^2 \); hence substitution in equation (5.1) gives

\[
\frac{dQ}{dt} = -k \cdot 4\pi r^2 \cdot \frac{d\theta}{dr},
\]

\[
\therefore \frac{dQ}{dt} \int_{r_1}^{r_2} \frac{dr}{r^2} = -4\pi k \int_{\theta_1}^{\theta_2} d\theta,
\]

\[
\therefore \frac{dQ}{dt} \left( \frac{1}{r_1} - \frac{1}{r_2} \right) = -4\pi k (\theta_2 - \theta_1),
\]

\[
\therefore \frac{dQ}{dt} = \frac{4\pi k (\theta_1 - \theta_2)}{r_1 - r_2}.
\]

Fig. 5.7

5.7 Conductivity of Liquids and Gases

Liquids (except mercury) are poorer conductors than solids, and gases are worse still.

In the measurement of the conductivity of fluids care must be taken to eliminate the effects of convection. For liquids the methods employed usually involve heating of the liquid from above so that no transfer of heat by convection may take place. A method similar to Lee's disc for solids may be used.

This technique may also be employed for gases. Another method used is to determine the rate of loss of heat from a straight wire through which an electric current is passing. The wire is surrounded by a jacket containing the gas under investigation. The effects of convection are eliminated by performing the experiment at a sufficiently low pressure for convection to be virtually inoperative. Except at very low pressures the conductivity of a gas does not depend on the pressure, so that this reduction of pressure does not reduce heat transfer by this means. The loss by radiation must also be measured and allowed for; this may be accomplished by evacuating the jacket, and repeating the experiment.

It has been mentioned above that the conductivity of a gas is almost independent of the pressure except at very low pressures. This point will be taken up again in Chapter 6 (see Section 6.10).

5.8 Introduction

The transfer of heat in a fluid by convection is a much more complex process to treat theoretically than is conduction; the rate at which heat is transferred depends on such a large number of factors. It is usual to distinguish between two types of convection in gases: forced convection, in
which the gas is caused to move past the hot object by some external agency (e.g. a current of air from a fan), and natural convection, in which the only movement of the gas is that caused by the change in density when the gas comes into contact with the hot body.

5.9 Forced Convection. Newton's Law of Cooling

The rate of cooling of a body of given size and surface depends on the temperature excess of the body over the current of gas.

An empirical law due to Newton states that

*The rate of loss of heat from a body subjected to forced convection is proportional to the difference in temperature between the body and the current of gas,*

i.e.

\[
\frac{dQ}{dt} = -k(\theta - \theta_0)
\]

where \(\theta\) is the temperature of the body and \(\theta_0\) that of the current of gas. The law is true for large temperature excesses provided that convection is forced.

The law is not applicable to cooling in still air, and although it is often applied in estimating cooling corrections, it is only an approximation in this case. However, provided the excess temperature is small (less than 30°C), the approximation is good enough in elementary work.

5.10 Natural Convection

The factors which need to be taken into account in the general case are numerous, e.g.:

(i) size, shape and position of the body;
(ii) temperature excess;
(iii) density, thermal conductivity, viscosity, specific heat, coefficient of expansion, of the gas.

It is possible to take these factors into account by the method of dimensional analysis, but the argument is too complicated to discuss here.

If we restrict the problem to that of a given body and gas, then it is possible experimentally, and theoretically, to investigate the relationship between rate of loss of heat and temperature excess. The results of theoretical work by Lorentz, and practical work by Dulong and Petit and by Lang suggest that the law should be

\[
\frac{dQ}{dt} = -k(\theta - \theta_0)^2.
\]

This is known as the *five-fourths power law.*
5.11 Thermal Radiation

We are concerned here with thermal radiation, i.e. radiation that is emitted from a body by virtue of its temperature. It is well known that this radiation forms but part of the complete electromagnetic spectrum, which also covers X-rays, visible radiation, radio waves, etc. These other parts of the spectrum are treated elsewhere. Because the majority of the radiation from a hot body has wavelengths in the region beyond the red of the visible spectrum, it is often referred to as infra-red radiation. The student should already be familiar with the methods of detection of thermal radiation and with its elementary properties.

All bodies, whatever their temperature, radiate energy; they also absorb some of the radiant energy emitted by neighbouring objects. If the body is in equilibrium in a constant-temperature enclosure, then its net loss of energy is zero, but it is still radiating and it is still absorbing; the equilibrium is a dynamic one. This is the principle known as Prevost’s theory of exchanges.

Radiation is not, of course, the only way in which a body may gain or lose energy, so that for equilibrium other than in a constant-temperature enclosure it is not necessary that the radiated and absorbed energies be equal. Some energy may be transferred by conduction or convection, or the body may itself be generating heat.

5.12 Definition of Terms

There is some confusion in text-books concerning the definitions of terms used. The ones adopted in this book are as follows.

A black body is a body which absorbs all the radiation falling on it.

The emissive power, \( e_\lambda \), of a body in the wavelength range \( \lambda \) to \( \lambda + \delta \lambda \) is defined such that \( e_\lambda \delta \lambda = \) energy emitted per cm\(^2\) per second in that wavelength range. The total emissive power, \( e \), is the total energy emitted per cm\(^2\) per second. It follows that

\[
e = \int_0^\infty e_\lambda \, d\lambda.
\]

\( E_\lambda \) and \( E \) denote the corresponding quantities for a black body.

The emissivity of a body is the ratio \( e_\lambda / E_\lambda \).

The absorptivity, \( a_\lambda \) (often called absorptive power), is the fraction of the radiation of wavelength \( \lambda \) falling on a body which is absorbed by it. It follows from the definition of a black body that its absorptivity is unity at all wavelengths.

5.13 Black-body Radiation

We have stated that a black body is one which absorbs all the radiation falling on it. The perfect black body is impossible to achieve in practice, but a good approximation to it is a hollow sphere with a small
hole in it – the hole is the black body. All the radiation falling on the hole enters the sphere and very little of it escapes again, particularly if the inside of the sphere is rough and blackened.

A black body is the perfect emitter of radiation, and radiation from it is known as black-body radiation. If a black body is in equilibrium in a constant-temperature enclosure, then it absorbs all radiation falling on it. It must therefore emit the same radiation as that present in the cavity, both in quantity and quality, so that it does not alter the radiation in the cavity. It follows that the radiation in the cavity is the same as that emitted by a black body – hence the alternative name cavity radiation. The terms full, or temperature, radiation are also used.

5.14 Kirchhoff’s Law

This is the most widely misapplied law in radiation theory. The law applies only under certain restricted conditions, as we shall see, and it is the failure to appreciate these conditions which leads to erroneous conclusions.

Consider a body which has emissive power $e_\lambda$ and absorptivity $a_\lambda$ for wavelength $\lambda$, at a certain temperature, and suppose that it is in equilibrium in a constant-temperature enclosure at this temperature. Suppose that the amount of radiation in the wavelength range $\lambda$ to $\lambda + \delta\lambda$ falling on it per cm$^2$ per second is $\delta Q$. Then the fraction of the incident radiation absorbed is $a_\lambda \delta Q$. As the body is in equilibrium, the energy radiated must be equal to the energy absorbed. Hence

$$e_\lambda \delta\lambda = a_\lambda \delta Q.$$  

(Note that it is unnecessary to consider conduction or convection, as we are dealing with a constant-temperature enclosure.)

$$\therefore \frac{e_\lambda}{a_\lambda} = \frac{\delta Q}{\delta\lambda},$$

a quantity which depends on the temperature and on the wavelength range, but not on the nature of the body. This is Kirchhoff’s law, which in words states that

*The ratio of the emissive power for a certain wavelength to the absorptivity at the same wavelength is the same for all bodies at the same temperature.*

For a black body $a_\lambda = 1$, so that the constant in Kirchhoff’s law is equal to $E_\lambda$, the emissive power of a black body;

i.e.

$$\frac{e_\lambda}{a_\lambda} = E_\lambda.$$

$$\therefore \frac{e_\lambda}{E_\lambda} = a_\lambda.$$

The ratio $e_\lambda/E_\lambda$ has been defined as the emissivity, so that we may say that the emissivity and absorptivity of a body are equal at a given temperature and wavelength.
The classic example of Kirchhoff’s law is the absorption spectrum of sodium vapour. If bright white light is passed through a sodium flame and then examined by a spectrometer it is found that the spectrum is crossed by two dark lines in the position normally occupied by the sodium emission lines. At each of these wavelengths (5,890 Å and 5,896 Å), at the temperature of the flame, sodium vapour has high values of $\epsilon_\lambda$, and hence by Kirchhoff’s law, high values of $a_\lambda$. Radiation of these wavelengths is absorbed from the white light and then re-emitted in all directions, so that there is a reduction in the intensity of this radiation entering the spectrometer. Note that the two conditions of same wavelength and same temperature are both fulfilled in this case.

If the student examines carefully many of the so-called ‘examples’ of Kirchhoff’s law he will find that one or other of these two conditions is not fulfilled. To be a true example of Kirchhoff’s law the emission and absorption must be taking place at the same wavelength and at the same temperature. It is not intended to imply that the phenomena described in these examples do not exist, but only that if they do exist they have nothing to do with Kirchhoff’s law.

5.15 Stefan’s Law

Stefan’s law applies to a black body; it states that the total energy radiated is proportional to the fourth power of the absolute temperature of the black body. This law was originally based on experimental evidence, but it has since been given a theoretical basis by thermodynamical reasoning.

Considering unit area of the black body, we may write for the energy, $E$, emitted per second at temperature $T^\circ K$,

$$E = \sigma T^4,$$

where $\sigma$ is known as Stefan’s constant.

The units of Stefan’s constant depend on the units used for measuring $E$:

- $E$ in cal, $\sigma$ in cal cm$^{-2}$ sec$^{-1}$ deg K$^{-4}$;
- $E$ in erg, $\sigma$ in erg cm$^{-2}$ sec$^{-1}$ deg K$^{-4}$;
- $E$ in joule, $\sigma$ in watt cm$^{-2}$ deg K$^{-4}$;
- or in watt m$^{-2}$ deg K$^{-4}$.

The value of $\sigma$ is $5\cdot67 \times 10^{-5}$ erg cm$^{-2}$ sec$^{-1}$ deg K$^{-4}$.

The black body also absorbs all the radiation falling on it from its surroundings. If these are at temperature $T_0$ it is absorbing energy at the rate $\sigma T_0^4$ per unit area of its surface, as the following argument shows. Suppose that the black body is at a temperature $T_0$ and situated in a constant-temperature enclosure also at temperature $T_0$. The body is absorbing energy at the same rate that it emits, which we know from Stefan’s law to be at the rate of $\sigma T_0^4$ per unit area. If the temperature of the black body is now raised to $T$ this can make no difference to the rate at which it absorbs radiation (the enclosure remaining at $T_0$), and
HEAT TRANSFER

hence this remains at $\sigma T_0^4$. Hence the net rate of loss of energy per unit area is

$$E_{\text{net}} = \sigma(T^4 - T_0^4).$$

Two extreme cases often arise in practice:

(i) $T \gg T_0$. In this case $T_0^4$ can be neglected in comparison with $T^4$, i.e. the heat absorbed from the surroundings is a negligible fraction of the heat emitted. In this case

$$E_{\text{net}} \simeq \sigma T^4.$$

(ii) $T \approx T_0$, i.e. $T - T_0$ small.

$$E_{\text{net}} = \sigma(T^4 - T_0^4)$$

$$= \sigma(T^2 + T_0^2)(T^2 - T_0^2)$$

$$= \sigma(T^2 + T_0^2)(T + T_0)(T - T_0).$$

But $T^2 + T_0^2 \simeq 2T_0^2$ and $T + T_0 \simeq 2T_0$,

$$\therefore E_{\text{net}} \simeq 4\sigma T_0^3(T - T_0),$$

i.e. $E_{\text{net}}$ is proportional to the temperature excess, provided this is small. It is worth noting that for small temperature excesses, conduction, convection and radiation losses are all approximately proportional to the temperature excess of the body above its surroundings.

### 5.16 Energy Distribution in the Black-body Spectrum

Measurements on the spectrum of black-body radiation have yielded
the results depicted in Fig. 5.8 for the intensity of radiation at various
wavelengths. The following points should be noted:

(i) Each curve lies completely under the one for a higher temperature, i.e. the intensity of the radiation at all wavelengths increases
with temperature.

(ii) At each temperature there is a wavelength for which the intensity is a maximum, and this wavelength becomes shorter as the
temperature rises; hence the change in colour as an object is heated.

(iii) The area under each curve must be proportional to $T^4$, in
accordance with Stefan’s law.

5.17 Wien’s Laws

From thermodynamic reasoning Wien was able to show that

$$E_\lambda = T^5 f(\lambda T),$$

$$= \lambda^{-5} g(\lambda T),$$

where $f(\lambda T)$ and $g(\lambda T)$ are functions of the product $\lambda T$. It is not possible
to determine the nature of the function by the methods of thermodynamics.

Two special cases may be deduced. Both apply to the maximum of
the distribution curve; the co-ordinates of which we denote by $\lambda_{\text{max}}$
and $E_{\text{max}}$.

(i) $\lambda_{\text{max}} T = \text{constant} = 0.293$.

This is known as Wien’s displacement law.

(ii) $E_{\text{max}} T^{-5} = \text{constant}$.

5.18 The Rayleigh–Jeans Formula

The first attempts to derive an expression for $f(\lambda T)$ in Wien’s law
were unsuccessful.

Rayleigh (and later Jeans), using the methods of statistical mech-
nanics, obtained the result

$$E_\lambda = 8\pi k T \lambda^{-4},$$

which satisfies Wien’s law, as can be seen by rewriting

$$E_\lambda = \lambda^{-5}(8\pi k T \lambda).$$

It is obvious, however, without even plotting the curve, that the Ray-
leigh–Jeans formula will not agree with the experimental results, be-
cause as $\lambda \rightarrow 0$, $E_\lambda \rightarrow \infty$. It thus predicts an increasing amount of
radiation at short wavelengths, often called the ‘ultra-violet cata-
trophe’.

The failure of this law, based on well-established classical theories,
led Planck to put forward the revolutionary new idea which is now the
basis of quantum theory, i.e. that an oscillator can possess energy only in certain discrete values.

Further discussion of this important topic will be found in Chapter 11 - The Quantum Theory.

QUESTIONS

1. Show that the time, \( t \), taken for ice on a pond to grow from thickness \( x_1 \) to \( x_2 \) is given by
   \[
   t = \frac{\rho L}{2k\theta}(x_2^2 - x_1^2),
   \]
   where \( \rho \) is the density of ice, \( L \) its latent heat, \( k \) its thermal conductivity and \( \theta \) the temperature of its surface. What assumptions are made in obtaining this result?

2. A body of thermal capacity \( W \) is suspended in an evacuated enclosure by a wire of length \( l \) and area of cross-section \( A \), made of a metal of thermal conductivity \( k \). If the top of the wire is maintained at a constant temperature, show that the temperature difference, \( \phi \), between the two ends of the wire after time \( t \) is given by
   \[
   \phi = \phi_1 e^{-\frac{ka}{W}}t,
   \]
   where \( \phi_1 \) is the value of \( \phi \) when \( t = 0 \). (Cooling by radiation may be neglected.)

3. A body cooling in a draught (Newton's law of cooling) has an excess temperature of \( \phi \) at a certain time. Show that the excess temperature at time \( t \) later is given by \( \phi = \phi_1 e^{-at} \), where \( a \) is a constant.

4. A black body is situated in a constant-temperature enclosure at temperature \( \theta_0 \), and cools by radiation only. If the temperature of the body is initially \( \theta_1 (\gg \theta_0) \), show that its temperature \( \theta \) after time \( t \) is given by
   \[
   \frac{1}{\theta^3} - \frac{1}{\theta_1^3} = \frac{3aAt}{W},
   \]
   where \( A \) is the surface area of the body and \( W \) its water equivalent.

5. In the above example, if \( \theta_1 \sim \theta_0 \) show that the expression becomes
   \[
   \theta - \theta_0 = (\theta_1 - \theta_0)e^{\frac{4\theta_0^3aAt}{W}}.
   \]

ESSAY TOPICS

6. Determination of the thermal conductivities of solids which are: (i) good conductors; (ii) bad conductors.

7. Determination of the thermal conductivities of: (i) liquids; (ii) gases.

8. Pyrometers.
6

KINETIC THEORY

6.1 Pressure of a Gas

The elementary kinetic theory of gases is based on several simplifying assumptions concerning the nature of a gas. These are:

(i) that a gas consists of a large number of molecules which are in constant random motion, and make many collisions with each other and with the walls of the container;
(ii) that the molecules themselves have negligible volume compared with the space they inhabit;
(iii) that all collisions are perfectly elastic;
(iv) that between collisions the molecules move in straight lines (i.e. no attractive forces).

The pressure exerted by a gas on the walls of the containing vessel is due to the bombardment of the walls by the molecules of the gas.

Consider an area $\delta S$ (assumed flat) of one wall of the containing vessel. The molecules which hit this area do so at many different angles and with many different velocities. We divide the molecules into groups so that all the molecules in a group have approximately the same velocity both in magnitude and direction. Let there be $n_1$, $n_2$, $n_3$, ... molecules per cc in each group and a total of $n$ molecules per cc in the gas. Suppose that the $n_1$ molecules of the first group move in the direction indicated by the arrow in Fig. 6.1, and have a component velocity $u_1$ in the direction of the normal to $\delta S$. All the molecules of this group which will hit the wall in time $\delta t$ are contained in the oblique cylinder indicated in Fig. 6.1. This has volume $u_1 \delta t \delta S$ and contains $n_1 u_1 \delta t \delta S$ of the molecules in question. When each of these molecules collides with the wall the component of its velocity perpendicular to the wall is reversed, so that its change in momentum is $2mu_1$, where $m$ is the mass of a molecule. The total change in momentum in time $\delta t$ due to the collisions of molecules in the first group is thus $(2mu_1)(n_1 u_1 \delta t \delta S)$.

By Newton's second law of motion, the force exerted by $\delta S$ on the molecules is equal to their rate of change of momentum. This is equal
(and opposite) to the force exerted by the molecules on $\delta S$, which is therefore $2mn_1u_1^2 \delta t \, \delta S/\delta t$, and the force per unit area is $2mn_1u_1^2$. This, then, is the pressure on $\delta S$ due to the bombardment of the molecules in the first group.

Now, only half of the groups of molecules are moving towards $\delta S$, so that the total pressure, $p$, on the wall is given by

$$p = 2m \cdot \frac{1}{2} \sum n_i u_i^2,$$

where the summation is over all groups.

$$\therefore p = mn \overline{u^2}, \quad \text{where} \quad \overline{u^2} = \frac{n_1u_1^2 + n_2u_2^2 + \ldots}{n}.$$

If we regard the direction of the normal to $\delta S$ as the $x$-axis each molecule also has components of velocity ($v$ and $w$) along the $y$- and $z$-axes respectively. If $c$ denotes the actual velocity of a molecule it follows that

$$\overline{c^2} = \overline{u^2} + \overline{v^2} + \overline{w^2}.$$ 

If it is assumed that the velocities of the molecules are entirely random in space, then

$$\overline{u^2} = \overline{v^2} = \overline{w^2} = \frac{1}{3} \overline{c^2}.$$ 

Hence

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

(6.1)

where $\rho$ is the density of the gas. $\overline{c^2}$ is the mean square velocity of the molecules, and $(\overline{c^2})^{\frac{1}{2}}$ the root mean square velocity. To define the mean square velocity by dividing the molecules into a finite number of groups is not entirely satisfactory. A better way is to denote the number of molecules with velocities between $c$ and $c + \delta c$ by $n_c \, \delta c$ and then define

$$\overline{c^2} = \frac{\int_0^\infty n_c c^2 dc}{\int_0^\infty n_c dc} = \frac{\int_0^\infty n_c c^2 dc}{n}.$$ 

In this proof of the pressure formula no account has been taken of the collisions between molecules; it can be shown that these collisions have no effect on the result obtained.

6.2 Deductions from the Pressure Formula

(a) Equation (6.1) may be written

$$p = \frac{3}{8} (n \cdot \frac{1}{2} m \overline{c^2}).$$

But $\frac{1}{2} m \overline{c^2}$ is the average kinetic energy of a molecule, and $n$ is the number of molecules per cc.

$$\therefore p = \frac{3}{8} \times \text{kinetic energy per cc}. $$
This result makes plausible the statement made above that \( p \) is not affected by collisions between molecules: such collisions cannot alter the total kinetic energy.

(b) If both sides of equation (6.1) are multiplied by the volume, \( V \), occupied by 1 gram-molecule of the gas at pressure \( p \) and temperature \( T^\circ K \) we obtain

\[
pV = \frac{1}{3}pVc^2 = \frac{1}{3}Mc^2,
\]

where \( M \) is the molecular weight of the gas.

Experimental results show that for an ideal gas

\[
pV = RT,
\]

where \( R \) is the gas constant.

This suggests that we put

\[
\frac{1}{3}Mc^2 = RT.
\]

Dividing by Avogadro's number (i.e. the number of molecules in a gram-molecule) gives

\[
\frac{1}{3}mc^2 = \frac{R}{N}T = kT
\]

where \( m \) is the mass of one molecule and \( k(= R/N) \) is Boltzmann's constant. Multiplying by \( \frac{3}{2} \) gives

\[
\frac{1}{2}mc^2 = \frac{3}{2}kT,
\]

i.e. the average kinetic energy of a molecule = \( \frac{3}{2}kT \).

### 6.3 Number of Molecules Crossing a Plane in a Gas

Molecules pass through any plane in a gas from both sides; we require to find the number passing through unit area from one side in one second.

Before proceeding to the main argument it is first necessary to obtain a mathematical result of which we shall make use. Suppose that a large number, \( N \), of particles are moving at random; we shall show that the number of these particles moving in a direction making an angle between \( \theta \) and \( \theta + \delta\theta \) with a given direction is \( (N/2) \sin \theta \delta\theta \).

We may represent the direction of motion of each particle by a line drawn from an origin \( O \), in a direction parallel to the actual path of the particle (Fig. 6.2). If a sphere is drawn with \( O \) as centre, then the number of lines passing through an area \( \delta S \) on the surface of the sphere is independent of where \( \delta S \) is drawn; this follows from the random motion of the particles. Thus the number of lines passing through \( \delta S \) is proportional to that area.

In Fig. 6.3 AOB is a fixed line. Those particles whose paths make an
angle between $\theta$ and $\theta + \delta \theta$ with the direction represented by AOB are those represented by lines passing through the shaded area of the sphere, and the number involved ($N_\theta$) is proportional to this area.

![Fig. 6.2](image)

![Fig. 6.3](image)

The portion of the surface concerned has area $2\pi(a \sin \theta)a \delta \theta$, where $a$ is the radius of the sphere. The total surface area of the sphere is $4\pi a^2$.

$$\therefore \frac{N_\theta}{N} = \frac{2\pi a^2 \sin \theta \delta \theta}{4\pi a^2}$$

$$\therefore N_\theta = \frac{N}{2} \sin \theta \delta \theta \quad \ldots \ldots \quad (6.2)$$

We proceed now with the calculation of the number of molecules crossing a plane in a gas which has $n$ molecules per cc. Divide the molecules of the gas into a large number of groups so that all the molecules in a group have velocities of approximately the same magnitude. Suppose that there are $n_1$ molecules per cc in the group with velocity $c_1$; $n_2$ with velocity $c_2$; and so on. Consider those molecules in the first group which arrive at an area $\delta S$ of the plane from one side, in a direction making an angle between $\theta$ and $\theta + \delta \theta$ with the normal. The number of these molecules that will cross $\delta S$ at the required angle during time $\delta t$ is the same as the number in a cylinder of height $c_1 \delta t \cos \theta$ whose base is $\delta S$ (Fig. 6.4). Some of the molecules in this cylinder which have the required velocity and direction do not go through $\delta S$, but these are exactly compensated for by those from outside the cylinder which do go through $\delta S$. The volume of the cylinder is $c_1 \delta t \cos \theta \delta S$, so that the number ($N$) of molecules of velocity $c_1$ in it is $n_1c_1 \delta t \cos \theta \delta S$. Of these only a fraction ($N_\theta$, given by equation (6.2)), cross the plane at the required angle.

$$\therefore N_\theta = \frac{1}{2}n_1c_1 \delta t \cos \theta \delta S \sin \theta \delta \theta.$$
The number of molecules with velocity $c_1$ passing through $\delta S$ from right to left in time $\delta t$, at all angles, is found by integrating this expression over the range $\theta = 0$ to $\theta = \pi/2$;

i.e. number of molecules \[ \frac{1}{2} n_1 c_1 \delta t \frac{\pi}{2} \cos \theta \sin \theta \, d\theta \]

\[ = \frac{1}{2} n_1 c_1 \delta t \delta S \left[ -\frac{\cos^2 \theta}{2} \right]_0^{\pi/2} \]

\[ = \frac{1}{4} n_1 c_1 \delta t \delta S. \]

The total number of molecules, $\nu$, of all velocities crossing unit area per second is thus given by

\[ \nu = \frac{1}{4} \left( n_1 c_1 + n_2 c_2 + \ldots \right) \]

\[ \therefore \nu = \frac{1}{4} nc \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad (6.3) \]

where

\[ \bar{c} = \frac{n_1 c_1 + n_2 c_2 + \ldots}{n} \]

$\bar{c}$ is known as the mean velocity; a better definition of this is:

\[ \bar{c} = \frac{\int_0^\infty n_c c dc}{\int_0^\infty n_c dc} = \frac{\int_0^\infty n_c c dc}{n}. \]

### 6.4 Maxwell’s Law of Distribution of Velocities

So far, although note has been taken of the fact that the molecules have a range of velocities, no attempt has been made to describe the nature of this distribution. This may be done theoretically by the methods of statistical mechanics. Maxwell obtained an expression for the number of molecules per cc, with velocities between $c$ and $c + \delta c$; if this number is denoted by $n_c \delta c$, then $n_c = A c^2 \exp(-bc^2)$, where $A$ and $b$ are constants for a certain gas under given conditions of temperature and pressure.

![Fig. 6.5](image-url)
A graph of $n_c$ against $c$ is shown in Fig. 6.5. The velocity corresponding to the maximum of this curve is known as the most probable velocity, $c_m$, of the molecules.

It may be shown from Maxwell’s distribution law that

$$
\langle c^2 \rangle = \left( \frac{3\pi}{8} \right)^{\frac{1}{2}} c = 1.09c
$$

and

$$
c = \left( \frac{4}{\pi} \right)^{\frac{1}{2}} c_m = 1.13c_m,
$$

i.e.

$$
\langle c^2 \rangle > \bar{c} > c_m.
$$

6.5 Equipartition of Energy

We have shown in Section 6.2 (b) that the average kinetic energy of a molecule is $\frac{3}{2}kT$. This was based on the assumption that the molecules are small hard spheres; they may therefore possess energy only by virtue of their translational motion in three dimensions. To describe the velocity of such a molecule it is necessary to specify its components in the directions of a set of three mutually perpendicular axes. If these components are $u$, $v$, $w$, then

$$
\langle c^2 \rangle = u^2 + v^2 + w^2
$$

and the kinetic energy $= \frac{1}{2}m\langle c^2 \rangle = \frac{1}{2}m(u^2 + v^2 + w^2)$.

The number of squared terms in the expression for the energy of a body is called the number of degrees of freedom, so that our point molecule is said to have three degrees of freedom.

We have seen that the average kinetic energy of such a molecule is $\frac{3}{2}kT$, i.e. $\frac{1}{2}kT$ per degree of freedom. The theory of equipartition of energy extends this to cover molecules with larger numbers of degrees of freedom, and postulates that in all cases each degree of freedom has associated with it an energy $\frac{1}{2}kT$. For example, if a molecule has five degrees of freedom it has, on average, an energy of $\frac{5}{2}kT$; we shall later see that this is the case for a diatomic molecule.

6.6 Specific Heats of Gases

If we denote the molar specific heats of a gas at constant pressure and constant volume by $C_p$ and $C_v$ respectively, then

$$
C_p - C_v = R . \quad . \quad . \quad . (6.4)\
$$

where $C$ and $R$ are measured in the same energy units (e.g. either ergs or calories). We write $\gamma = C_p/C_v$.

Now, for a monatomic gas (which almost fulfils the assumption of point molecules) the average energy per molecule is $\frac{3}{2}kT$, and the energy per gram-molecule is $\frac{3}{2}RT$. Suppose that the temperature of one gram-molecule of the gas is raised by 1 deg K. Then the energy increases by $\frac{3}{2} R$. But the increase in internal energy for 1 deg K rise in

* The proof of this relation will be found in A-level texts.
temperature is the specific heat at constant volume, $C_v$, hence

$$C_v = \frac{5}{2}R$$

It follows from equation (6.4) that

$$C_p = \frac{5}{3}R,$$

and hence $\gamma = \frac{5}{3} = 1.67$ for a monatomic gas. Table 6.1 shows how well actual gases conform to this.

<table>
<thead>
<tr>
<th>Gas</th>
<th>Atomicity</th>
<th>$\gamma$ at 20°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argon</td>
<td>1</td>
<td>1.67</td>
</tr>
<tr>
<td>Helium</td>
<td>1</td>
<td>1.67</td>
</tr>
<tr>
<td>Neon</td>
<td>1</td>
<td>1.64</td>
</tr>
<tr>
<td>Nitrogen</td>
<td>2</td>
<td>1.41</td>
</tr>
<tr>
<td>Oxygen</td>
<td>2</td>
<td>1.40</td>
</tr>
<tr>
<td>Carbon Dioxide</td>
<td>3</td>
<td>1.30</td>
</tr>
<tr>
<td>Methane</td>
<td>5</td>
<td>1.31</td>
</tr>
</tbody>
</table>

Table 6.1

For a diatomic gas the number of degrees of freedom is greater. If we picture the molecule as a ‘dumb bell’ (Fig. 6.6), then in addition to the kinetic energy of translation the molecule may also possess kinetic energy of rotation. This introduces two more degrees of freedom for rotation about each of two axes perpendicular to the line joining the atoms. Rotation about the line joining the atoms is ruled out because the atoms are considered as point masses, and hence the moment of inertia about this line is vanishingly small.

As there are five degrees of freedom,

$$C_v = \frac{5}{2}R,$$

and

$$C_p = \frac{7}{2}R.$$

Thus $\gamma = \frac{7}{5}$ for a diatomic gas without vibration. A further two degrees of freedom are introduced if the atom vibrates along the line of centres. This introduces two degrees, because in the expression for the energy of vibration there is a squared term corresponding to the potential energy and a squared term for the kinetic energy.

Hence

$$C_v = \frac{7}{2}R,$$

and

$$C_p = \frac{9}{2}R.$$

Thus $\gamma = \frac{9}{5}$ for a diatomic gas with vibration. It will be clear from the above that if, in general, a molecule has $n$ degrees of freedom, then

$$C_v = \frac{n}{2}R,$$

$$C_p = \frac{n + 2}{2}R,$$

and

$$\gamma = \frac{n + 2}{n}.$$
For a polyatomic molecule there are at least six degrees of freedom (three translational and three rotational) and there may be many more if vibrations take place.

6.7 Variation of Specific Heats of Gases with Temperature

The theory outlined above does not predict any variation in the specific heat of a gas with temperature. Experiment has shown that variations do occur, particularly at low temperatures. An interesting and informative case is that of hydrogen, which is a diatomic gas and accordingly has the values: \( C_v = \frac{3}{2}R, \gamma = \frac{7}{5} \) at room temperature. At very low temperatures, however, the value of \( C_v \) decreases to almost \( \frac{3}{2}R \) and \( \gamma \) increases to \( \frac{5}{3} \), the values for a monatomic gas. It would appear, then, that at very low temperatures rotation has ceased and the molecule has only the three translational degrees of freedom. This is impossible to explain on the classical theory of equipartition of energy, especially as the reduction in specific heat is a gradual one; it appears that the number of degrees of freedom changes continuously from five to three.

This phenomenon has been convincingly explained in terms of the quantum theory (see Chapter 11). Rotation in the molecule is subject to the rules of quantization of energy, which means that the rotational energy possessed by the molecule can take on only certain discrete values. The rotation cannot have any intermediate values of energy and, in particular, can have no value of energy less than a certain value \((\epsilon)\).

Suppose first that the temperature of the gas is high, so that \( kT \gg \epsilon \). This means that the average energy per degree of freedom, \( \frac{1}{2}kT \), is well above the lowest allowed energy. The ‘steps’ between allowed values of energy are then very small compared with the energy itself. The actual rotational energies possessed by the individual molecules are thus distributed in almost the same way as predicted by classical theory.

If the temperature is now reduced until \( \epsilon \sim kT \) the average kinetic energy of rotation of a molecule according to the classical theory is of the same order of magnitude as the smallest energy it may possess. The theorem of the equipartition of energy now breaks down, because those molecules which according to the statistical distribution should have had energy less than the average will not now have any energy at all. The total energy of rotation is thus much less than predicted by the classical theory, with the result that \( C_v \) falls.

When the temperature is so low that \( \epsilon \gg kT \) only a very small number of molecules possess enough energy to rotate and the gas behaves almost as a monatomic one.

The quantum theory also explains why we may disregard rotation in a monatomic gas and rotation about the line of centres in a diatomic molecule, even though the atoms do have finite sizes. The moments of inertia, \( I \), for these rotations are very small and this results in a large
value of $\epsilon$ ($= \hbar^2/8\pi^2l$), so that $\epsilon \gg kT$, even at high temperatures. Very few molecules ever acquire enough energy to rotate about these axes, and the degrees of freedom corresponding to these rotations may be ignored.

**6.8 Mean Free Path**

The mean free path may be defined as the average distance a molecule travels between collisions; it will be denoted by $\lambda$. The value of $\lambda$ in a given case depends on the density of the gas and on the diameter of the molecules.

Imagine that all but one of the molecules of the gas are stationary, and that this molecule is moving in a straight line at constant velocity. Let the diameter of the molecules be $\sigma$ and let there be $n$ molecules per cc. The moving molecule collides with any molecule whose centre comes within a distance $\sigma$ of its centre. Thus we may draw a sphere of radius $\sigma$ centred on the moving molecule, and any molecule whose centre comes within this sphere undergoes a collision (Fig. 6.7).

If the molecule moves for distance $l$ the sphere traces out a volume $\pi\sigma^2l$, which includes $n\pi\sigma^2l$ stationary molecules. This is the number of collisions made by the moving molecule in distance $l$, so that the average distance between collisions is

$$\frac{l}{n\pi\sigma^2l} = \frac{1}{n\pi\sigma^2l}$$

Hence

$$\lambda = \frac{1}{n\pi\sigma^2} \quad \quad \quad \quad \quad (6.5)$$

For a given gas $n \propto \rho$, the density of the gas

$$: \quad \lambda \propto \frac{1}{\rho}$$

The pressure, $p$, of the gas is proportional to $\rho$, so that

$$\lambda \propto \frac{1}{p}$$
The expression (equation (6.5)) for \( \lambda \) is an approximate one, because it has been deduced on the assumption that only one molecule is moving. If the velocities of all the molecules are taken into account and use made of Maxwell’s distribution law the result obtained is

\[
\lambda = \frac{1}{\sqrt{2\pi \sigma^2}}.
\]

The mean free path for hydrogen at N.T.P. is about \( 1.1 \times 10^{-5} \) cm and at a pressure of \( 10^{-3} \) cm of mercury is about 8 mm.

### 6.9 Viscosity

Fig. 6.8 illustrates the laminar flow of a gas over a flat surface; the lengths of the arrows represent the velocities of various layers of the gas (see Section 2.12).

![Fig. 6.8](image)

Now consider the picture in terms of the kinetic theory of gases. All the molecules are undergoing random movements due to their thermal energy, but superimposed on these is an additional constant velocity in the direction of flow of the gas as a whole. The path of an individual molecule may be as shown in Fig. 6.9.

Notice that, although the molecule still pursues a zig-zag path due to collisions with other molecules, the component of its velocity in the direction of the arrow is on the whole greater than the components in other directions, so that it is progressing towards the right. During its random motion the molecule moves from one layer of gas to another. Suppose it moves to a region of lower flow velocity – e.g. from A to B in Fig. 6.8, crossing the plane S indicated by the dotted line. As a result of this movement the average velocity of the molecules at B in the direction of flow is increased. Similarly, molecules that move the opposite way across S cause a reduction in the flow velocity in the region of A. We see, then, that the random motion of the molecules results in a tendency to decrease the velocity in the region of A and increase it in the region of B. In equilibrium the flow velocities in the two layers are constant; the decrease in velocity at A due to the effect described above is compensated for by interchange of molecules between A and the faster-moving layers above it. Similarly, the increase in velocity at B is counteracted by the transfer of slower-moving molecules from below.
Consider now the effect at the boundary S in Fig. 6.8. We have already stated that the molecules from A which move across this boundary have a higher velocity (in the direction of flow) than those in the region B to which they are crossing. This means that they cause an increase in the momentum of layer B. Returning now to a macroscopic picture (i.e. ignoring the structure of the gas), we would say that if the layer A causes an increase in the momentum in layer B it is exerting a force on that layer in the direction of the increase in momentum. This is the tangential viscous force previously discussed in Section 2.12.

To sum up, we have explained the phenomenon of viscosity (which is a macroscopic effect) in terms of the microscopic structure of the gas itself (kinetic theory).

We now proceed to calculate an expression for the coefficient of viscosity of a gas in terms of the kinetic theory. An approximate treatment only will be given as indicated by the use of the sign \( \sim \) to indicate orders of magnitude. Consider the transfer of momentum across the plane S (Fig. 6.10), where the flow velocity is \( v_0 \). The molecules crossing

\[ v_0 + \lambda \frac{dv}{dx} \]

\[ v_0 - \lambda \frac{dv}{dx} \]

S will, on the average, have come from a distance \( \lambda \) on either side of S, where \( \lambda \) is the mean free path. If the velocity gradient in the gas is \( \frac{dv}{dx} \), then the molecules crossing from above will, on the average, have velocity \( v_0 + \lambda \frac{dv}{dx} \), and those from below velocity \( v_0 - \lambda \frac{dv}{dx} \) in the direction of flow.

Suppose that there are \( n \) molecules, each of mass \( m \), in 1 cc of the gas. The number of molecules crossing unit area of S per sec \( \sim n \bar{c} \) (see equation (6.3))

\[ \therefore \] increase in momentum at B per sec \( \sim n \bar{c} m \lambda \frac{dv}{dx} \).

and decrease in momentum at A per sec \( \sim n \bar{c} m \lambda \frac{dv}{dx} \).

But, Force = Change in momentum per second,

hence the tangential viscous force per unit area, \( F/A \sim n \bar{c} \mu \lambda \frac{dv}{dx} \).

But \[ \frac{F}{A} = \eta \frac{dv}{dx} \] (see equation (2.2))
where $\eta$ is the coefficient of viscosity of the gas.
\[
\therefore \eta \sim mn^2a,
\]
i.e.
\[
\eta \sim \rho \bar{c} \lambda
\]  \hspace{1cm} (6.6)

where $\rho$ is the density of the gas. The full calculation yields $\eta = \frac{1}{3} \rho \bar{c} \lambda$.

Now, $\lambda = 1/n\pi a^2$, so that $n\lambda = \text{constant}$, for any given gas.

Thus $\eta \propto \bar{c}$ for a given gas.

This leads to the surprising prediction that $\eta$ does not depend on the density of the gas. This is quite well confirmed by experiment except at very low pressures, when $\lambda$ becomes comparable with the dimensions of the vessel; and at high pressures, when the assumptions of the kinetic theory cease to be valid. Also $\eta$ should be proportional to $T^{1/4}$, where $T$ is the absolute temperature. This follows because $\bar{c} \propto (\bar{c}^2)^{1/4}$ and $\bar{c}^2 \propto T$.

The relation between $\eta$ and $T$ is not well obeyed in practice. The theory breaks down because we have made the assumption that $\sigma$ is independent of the velocity of the molecules.

### 6.10 Thermal Conductivity

A and B (Fig. 6.11) are two regions in a gas where the temperatures are $T_A$ and $T_B$ respectively ($T_A > T_B$). This means that the molecules

![Diagram](image)

Fig. 6.11

at A have higher velocities (on average) than those at B, and hence the random movement of molecules from A to B and from B to A causes a net increase in the energy at B and a net decrease at A, i.e. thermal conduction takes place from A to B.

Assume that the molecules at S have an average energy $E_0$, and that the energy gradient is $dE/dx$. If there are $n$ molecules per cc, each of mass $m$, then:

number of molecules crossing unit area of S (from A to B) per second $\sim ne$.

On the average these have come from distance $\lambda$, and hence have energy $E_0 + \lambda \frac{dE}{dx}$. 
ADVANCED THEORETICAL PHYSICS

the rate at which energy is transferred across unit area

\[ S \sim \bar{c} \lambda \frac{dE}{dx}. \]

But from the conduction equation this rate of increase of energy

\[ = k \frac{dT}{dx}, \]

where \( k \) is the coefficient of thermal conductivity.

\[ \therefore k \sim \bar{c} \lambda \frac{dE}{dT}. \]

(6.7)

In one gram of the gas there are \( \frac{1}{m} \) molecules, so that the kinetic energy, \( U \), in one gram is \( \frac{1}{m} \cdot E \). Now \( dU/dT = c_v \), the specific heat of the gas at constant volume.

\[ \therefore \frac{dE}{dT} = mc_v \]

\[ \therefore k \sim m\bar{c} \lambda c_v \sim \rho \bar{c} \lambda c_v. \]

But \( \eta \sim \rho \bar{c} \lambda \) (see equation (6.6)),

\[ \therefore k \sim \eta c_v. \]

(6.8)

For any given gas, then, \( k \) should vary in the same way as \( \eta \); i.e. it should be independent of the pressure and proportional to \( T^4 \). The prediction that \( k \) does not vary with pressure is verified by experiment except at low and high pressures. The dependence on temperature is not well obeyed in practice. The reasons for the breakdown of the theory in these cases are the same as mentioned in the previous section.

The conclusion that \( k \) is independent of pressure is a very surprising one, and means, for example, that the removal of air from between the walls of a Dewar flask is not effective in reducing conduction losses until the pressure reaches the region where the theoretical results break down (i.e. when \( \lambda \) becomes comparable with the dimensions of the vessel).

It follows from equation (6.8) that \( k/\eta c_v \) should be the same for all gases. Experimental results show that this ratio is almost the same for all monatomic gases.

LIQUIDS

6.11 Brownian Motion

If a colloidal suspension of very small particles in a liquid is examined under a high-power microscope, using special illumination, the colloidal particles are observed to be in constant irregular motion. This motion is due to the bombardment of the particles by the molecules of the liquid. A given particle is hit by more molecules on one side than on the other in a very short time interval, so that it is given momentum in a particular direction.

The motion was first observed by the botanist Brown (hence the name given to the motion), who used a suspension of pollen in water. The
existence of Brownian motion is strong evidence in support of the
kinetic theory of matter.

6.12 Perrin’s Experiments. Determination of Avogadro’s Number

A classical series of experiments on Brownian motion was carried out
by Perrin in 1908. The simplest of these to understand is one he used to
determine Avogadro’s number.

The experiment was based on finding the vertical distribution of a
suspension of particles in a fluid (Perrin used emulsions of gamboge and gum-mastic). Because of the
combined effects of gravity and molecular bombard-
ment the concentration of particles at equilibrium is
greatest at the bottom of the vessel and decreases
steadily towards the top. It is assumed that the dis-
tribution is the same as that of air molecules in the
atmosphere. This assumption is borne out by experi-
ments on osmotic pressure which show that this pres-
sure obeys the same laws as a perfect gas.

We proceed first, then, to obtain an expression for
the variation of density with height in the atmo-
sphere.

Consider a column of air, of area of cross-section $A$
(Fig. 6.12), and in particular the portion of it between
the heights $h$ and $h + \delta h$ above the ground. Suppose the pressures are
respectively $p$ and $p + \delta p$ at these heights.

This air is in equilibrium under the action of two forces, due to:

(i) the pressure difference, $\delta p$, between top and bottom, producing
a downward force $\delta p A$;

(ii) its own weight, $\delta h A \rho g$, where $\rho$ is the density of air at height $h$.

Hence, for equilibrium

$$\delta p A + \delta h A \rho g = 0$$

$$\therefore \frac{dp}{dh} = -\rho g \quad \cdots \cdots \quad (6.9)$$

This equation contains three variables, $p$, $h$, $\rho$, and so we first obtain an
expression for $p$ in terms of $\rho$. If we assume that air behaves as a perfect
gas, then

$$pV = RT,$$

where $V$ is the volume of 1 gram-molecule,

i.e.

$$V = \frac{M}{\rho},$$

where $M$ is the molecular weight of the gas.

$$\therefore p = p \frac{RT}{M}.$$
If we now assume that the temperature is independent of the height, then, differentiating with respect to \( h \),

\[
\frac{d\rho}{dh} = \frac{RTd\rho}{M} \frac{1}{dh'}
\]

and substituting in equation (6.9),

\[
\frac{RTd\rho}{M} \frac{1}{dh} = -\rho g.
\]

\[
\therefore \int_{\rho_0}^{\rho} \frac{d\rho}{\rho} = -\frac{Mgh}{RT} \biggr|_0^h dh,
\]

where \( \rho_0 \) is the density at ground level.

\[
\therefore \ln \frac{\rho}{\rho_0} = -\frac{Mgh}{RT},
\]

\[
\therefore \rho = \rho_0 e^{-\frac{Mgh}{RT}} \quad \ldots \quad (6.10)
\]

To include Avogadro's number, \( N \), we write \( M = mN \), where \( m \) is the mass of a molecule.

Hence,

\[
\ln \frac{\rho}{\rho_0} = -\frac{mNgh}{RT} \quad \ldots \quad (6.11)
\]

If equation (6.11) is applied to obtain the variation with height of the number of colloidal particles per cc in a suspension, we must write

\[
\rho_0 = mn_1,
\]

\[
\rho = mn_2,
\]

where \( n_1, n_2 \) are the numbers of particles per cc at two heights separated by a distance \( h \); and \( m \) is the mass of a particle.

\[
\therefore \ln \frac{n_2}{n_1} = -\frac{m'Ngh}{RT} \quad \ldots \quad (6.12)
\]

\( m'g \) is the effective weight of the particle, which is less than its true weight because of the upthrust of the liquid. If \( d \) is the density of the substance of the particles, and \( \sigma \) that of the liquid, then the upthrust is \( mg/d \), so that the effective weight, \( m'g \) is \( mg \left( 1 - \frac{\sigma}{d} \right) \).

Substituting this in equation (6.12) and rewriting gives

\[
N = \frac{RTd}{mgh(d - \sigma)} \ln \frac{n_1}{n_2},
\]

as the required expression for Avogadro's number.

Perrin used a suspension of gamboge particles. By repeated centrifuging he obtained a suspension in which all the particles were approximately the same size. A column of liquid about 0·1 mm high was viewed
from above, using a high-power microscope of very small depth of focus. The microscope was focused on a particular layer in the fluid and the number of particles in the field of view counted. This was repeated for a different layer, distance \( h \) from the first, and hence \( n_1/n_2 \) was calculated.

The density and size of the particles were determined in several ways. In one experiment the density was determined by measuring the density of a liquid in which the particles neither floated nor sank, and the size of the particles was found by observing their terminal velocity and applying Stokes’ formula (Section 2.18).

The value obtained for \( N \) by Perrin was \( 6.8 \times 10^{23} \) compared with the present accepted value, \( 6.02 \times 10^{23} \).

**SOLIDS**

**6.13 Atomic Heat. Dulong and Petit’s Law**

An empirical law due to Dulong and Petit states that for a solid element the atomic heat is approximately \( 6.3 \text{ cal (gram-atom)}^{-1} \text{ deg C}^{-1} \). The atomic heat is defined as the thermal capacity of a gram-atom of the element and is thus the product of the specific heat (in cal g\(^{-1}\) deg C\(^{-1}\)) and the atomic weight.

This law may be justified theoretically by application of the theorem of equipartition of energy. An atom in a solid possesses energy of vibration and has six degrees of freedom (one corresponding to potential energy and one to kinetic energy in each of three directions). At temperature \( T \) the average energy of each atom is therefore \( 6 \times \frac{1}{2} kT \), so that the total energy of a gram-atom of the solid element is \( 3NkT \), where \( N \) is Avogadro’s number. But \( Nk = R \), the gas constant, hence:

\[
\text{Internal energy of solid} = 3RT \text{ per gram-atom.}
\]

The atomic heat of the solid is the increase in internal energy of one gram-atom when the temperature is raised by one degree so that:

\[
\text{Atomic heat} = 3R = 6.0 \text{ cal (gram-atom)}^{-1} \text{ deg C}^{-1}.
\]

The above argument applies to the atomic heat at constant volume, \( C_v \). The measured quantity will be the atomic heat at constant pressure, \( C_p \), and this is somewhat higher (\( C_p - C_v \approx 0.1 \) for copper at room temperature).

Some elements do not obey Dulong and Petit’s law at room temperature: carbon has an atomic heat of about \( 2 \text{ cal (gram-atom)}^{-1} \text{ deg C}^{-1} \) at room temperature. It is found, however, that as the temperature is raised \( C_v \) approaches a value of about \( 6 \text{ cal (gram-atom)}^{-1} \text{ deg C}^{-1} \) for all elements. At low temperatures the value of \( C_v \) is much reduced and tends to zero as \( T \) tends to zero. Curves showing \( C_v \) against \( T \) are similar in shape for all elements (Fig. 6.13).
Debye obtained an expression for the variation of \( C_v \) with temperature, in fair agreement with experimental results. This expression shows that at high temperatures \( C_v \to 3R \), and at low temperatures \( C_v \simeq \text{const.} \left( \frac{T}{\theta} \right)^3 \), where the constant is the same for all elements, and \( \theta \) depends on the element. \( \theta \) is known as the Debye temperature. The low-temperature approximation gives a good fit for most non-metals; for metals the law breaks down because of the energy possessed by the 'free' electrons (see Chapter 13), which has not been taken into account.

**QUESTIONS**

1. A vessel is divided into two parts by a thin insulating partition containing a small hole. The pressure is such that the diameter of the hole is much less than the mean free path of the molecules, and the two compartments are maintained at temperatures \( T_1, T_2 \). Show that the pressures \( p_1, p_2 \) in the two compartments at equilibrium are related by \( p_1/p_2 = (T_1/T_2)^{3/4} \).

2. If the density of a gas on one side of a small hole (of diameter much less than the mean free path) is \( \rho_1 \) and that on the other is \( \rho_2 (\rho_1 > \rho_2) \), show that the net mass of gas transferred through the hole per second is \( \frac{A\tilde{c}}{4} (\rho_1 - \rho_2) \), where \( A \) is the area of the hole.

Using the relation \( \tilde{c} = \left( \frac{8}{3\pi} \right)^{\frac{3}{2}} \), show that this expression may also be expressed as \( A (\rho_1 - \rho_2) \left( \frac{mN}{2\pi RT} \right)^{\frac{1}{2}} \), where \( m \) is the mass of each molecule.
THERMODYNAMICS

7.1 Introduction

Thermodynamics is the study of the relationship between heat and other forms of energy. One very important aspect of thermodynamics, and one with which we shall largely be concerned in this chapter, is the relation between heat and mechanical energy. A device which converts heat into mechanical energy is called a heat engine; the steam engine and the internal-combustion engine are obvious examples.

7.2 The First Law of Thermodynamics

The reader will already be familiar with the experiments, starting with the famous ones by Joule, which have shown that heat is a form of energy, and that when heat is converted into work, or vice-versa, there is a definite ratio, the mechanical equivalent of heat, between the amount of heat energy and the amount of mechanical energy. This is formally stated in the First Law of Thermodynamics as follows:

When a certain quantity of mechanical energy is completely converted into heat energy by any means the same quantity of heat energy is always obtained. When a certain quantity of heat energy is completely converted into mechanical energy by any means the same quantity of mechanical energy is always produced.

This is, of course, a particular case of the more general principle of the conservation of energy.

As an example of the application of the first law, consider a quantity of gas enclosed in a cylinder by a close-fitting piston, which is free to move without friction. If heat is supplied to the gas two things happen:

(i) the gas expands, pushing out the piston and doing a certain amount of external work;
(ii) the temperature of the gas rises, i.e. the energy of the molecules is increased. If we apply the first law to this example, denoting the heat supplied by \( \delta Q \), the work done by the gas by \( \delta W \) and the increase in internal energy (energy of the molecules) by \( \delta U \) then

\[
\delta Q = \delta U + \delta W . \quad \quad \quad (7.1)
\]

It is assumed that all the quantities are measured in the same units (e.g. all in ergs or all in calories).

7.3 Work Done during an Expansion

The expansion of a gas may be represented graphically by plotting the pressure against the volume of the gas as the expansion proceeds. Such a graph is known as an indicator diagram (Fig. 7.1).
Suppose the gas starts at a pressure and volume indicated by the point A and then expands until its pressure and volume are the coordinates of the point B. The course of the expansion is described by the curve AB.

The reader will already be familiar with the result that the work, $W$, done in such an expansion is given by

$$W = \int_A^B p \, dv.$$  

This is seen to be the area between the $p$–$v$ curve and the $v$-axis, i.e. the area LABM shown shaded in Fig. 7.1. This area depends not only on the positions of A and B on the diagram but also on the shape of the curve between A and B. In other words, the work done during an expansion does not depend solely on the initial and final states of the gas but also on the way in which the expansion takes place.

There are an infinite number of ways in which a gas may expand, but two special cases are of extreme importance: isothermal and adiabatic expansions.

An isothermal expansion is one in which the temperature of the gas remains constant throughout the change. For a perfect gas this means that $pv$ is a constant ($K$), so that the indicator diagram is a graph of the curve $p = K/v$ (Fig. 7.2). As the temperature remains constant, the internal energy also remains constant (again assuming a perfect gas), so that in the mathematical expression of the first law (equation (7.1)),

$$\delta U = 0,$$

i.e.

$$\delta Q = \delta W;$$

all the heat supplied is converted into work.

The work done in a reversible isothermal expansion of a perfect gas, in which the volume changes from $v_1$ to $v_2$, may be found as follows.

Work done,

$$W = \int_{v_1}^{v_2} p \, dv$$

$$= \int_{v_1}^{v_2} \frac{K}{v} \, dv$$

$$= K \ln \left( \frac{v_2}{v_1} \right).$$

The constant, $K$, depends on the mass of gas and the temperature; $K = RT$ for 1 gram-molecule of a gas at temperature $T^\circ$K.

An adiabatic expansion is one in which no heat enters or leaves the
system during the change, i.e. $\delta Q = 0$, and therefore $\delta U + \delta W = 0$. This means that the work done by the gas is entirely at the expense of the internal energy, and so the temperature falls. For a perfect gas the volume and pressure are related by $p\nu^\gamma = \text{constant}$, where $\gamma$ is the ratio of the principal specific heats of the gas, $C_p/C_v$. Part of an adiabatic curve is shown in Fig. 7.2. It may be shown that at the intersection of an isothermal and an adiabatic curve the latter always has the steeper slope. The proof of this is left as an exercise for the student.

The work done in a reversible adiabatic expansion of a perfect gas, in which the pressure, volume and temperature change from $p_1$, $v_1$, $T_1$ to $p_2$, $v_2$, $T_2$, respectively, is calculated as follows.

Work done

\[ W = \int_{v_1}^{v_2} \frac{p}{\nu^{\gamma}} \, dv \]

But

\[ p\nu^\gamma = C \quad \text{(a constant)} \]

\[ \therefore \quad W = \int_{v_1}^{v_2} \frac{C}{\nu^{\gamma}} \, dv \]

\[ = \left[ \frac{C\nu(1 - \gamma)}{1 - \gamma} \right]_{v_1}^{v_2} = \left[ \frac{p\nu}{1 - \gamma} \right]_{p_1,v_1}^{p_2,v_2} \]

\[ = \frac{1}{\gamma - 1} (p_1v_1 - p_2v_2). \]

For 1 gram-molecule of a gas, this may be written,

\[ W = \frac{R}{\gamma - 1} [T_1 - T_2]. \]

### 7.4 Reversible and Irreversible Changes

The equations for work done obtained in the previous section apply only if the changes are reversible.

In thermodynamics a change is reversible if it can be made to take place in the opposite direction by an infinitesimal change in the conditions under which it is taking place in the forward direction. An example will help to clarify this statement.

Consider the isothermal expansion of a gas. If this expansion is to be reversible, then an infinitesimal increase in the external pressure on the piston must cause the change to reverse, i.e. become a compression. This means that at no time must there be any appreciable difference
between the pressure inside the cylinder and the external pressure, and the piston must have no momentum; in other words, the change must occupy an infinite time. Further, the gas is absorbing heat during the expansion; on the change to compression it must give out heat. If this is to happen, then there must be no difference between the temperature of the source and that of the gas, so that the absorption of heat must also take an infinite time. It is obvious that a reversible change of this type can never occur in practice.

A further effect which prevents reversible changes occurring in practice is that of friction. Frictional forces always act against motion, so that if a certain change is being opposed by friction, then an infinitesimal change in the external conditions will not cause the change to reverse; the frictional force has still to be overcome in the opposite direction.

7.5 Cycles of Changes

In the example discussed in Section 7.3 and illustrated by Fig. 7.1 a gas was made to undergo an expansion by supplying heat to it, and as a result it performed a certain amount of external work and underwent a change in internal energy. If this expansion (assumed reversible) is now exactly reversed by supplying an equal amount of external work the gas gives out the same amount of heat that it absorbed during the expansion, and its internal energy returns to its original value. If, however, the conditions are changed in some way the gas can be returned to its original state at A, but by a different reversible path, such that the amount of work to be done on it is less and the heat given out also less. This is illustrated in Fig. 7.3.

When the gas expands along the path ACB it does work equal to the area ACBML, and when the gas contracts along the path BDA work is done on it equal to the area ADBML. The gas is now said to have undergone a cyclic process and ACBDA represents a cycle of changes. At the end of the cycle a net amount of external work has been done by the gas. This is given by the difference between the areas ACBML and ADBML, i.e. the shaded area ACBDA. The internal energy of the gas at the end of the cycle is the same as at the beginning, because the gas is in exactly the same state, so that one may deduce from the first law of thermodynamics that the work done is equal to the net amount of heat absorbed. Thus if \( Q_1 \) units of heat are absorbed by the gas along ACB
and $Q_2$ given out along BDA, and if a net amount $W$ units of work is done during the cycle, then

$$W = Q_1 - Q_2.$$ 

A machine constructed to act in such a cycle could continuously convert heat energy into mechanical energy, and is thus a heat engine. We have discussed the particular example of the expansion of a gas, but one is not limited to its use, and in general the substance which undergoes the cycle of changes is called the working substance. In practice, this is either steam (in the reciprocating engine or turbine), or air (in the internal-combustion engine).

### 7.6 Carnot Cycle

Before studying actual heat engines and their cycles of operation it is useful to consider an ideal heat engine which does not suffer from any mechanical or other imperfections. Such an ideal heat engine was described by a young French engineer, Sadi Carnot, in 1824, and the cycle in which this engine works is now known as the Carnot cycle. It is an extremely simple cycle, consisting of two reversible isothermal, and two reversible adiabatic, changes.

The ideal engine consists of a cylinder of perfectly insulating material closed at one end by a frictionless, perfectly insulating piston and at the other by a perfectly conducting base (Fig. 7.4).

The cylinder contains the working substance. Also available are a source at a constant high temperature $T_1$, a sink at a constant low temperature $T_2$ and a perfectly insulating stand. The indicator diagram for the cycle is shown in Fig. 7.5.

The cycle of operation is as follows.

(i) The working substance is in the state represented by A, at temperature

![Fig. 7.4](image)

![Fig. 7.5](image)
The cylinder is put in contact with the source and a reversible isothermal expansion allowed to take place, the temperature remaining constant at $T_1$. This is continued until the gas has the pressure and volume corresponding to the point B.

(ii) The cylinder is placed on the insulating stand and the reversible expansion allowed to continue until the temperature falls to $T_2$ (point C on diagram). This expansion is adiabatic, because the working substance is thermally isolated from its surroundings.

(iii) The cylinder is transferred to the sink, at temperature $T_2$, and the piston driven in under reversible conditions. The work supplied is converted into heat, which is transferred to the sink, the temperature remaining constant at $T_2$. This reversible isothermal compression is continued until the point D on the indicator diagram is reached. This is chosen such that if the compression is continued adiabatically the working substance returns to its original state at A.

(iv) The cylinder is moved on to the insulating stand and the reversible compression continued until the point A on the indicator diagram is reached.

Suppose the working substance takes in $Q_1$ units of heat from the source along AB and delivers $Q_2$ units of heat to the sink along CD. The net work done during the cycle is given by $Q_1 - Q_2$.

The efficiency of a heat engine is defined as:

$$\text{efficiency} = \frac{\text{work done during cycle}}{\text{heat absorbed at high temperature}}$$

Hence the efficiency, $\eta$, of a Carnot engine is given by

$$\eta = \frac{Q_1 - Q_2}{Q_1} = 1 - \frac{Q_2}{Q_1} \quad . \quad \quad (7.2)$$

### 7.7 The Carnot Refrigerator

If the Carnot cycle is performed in reverse, then a quantity of heat, $Q_2$, is abstracted from the sink at temperature $T_2$, and a quantity of heat, $Q_1$, transferred to the source at higher temperature $T_1$, by the performance of work $Q_1 - Q_2$ on the engine. The engine is now acting as a refrigerator or heat pump because it is removing heat from a cool object and giving heat to a hotter one. The effectiveness of the refrigerator is measured by its coefficient of performance, which is defined as:

$$\text{coefficient of performance} = \frac{\text{heat abstracted from cold body}}{\text{work done on engine}}$$

In the case of a Carnot refrigerator the coefficient of performance is

$$\frac{Q_2}{Q_1 - Q_2}.$$
7.8 The Second Law of Thermodynamics

The first law of thermodynamics tells us what happens when a certain quantity of heat is absorbed by a body and totally or partially converted into work; it does not tell us under what conditions this conversion of heat energy into work may take place. These conditions are stated in the second law.

We have seen that, in the performance of a Carnot cycle, it is necessary to have available not only a source of heat at a high temperature but also a sink at a lower temperature to which heat may be transferred during one stage of the cycle. Thus, not all the heat absorbed by the engine is converted into work, but only some fraction of it, the rest being rejected to the sink. It is impossible for the engine to operate if such a sink is not available. Experience shows that the same conditions apply to all heat-engine cycles that can be devised. If this were not so it would be possible, for example, to design an engine which would run simply by abstracting heat from its surroundings (e.g. the air) and converting all this heat into work; no such engine has been devised. For the completion of the cycle of such an engine it is necessary to have an object at a lower temperature than the air, to which heat can be delivered during part of the cycle.

These experiences may be expressed formally as follows:

No engine can be devised which, operating in a cycle, produces mechanical work by abstracting heat from the coldest part of its surroundings.

This is one way of stating the second law of thermodynamics and is based on the original statement of the law by Kelvin.

The phrase 'operating in a cycle' is a most important part of this statement; it is possible for an engine to abstract heat from the coldest part of its surroundings and completely convert this heat into work, but only in a single change which leaves the working substance in a different state at the end. For example, if a perfect gas in thermal contact with the coldest part of its surroundings undergoes a reversible isothermal expansion all the heat absorbed is converted into work, but at the end of the expansion the gas is not in the same state as at the beginning; its volume is greater and its pressure less. It is impossible to return the gas to its original state by any path other than the isothermal reversible one, as no sink at a lower temperature is available, and if it does retrace this path the net work done in the cycle is zero.

The second law of thermodynamics cannot be verified experimentally because it is a negative law, so that, no matter how many experiments one does to test its validity, there always remains the possibility that an engine which does contravene the law may yet be devised. Neither is it possible to deduce the law from other known laws. Nevertheless, the law is very strongly held to be true for two reasons: first, no engine which contradicts it has yet been devised; and secondly, all deductions which have been made from the second law have been found to hold good.
The law may be expressed in an equivalent form, due to Clausius. This statement is a formal expression of the common experience that heat cannot flow from a cold body to a hotter one, thus:

_No self-acting engine, operating in a cycle, can transfer heat from one body to another at a higher temperature._

Again notice two important conditions which have been included in this statement; first, the device must be self-acting, i.e. it must not have connection with any outside agency (e.g. a motor); and secondly, the device must operate in a cycle.

It is not immediately obvious that the two statements given above are alternative ways of saying the same thing; the proof that they are equivalent will not be given here, but it may be found in Zemansky (see bibliography).

### 7.9 Carnot’s Theorem

Carnot’s theorem may be stated as follows:

_No heat engine operating between two given temperatures can be more efficient than a reversible engine operating between the same two temperatures._

This theorem is proved by showing that if it were not true, then an engine could be devised which would violate the second law. Consider a reversible engine, R, and any other engine, S, designed so that they both produce the same amount of work, $W$, in one cycle when they operate between the same two reservoirs. Suppose that R takes in heat $Q_1$ from the hot reservoir, produces work $W$ and rejects heat $Q_1 - W$ to the cold reservoir; its efficiency is thus $W/Q_1$. The corresponding quantities for S are denoted by $Q'_1$, $W$ and $Q'_1 - W$; its efficiency is $W/Q'_1$.

Fig. 7.6 shows R being driven backwards by the work supplied by S. As R is reversible, it absorbs heat $Q_1 - W$ from the cold reservoir and delivers heat $Q_1$ to the hot reservoir in each reverse cycle. Now suppose that S is more efficient than R, i.e. $W/Q'_1 > W/Q_1$, which means that $Q_1 > Q'_1$. The combination R–S is a self-acting engine which abstracts heat $(Q_1 - W) - (Q'_1 - W) = Q_2 - Q'_1$ from the cold reservoir and rejects heat $Q_1 - Q'_1$ to the hot reservoir in each cycle. This is a positive quantity of heat because $Q_1 > Q'_1$, and hence the engine contravenes the Clausius’ statement of the second law.

The assumption that S is more efficient than R thus leads to a contradiction of the second law, and hence Carnot’s theorem is proved.

An important corollary of Carnot’s theorem is:
All reversible engines working between the same two temperatures have the same efficiency.

This may be proved by following the above argument using two reversible engines $R_1$ and $R_2$. It may then be shown that $R_1$ is not more efficient than $R_2$ and, by interchanging the two engines, that $R_2$ is not more efficient than $R_1$. It follows that $R_1$ and $R_2$ have the same efficiency.

We have now shown that the efficiency of a reversible engine does not depend on the working substance, but only on the temperatures of the reservoirs between which it works. This important fact suggests a way of defining a temperature scale which does not depend on the properties of a particular substance. The way this has been done is described in the next section.

7.10 Kelvin Absolute Temperature Scale

The Kelvin temperature scale, also known as the thermodynamic scale of temperature, is defined with reference to a reversible engine working between two reservoirs. The efficiency, $\eta$, of this engine is given by

$$\eta = \frac{Q_1 - Q_2}{Q_1} = 1 - \frac{Q_2}{Q_1}$$

(see equation (7.2))

where $Q_1, Q_2$ have their usual meanings.

$$\therefore \frac{Q_1}{Q_2} = \frac{1}{1 - \eta}$$

We have seen in the previous section that the efficiency of the engine depends only on the temperatures of the two reservoirs between which it works, so that the right-hand side of the above equation is a function of these two temperatures only. If we denote the temperatures of the hot and cold reservoirs by $\theta_1, \theta_2$, respectively, then we may write

$$\frac{Q_1}{Q_2} = f(\theta_1, \theta_2),$$

where $f(\theta_1, \theta_2)$ is some function of $\theta_1$ and $\theta_2$. It would be possible to use this equation to define any number of different temperature scales by suitable choice of the function $f(\theta_1, \theta_2)$. The Kelvin scale is defined by putting $f(\theta_1, \theta_2) = \theta_1/\theta_2$.

$$\therefore \frac{Q_1}{Q_2} = \frac{\theta_1}{\theta_2}$$

(7.3)

It should be emphasized that this is merely an arbitrary choice, but a very convenient one, as we shall see later. The temperatures $\theta_1, \theta_2$ are then said to be measured on the Kelvin temperature scale and are denoted by °K. To complete the scale it is necessary to fix the size of the degree.

This may be done by having two fixed points, the ice point and the steam point, and representing the interval between them by 100 deg.
Because of the practical difficulties involved in achieving a precise standard for the ice point, this method has been superseded by one in which only one fixed point is needed. In 1954 it was agreed to use the triple point of water as the fixed point and to denote this temperature by 273.16°K. To establish a temperature scale in this way it is merely necessary to define the ratio between two temperatures in terms of a property of the thermometric substance.

For example, for the constant-volume gas scale one defines the ratio between two temperatures $T_1$ and $T_2$ by the equation

$$\frac{T_1}{T_2} = \frac{p_1}{p_2},$$

where $p_1$, $p_2$ are the pressures of the gas at the temperatures $T_1$, $T_2$ respectively. In particular, if $P_t$ denotes the pressure of the gas when at the temperature of the triple point of water, and $p$ the pressure at temperature $T$, then

$$\frac{T}{273.16} = \frac{p}{P_t},$$

which defines the scale.

To define a temperature, $\theta$, on the Kelvin scale (which is not referred to a property of any substance) equation (7.3) is used. This becomes

$$\frac{\theta}{273.16} = \frac{Q}{Q_t},$$

where $Q$ is the heat absorbed from a reservoir at temperature $\theta$, and $Q_t$ the heat rejected to a reservoir at the temperature of the triple point of water.

### 7.11 Relation between the Kelvin Scale and the Perfect Gas Scale

Suppose that the working substance in a Carnot engine is one gram-molecule of a perfect gas. Fig. 7.7 shows one complete cycle of the en-

![Diagram](image-url)

Fig. 7.7

engine. AB and CD are isothermals, BC and DA are adiabatics. Let the temperatures along AB and CD be $T_1$, $T_2$ respectively, measured on the perfect gas scale.
THERMODYNAMICS

Work done by gas along $AB = RT_1 \ln \left( \frac{V_2}{V_1} \right)$

$= Q_1$.

Work done by gas along $CD = RT_2 \ln \left( \frac{V_4}{V_3} \right)$

$= -Q_2$

$\therefore Q_2 = RT_2 \ln \left( \frac{V_3}{V_4} \right)$.

$\therefore \frac{\theta_1}{\theta_2} = \frac{Q_1}{Q_2} = \frac{T_1}{T_2} \left( \frac{\ln \frac{V_2}{V_1}}{\ln \frac{V_3}{V_4}} \right)$,

where $\theta_1$, $\theta_2$ are the temperatures along $AB$, $CD$, respectively, measured on the Kelvin scale.

Now, for the isothermal $AB$: $p_1v_1 = p_2v_2$,

and for the adiabatic $BC$: $p_2v_2^\gamma = p_3v_3^\gamma$.

For the isothermal $CD$: $p_3v_3 = p_4v_4$,

and for the adiabatic $DA$: $p_4v_4^\gamma = p_1v_1^\gamma$.

Multiplying all these equations together gives

$v_1v_2(v_2v_4)^\gamma = v_2v_4(v_1v_3)^\gamma$,

$\therefore v_1v_2 = v_2v_4$,

$\therefore \frac{v_3}{v_4} = \frac{v_2}{v_1}$

$\therefore \frac{\ln \frac{v_2}{v_1}}{\ln \frac{v_3}{v_4}} = 1$,

so that

$\frac{\theta_1}{\theta_2} = \frac{T_1}{T_2}$.

Let $T$ refer to any temperature measured on the perfect gas scale, and $\theta$ refer to the same temperature measured on the Kelvin scale. Let $T_t$, $\theta_t$ denote the temperature of the triple point of water on the two scales.

From the result just obtained it follows that

$\frac{\theta}{\theta_t} = \frac{T}{T_t}$.

But, by definition, $\theta_t = T_t = 273.16$.

$\therefore \theta = T$,

and the two scales are identical.
We now drop the use of $\theta$, and use $T$ to denote temperature measured on the Kelvin scale.

### 7.12 Entropy

Suppose that, in a certain Carnot cycle, the working substance takes in heat $Q_1$ at temperature $T_1$, and gives out heat $Q_2$ at temperature $T_2$. Then, from the definition of the Kelvin scale of temperature,

$$\frac{Q_1}{Q_2} = \frac{T_1}{T_2},$$

$$\therefore \frac{Q_1}{T_1} - \frac{Q_2}{T_2} = 0.$$

If we now use positive numbers to denote heat absorbed and negative numbers to denote heat given out, this equation becomes

$$\frac{Q_1}{T_1} + \frac{Q_2}{T_2} = 0,$$

where the negative sign has been included in the $Q_2$.

Now consider any reversible cycle (not necessarily a Carnot cycle) such as the one represented by the continuous curve in Fig. 7.8.

The dotted lines (such as $aa'$) crossing the cycle are reversible adiabatics and the short lines between them (such as $ab$) are portions of reversible isothermals.

Cycles like $aa'b'b$ are Carnot cycles, and if we denote the heat taken in at temperature $T_{ab}$, along $ba$, by $Q_{ab}$, and that taken in at temperature $T_{ab'}$ along $a'b'$ by $Q_{a'b'}$, then

$$\frac{Q_{ab}}{T_{ab}} + \frac{Q_{a'b'}}{T_{a'b'}} = 0.$$

Writing similar equations for the other Carnot cycles and adding gives

$$\frac{Q_{ab}}{T_{ab}} + \frac{Q_{cd}}{T_{cd}} + \ldots + \frac{Q_{a'b'}}{T_{a'b'}} + \frac{Q_{c'd'}}{T_{c'd'}} + \ldots = 0 \quad (7.5)$$

Now if we take the working substance round the complete zig-zag cycle $a'b'c'd'\ldots dcbaa'$ we shall include all the isothermals of equation (7.5). There is no heat transfer along the adiabatics, so this equation must hold for the complete zig-zag cycle; it may be written

$$\sum_{T} Q_{T} = 0.$$
By increasing the number of adiabatics the zig-zag cycle may be made as near as one likes to the continuous curve. The heat transfer along each isothermal becomes very small (δQ), so that we may write
\[ \sum \frac{\delta Q}{T} = 0. \]

In the limit, as the lengths of the isothermals tend to zero, δQ \rightarrow 0 and the equation becomes
\[ \oint \frac{dQ}{T} = 0, \text{ for any reversible cycle.} \]
(The symbol \( \oint \) indicates integration round a closed path.)

This important result (known as Clausius' theorem) leads to the concept of entropy as follows.

In Fig. 7.9 \( R_1, R_2 \) represents two possible reversible paths between the states represented by A and B. The cycle formed by going from A to B along \( R_1 \) and then back to A along \( R_2 \) is reversible, so that, according to Clausius' theorem,
\[ \int_{A}^{B} \left( \frac{dQ}{T} \right)_{R_1} + \int_{B}^{A} \left( \frac{dQ}{T} \right)_{R_2} = 0, \]
where the first integration is along path \( R_1 \) and the second along \( R_2 \).

But
\[ \int_{B}^{A} \left( \frac{dQ}{T} \right)_{R_2} = -\int_{A}^{B} \left( \frac{dQ}{T} \right)_{R_1}, \]
so that
\[ \int_{A}^{B} \left( \frac{dQ}{T} \right)_{R_2} = \int_{A}^{B} \left( \frac{dQ}{T} \right)_{R_1}. \]

The value of \( \int_{A}^{B} \frac{dQ}{T} \) is seen to be independent of the reversible path between A and B along which it is calculated, and is thus a function only of the states of the substance at A and B. It is called the change of entropy between A and B. The zero of entropy is arbitrary (cf. the zero of energy), but we may write
\[ S_B - S_A = \int_{A}^{B} \frac{dQ}{T}. \]  \hspace{1cm} (7.6)
where \( S_A, S_B \) are the entropies at A and B respectively.

It is very important that the integration in equation (7.6) be calculated along a reversible path between A and B. We have seen that the entropy change of a substance depends only on its initial and final states, so that, even if the path taken is an irreversible one, the change in
entropy is still the same as if the change had been reversible. The entropy change along an irreversible path is calculated by finding the value of $\int \frac{dQ}{T}$ along any reversible path between the initial and final states, but this is not equal to the value of $\int \frac{dQ}{T}$ along the actual (irreversible) path.

7.13 Reversible Processes

We have seen that in a reversible cycle the value of $\int \frac{dQ}{T}$ is zero, i.e. the change in entropy of the working substance is zero. Let us now consider what happens in a single reversible process, so that the state of the substance at the end of the process is not the same as that at the beginning. Obviously the entropy of the substance changes during the process.

Suppose that, at a certain stage in the process, the substance is at temperature $T$ and absorbs heat $\delta Q$ from a reservoir also at temperature $T$ (a condition of reversibility). The entropy gained by the substance is $\delta Q/T$ and the entropy lost by the reservoir is $\delta Q/T$. Thus the net change of entropy of the universe (i.e. substance + surroundings) is zero. Similarly, if the substance rejects heat it loses a certain amount of entropy and the surroundings gain an equal amount: the change in entropy of the universe is again zero. In any reversible adiabatic stages there is no exchange of heat with the surroundings, and hence no change in entropy of either the substance or the surroundings.

We have shown that there is no change in the entropy of the universe at any stage of a reversible process, so that

$$\Delta S = 0$$

for any reversible process,

where $\Delta S$ stands for the change in entropy of the universe.

7.14 Entropy Change in an Irreversible Cycle

It will be shown that the performance of any irreversible cycle results in an increase in the entropy of the universe.

First we must show that the efficiency of any engine working in an irreversible cycle between two reservoirs is less than that of a reversible engine working between the same two reservoirs. In Section 7.9 it was shown that no engine can be more efficient than a reversible one, so that we now have to prove that the irreversible one cannot be as efficient as the reversible one.

Suppose that the engine $S$ in Section 7.9 works in an irreversible cycle and has the same efficiency as $R$. This assumption leads to $Q_1 = Q_1'$, so that in one cycle the combination $R-S$ absorbs no heat from either reservoir, rejects no heat to either reservoir and does no external work. At the end of the cycle, then, the surroundings have undergone no change and the engine $R-S$ is in the same state as at the beginning.
This implies that R–S is reversible, which is not true, because S is irreversible; the assumption that R and S can have the same efficiency is false. We may now conclude that S is less efficient than R.

Now, for a reversible engine which absorbs heat $Q_1$ at temperature $T_1$ and rejects heat $Q_2$ at temperature $T_2$:

$$\frac{Q_1}{Q_2} = \frac{T_1}{T_2}$$

by the definition of the Kelvin scale. Thus for the efficiency, $\eta_R$, of a reversible engine we may write

$$\eta_R = 1 - \frac{Q_2}{Q_1} = 1 - \frac{T_2}{T_1}.$$

Suppose that an engine working in an irreversible cycle between the same two reservoirs absorbs heat $Q_1'$ and rejects heat $Q_2'$; its efficiency

$$\eta_I = 1 - \frac{Q_2'}{Q_1'}.$$

But

$$\eta_I < \eta_R,$$

$$\therefore 1 - \frac{Q_2'}{Q_1'} < 1 - \frac{T_2}{T_1},$$

$$\therefore \frac{Q_1'}{T_1} < \frac{Q_2'}{T_2}.$$

At the end of such a cycle the hot reservoir has lost entropy $Q_1'/T_1$, the cold reservoir has gained entropy $Q_2'/T_2$ and the entropy of the working substance is unchanged. Thus the total gain in the entropy of the universe,

$$\Delta S = \frac{Q_2'}{T_2} - \frac{Q_1'}{T_1} > 0.$$ 

Thus $\Delta S > 0$ for an irreversible cycle.

### 7.15 Irreversible Processes – The Principle of Increase of Entropy

A common irreversible process is one in which heat $Q$ flows from a body at temperature $T_1$ to a body at temperature $T_2$ ($T_1 > T_2$). In this case

$$\Delta S = \frac{Q}{T_2} - \frac{Q}{T_1} > 0.$$ 

We now proceed to show that $\Delta S > 0$ applies to all irreversible processes.

Suppose that a substance undergoes any irreversible process. It is possible to bring the substance back to its original state by a combination of reversible adiabatic and isothermal processes* so that it has com-

* This statement is easily seen to be true for the expansion and compression of gases, but in other cases (e.g. chemical changes) it needs more careful verification, which it is not possible to give here.
pleted a cycle. This cycle is an irreversible one (because one part of it is irreversible) so that for the complete cycle $\Delta S > 0$. But, for all the reversible processes in the cycle $\Delta S = 0$,

so that $\Delta S > 0$ for an irreversible process.

This important conclusion is known as the principle of the increase of entropy and shows that in any natural process the entropy of the universe always increases. The principle has been deduced from Carnot's theorem, which is a consequence of the second law of thermodynamics.

### 7.16 Unavailable Energy

Suppose that a quantity of heat, $Q$, is available at temperature $T_1$ and that the temperature of the coldest reservoir available is $T_0$. The maximum amount of work which can be obtained is that produced by a reversible engine working between the temperatures $T_1$ and $T_0$,

$$W_{\text{max.}} = \eta R Q$$
$$= \left( \frac{T_1 - T_0}{T_1} \right) Q$$
$$= Q \left( 1 - \frac{T_0}{T_1} \right).$$

This is a fraction of $Q$, the remainder of the heat (i.e. $\frac{T_0}{T_1} Q$) now being at temperature $T_0$. As this is the lowest temperature available, this heat cannot be used to operate a heat engine and is unavailable energy.

Further, it may be shown that whenever an irreversible process takes place a certain quantity of energy becomes unavailable. The complete proof of this is too difficult to include here, but we give one example, that of conduction.

Suppose that we have available three reservoirs at temperatures $T_1$, $T_2$, $T_0$ ($T_1 > T_2 > T_0$) and that $T_0$ is the lowest attainable temperature. Now if there is a quantity of heat $Q$ available at the highest temperature $T_1$ the maximum amount of work into which it may be converted is

$$W_1 = Q \left( 1 - \frac{T_0}{T_1} \right), \text{ as above.}$$

If the heat is first conducted from the reservoir at $T_1$ to that at $T_2$ the maximum amount of work available is

$$W_2 = Q \left( 1 - \frac{T_0}{T_2} \right).$$

$$\therefore \quad W_1 - W_2 = Q \left( \frac{T_0}{T_2} - \frac{T_0}{T_1} \right)$$
$$= Q T_0 \left( \frac{1}{T_2} - \frac{1}{T_1} \right)$$
$$> 0.$$
As a result of the conduction process, then, the amount of energy available is decreased by

\[ QT_0 \left( \frac{1}{T_2} - \frac{1}{T_1} \right) \]

\[ = T_0 \left( \frac{Q}{T_2} - \frac{Q}{T_1} \right) \]

\[ = T_0 \Delta S, \]

where \( \Delta S \) is the change in entropy of the universe caused by the conduction process. This result, that the decrease in available energy is equal to \( T_0 \) times the increase in entropy of the universe produced by the change, is true of all irreversible changes.

All the natural processes that take place, then, result in a decrease in the store of available energy in the universe. It must be emphasized that this is not a contradiction of the principle of conservation of energy. The total amount of energy in the universe remains constant; but the portion of it which is available for use is decreasing, whilst that which is unavailable is increasing.

### 7.17 Theoretical Use of Reversible Cycles

It has been shown that all reversible engines working between the same two temperatures have the same efficiency, which is given by

\[ \eta = \frac{Q_1 - Q_2}{Q_1} = \frac{T_1 - T_2}{T_1}, \]

the symbols having their usual meanings.

If we put

\[ Q_1 = Q \quad ; \quad T_1 = T; \]

\[ Q_2 = Q - \delta Q \quad ; \quad T_2 = T - \delta T; \]

then

\[ \eta = \frac{\delta Q}{Q} = \frac{\delta T}{T}. \]

But the amount of work, \( \delta W \), done in the cycle is equal to \( \delta Q \), provided these are measured in the same units.

Hence

\[ \frac{\delta W}{Q} = \frac{\delta T}{T} \quad . \quad . \quad . \quad . \quad (7.7) \]

Two examples will now be given of the application of this equation.

### 7.18 The Clausius–Clapeyron Equation

Fig. 7.10 shows two isothermals of unit mass of a substance below its critical temperature. The upper one is for temperature \( T \) and the lower for \( T - \delta T \). Let the difference between the saturated vapour pressures of the liquid at these two temperatures be \( \delta \phi \). \( v_2 \) and \( v_1 \) denote the volumes of 1 g of vapour at C, and 1 g of liquid at D, respectively. Through D and C are drawn two reversible adiabatics, which intersect the lower curve at points close to A and B, respectively.
The cycle ABCD is now a reversible one, so that we may apply equation (7.7).

The heat supplied along DC is that required to vaporize 1 g of the substance, i.e. the latent heat \( L \text{ cal g}^{-1} \) at \( T^\circ\text{K} \).

\[ \therefore Q = L \text{ cal} = LJ \text{ erg.} \]

The work done, \( \delta W \), in the cycle is the area of the figure ABCD, which is approximately \( \delta p(v_2 - v_1) \) ergs.

![Diagram](image)

**Fig. 7.10**

Hence, substituting in equation (7.7) gives

\[ \frac{\delta p(v_2 - v_1)}{LJ} = \frac{\delta T}{T}. \]

If now \( \delta T \to 0 \), the equation becomes

\[ \frac{dp}{dT} = L \frac{T}{(v_2 - v_1)}. \]

This is known as the Clausius–Clapeyron equation.

**Example.** Saturated vapour pressure of water at 99.5\(^\circ\text{C} \) = 746.5 mm of Hg; saturated vapour pressure of water at 100.5\(^\circ\text{C} \) = 773.7 mm of Hg;

\[ \therefore \text{ at } T = 373^\circ\text{K}, \frac{dp}{dT} \simeq 773.7 - 746.5 \simeq 27.2 \text{ mm of Hg}/^\circ\text{K} \simeq 2.72 \times 13.6 \times 981 \text{ dyn cm}^{-2} \text{ deg K}; \]

volume of 1 g of water at 100\(^\circ\text{C} \) = 1 cc;

volume of 1 g of steam at 100\(^\circ\text{C} \) = 1,670 cc.

Substitution of these values in the Clausius–Clapeyron equation yields \( L = 540 \text{ cal g}^{-1} \).
The Clausius–Clapeyron equation may also be applied to melting, \( v_1 \) now being the volume of 1 g of solid and \( v_2 \) that of 1 g of liquid. \( L \) is the latent heat of fusion at \( T^\circ \text{K} \).

If the substance contracts on melting (e.g. ice) \( v_1 \) is greater than \( v_2 \) and \( \frac{dp}{dT} \) is negative, so that an increase in pressure lowers the melting point. On the other hand, for those substances which expand on melting \( v_2 \) is greater than \( v_1 \) and \( \frac{dp}{dT} \) is positive, so that an increase in pressure raises the melting point. This is a conclusion that is completely borne out in practice.

### 7.19 Surface Energy

In Section 2.3 it was shown that the work done to form a film of area \( A \) cm\(^2\) and surface tension \( \gamma \) dyn cm\(^{-1}\) is \( \gamma A \) ergs. The surface energy of the film is this work, plus the heat absorbed when the film is formed isothermally. We now proceed to find an expression for this heat. Fig. 7.11 represents a reversible cycle performed on a liquid film.

![Diagram](image-url)

Fig. 7.11

Starting at A the film is stretched isothermally at temperature \( T \) until point B is reached, its area having increased by \( \delta A \). Suppose that the heat absorbed during this process is \( Q \). The film is then further stretched adiabatically until its temperature has fallen to \( T - \delta T \) (point C). The film is now allowed to contract isothermally until the point D is reached such that when the film is allowed to continue to contract adiabatically point A is reached. The cycle ABCD is a reversible cycle, and so equation (7.7) applies. Work is done on the film along AB and by the film along CD, so that the total work done by the film in the cycle

\[
= - \text{(area of cycle)}
= -\gamma A . \delta A.
\]

\[\therefore \text{substituting in equation (7.7),}\]

\[
\frac{\gamma \delta A}{Q} = \frac{\delta T}{T'}
\]

\[\therefore \frac{Q}{\delta A} = - T \frac{\delta \gamma}{\delta T}.\]
Now $Q/\delta A$ is the heat absorbed per unit area of new surface formed ($= q$, say).

If $\delta T \to 0$, the above equation becomes

$$q = -T \frac{d\gamma}{dT}.$$ 

$d\gamma/dT$ is a negative quantity (surface tension decreases with rise in temperature), so that $q$ is a positive quantity.

Total surface energy per unit area $= \gamma - T \frac{d\gamma}{dT}$. 

LOW TEMPERATURES

8.1 Introduction

One of the most fruitful fields of research at the present time is the study of the properties of the solid state, and much of this work has been conducted at very low temperatures. Methods of attaining and measuring such temperatures are, therefore, very important, and in this chapter a survey of the techniques used will be given. It is hoped that the reader will be encouraged to pursue the subject in more detail by consulting the books mentioned in the bibliography.

A preliminary to any method of producing low temperatures is the liquefaction of certain gases: e.g. air, hydrogen and helium. We begin with an outline of some of the liquefaction processes used.

LIQUEFACTION OF GASES

8.2 Early Methods

Some gases (e.g. ammonia and sulphur dioxide) may easily be liquefied by sufficient increase in pressure at room temperature. Early workers had assumed that all gases could be liquefied in this way, provided a high enough pressure were reached. The method of procedure was simply to cool the gas somewhat, apply the greatest possible pressure and hope that liquefaction would take place. It soon became apparent that certain gases (e.g. oxygen, nitrogen and hydrogen) would not respond to this treatment, and for this reason they were often referred to as the permanent gases.

Following the work by Andrews on carbon dioxide, it was realized that the reason for the failure of attempts to liquefy certain gases was that they had not been cooled below their critical temperatures. Subsequent research was therefore directed towards finding methods of cooling each of the ‘permanent’ gases below their respective critical temperatures so that the application of pressure would then liquefy the gas.

In 1877–8 Pictet developed a method of liquefying oxygen by producing cooling in stages using a series of cooling agents. A gas, which had a critical temperature well above room temperature, was liquefied by the application of pressure only; this liquid was then boiled under reduced pressure and the resultant cooling effect used to bring a second gas well below its critical temperature. This second gas was then liquefied by compression, boiled under reduced pressure and used to cool the oxygen below its critical temperature. An increase in pressure
then resulted in liquefaction of the oxygen. The process is indicated schematically in Fig. 8.1. This method, often known as the cascade process, has also been used to liquefy air. Series of gases which have been used by various workers include:

sulphur dioxide; carbon dioxide; oxygen (Pictet);
sulphur dioxide; ethylene; oxygen (continuous production of liquid oxygen);
methyl chloride; ethylene; air (continuous production of liquid air).

![Diagram](image)

**Fig. 8.1**

Table 8.1 shows the boiling points and critical temperatures of these gases.

<table>
<thead>
<tr>
<th>Gas</th>
<th>Critical temperature, °K</th>
<th>Normal boiling point, °K</th>
<th>Inversion temperature, °K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulphur dioxide</td>
<td>428</td>
<td>265</td>
<td></td>
</tr>
<tr>
<td>Carbon dioxide</td>
<td>304</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methyl chloride</td>
<td>416</td>
<td>249</td>
<td></td>
</tr>
<tr>
<td>Ethylene</td>
<td>283</td>
<td>170</td>
<td></td>
</tr>
<tr>
<td>Oxygen</td>
<td>155</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>Nitrogen</td>
<td>127</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>Hydrogen</td>
<td>33</td>
<td>20</td>
<td>~ 200</td>
</tr>
<tr>
<td>Helium</td>
<td>5.2</td>
<td>4.2</td>
<td>~ 30</td>
</tr>
</tbody>
</table>

**Table 8.1**

The cascade process is not applicable to the liquefaction of hydrogen, because the lowest temperature that can be produced by boiling liquid nitrogen under reduced pressure is above the critical temperatures of
the remaining 'permanent' gases (neon, hydrogen, helium). An alternative method of cooling must therefore be employed; this is discussed in the next section.

8.3 The Joule–Kelvin Effect

It was discovered by Joule and Lord Kelvin that when a gas is allowed to expand through a porous plug the temperature of the gas changes. The gas was maintained at a constant pressure, \( p_1 \), on one side of the plug by means of a compressor, and at pressure \( p_2 \) (<\( p_1 \)) on the other side. The temperatures on each side of the plug were measured when the conditions had become steady.

The results of these, and later, experiments show that:

(i) for most gases at ordinary temperatures there is a cooling effect, but for hydrogen and helium a rise in temperature results;
(ii) the change in temperature is proportional to \( p_1 - p_2 \);
(iii) all gases exhibit a cooling effect if below a certain temperature (different for each gas), and a heating effect if above that temperature. This temperature is known as the inversion temperature (see Table 8.1).

We are really concerned here with the application of the Joule–Kelvin effect to the liquefaction of gases, but a few words will first be said about the theory of the effect. What follows is not necessary to the main development of this chapter, and on a first reading the student may wish to continue with Section 8.4.

The change in energy of the gas during the throttling process through the porous plug may be due to one or both of two causes: external work and internal work. The latter arises if there is a force of attraction between the molecules.

Suppose that volume \( v_1 \) of gas at pressure \( p_1 \) on one side of the plug becomes \( v_2 \) at pressure \( p_2 \) after the throttling. The external work done on the gas by the compressor in passing it through the plug is \( p_1 v_1 \) and the work done by the gas on the low pressure side is \( p_2 v_2 \). Thus the net amount of external work done on the gas is \( p_1 v_1 - p_2 v_2 \).

The internal work, \( u \), is that required to separate the molecules against their mutual attractions. This results in a decrease in the kinetic energy of the molecules of the gas. Thus the total heat energy gained by the gas is

\[
Q = p_1 v_1 - p_2 v_2 - u
\]  

If the gas is to be maintained at constant temperature an amount of heat energy equal to \( Q \) must be removed from the gas on throttling. It follows that if \( Q \) is positive, and heat is not removed, the temperature rises; if \( Q \) is negative, and heat is not supplied, the gas cools.

The inversion temperature is characterized by a zero value of \( Q \), temperatures above inversion (heating effect) by a positive value of \( Q \) and temperatures below (cooling effect) by a negative value of \( Q \).
Note that for a perfect gas $Q$ is zero at all temperatures, because $p_1v_1 = p_2v_2$ (Boyle's Law), and $u = 0$ (no inter-molecular attraction). The Joule–Kelvin effect occurs only because of departures of real gases from ideal.

For real gases there is a temperature known as the **Boyle temperature** at which $p_1v_1 = p_2v_2$ over a wide range of pressures. Above this temperature $p_1v_1 > p_2v_2$, and below it $p_1v_1 < p_2v_2$ (provided the pressure is not high). It follows that if a gas is below its Boyle temperature, $Q$ in equation (8.1) is necessarily negative and the Joule–Kelvin effect is a cooling.

If a gas is above its Boyle temperature the sign of $Q$ depends on whether $p_1v_1 - p_2v_2$ is greater or less than $u$. If $(p_1v_1 - p_2v_2) < u$, $Q$ is negative and the effect is a cooling, and if $(p_1v_1 - p_2v_2) > u$ there will be a rise in temperature during the throttling process.

The inversion temperature is that temperature for which $p_1v_1 - p_2v_2 = u$.

### 8.4 The Hampson and Linde Processes

Both the Hampson and Linde processes for liquefaction of gases use a throttling process as the means of producing cooling. The porous plug of the Joule–Kelvin experiment is replaced by a special valve. The cooling produced by the Joule–Kelvin effect is relatively small, e.g. for air the cooling is about 0.2°C per atmosphere change in pressure at room temperature, but this increases at lower temperatures. To obtain sufficient cooling to produce liquefaction it is necessary to use a *regenerative process*, in which the high-pressure gas is cooled by the low-pressure gas which has previously undergone the throttling process. This is accomplished in a *heat exchanger* in which the two streams of gas are caused to pass through coils of copper tube placed in very close contact over a large area; often the two tubes are coaxial. This is indicated schematically in Fig. 8.2. After a time the temperature of the high-pressure gas is low enough for some of it to liquefy after throttling. Only a fraction of the gas liquefies; the rest is used to cool the incoming gas in the heat exchanger. The preliminary cooling is necessary to remove the heat produced during compression in the pump.
LOW TEMPERATURES

In the Hampson process for liquefaction of air this cooling is by water at room temperature, but in the Linde process (for air) there is further preliminary cooling (to about $-40^\circ$C) in liquid ammonia boiling under reduced pressure.

The other basic difference between the two processes is that, whereas in the Hampson process the low-pressure side is at about 1 atmosphere, in the Linde process it is about 40 atmospheres. In both cases the high pressure is in the region of 200 atmospheres. The smaller drop in pressure in the Linde process results in slightly less cooling at each expansion, but this is more than offset by the large saving in energy accomplished by having to compress the air only from 40 to 200 atmospheres. The Linde process is much more efficient than the Hampson process.

The Linde process is also used to liquefy hydrogen. In this case, to obtain the hydrogen below its inversion temperature pre-cooling in liquid nitrogen is used. Care must be taken to ensure that the hydrogen is free from all impurities (water vapour, oxygen, etc.) which would freeze and block the apparatus.

Helium may also be liquefied by this means, the pre-cooling being in liquid hydrogen boiling under reduced pressure.

8.5 The Claude and Heylandt Processes

A big disadvantage inherent in using the Joule–Kelvin effect for liquefying hydrogen and helium is that pre-cooling is necessary in order to get the gases below their inversion temperatures. This involves the use of other liquefied gases.

Another way of cooling a gas is to allow it to undergo an adiabatic expansion and perform external work. This may be accomplished either by causing the gas to push against a piston in a cylinder or against the blade of a turbine. A cooling always results, whatever the initial

---

Fig. 8.3
temperature of the gas; no pre-cooling is necessary. Cooling by adiabatic expansion is the basis of the Claude and Heylandt processes.

Fig. 8.3 illustrates the principle of an air liquefier using the Claude method. Purified air at high pressure passes through a heat exchanger (A) where it is cooled by the low-pressure gas returning to the pump. The stream of high-pressure gas then divides, a fraction (less than half) of it going through the cylinder (C), where it undergoes adiabatic expansion, performs external work and hence cools; the work produced is used to supply some of the energy to operate the compressor. The expanded gas then passes through the heat exchanger B and cools the remaining high-pressure gas. The conditions are so adjusted that the gas from the expansion cylinder is not at a low enough temperature to liquefy at its pressure (1 atmosphere), but is cold enough to cause the high-pressure gas to liquefy. The low-pressure gas from the heat exchanger B passes through heat exchanger A to cool the incoming gas, before returning to the low-pressure side of the compressor.

In the Heylandt liquefier the conditions are arranged so that the adiabatic expansion is not sufficient to cause liquefaction, but is sufficient to cool the gas below its inversion temperature. The final cooling is produced by a throttling process.

8.6 Helium Liquefiers

Helium liquefiers based on the Heylandt method have been designed by Kapitza and Collins. One of the biggest problems is the lubrication of the piston in the expansion cylinder because, at the temperatures involved, no liquid lubricant is available. This was solved in Kapitza's engine by employing no lubricant and leaving a very small gap between piston and cylinder. The escaping helium ensured that the piston did not come into contact with the sides of the cylinder. By employing a very rapid stroke the amount of helium escaping was kept low.

In the Collins liquefier the expansion takes place in two adiabatic stages followed by a Joule–Kelvin expansion.

A method due to Simon uses a single adiabatic expansion of compressed helium to produce liquefaction. Helium at a pressure of about 150 atmospheres is cooled in a bath of liquid hydrogen boiling under reduced pressure. The helium is then thermally isolated from its surroundings by a vacuum jacket and allowed to expand slowly to achieve reversible conditions. The resultant cooling produces liquefaction. More than 1 litre per expansion has been produced by this means.

PRODUCTION AND MEASUREMENT OF VERY LOW TEMPERATURES

8.7 Liquid Helium under Reduced Pressure

The normal boiling point of liquid helium is 4.2°K, but by reducing the pressure above the liquid lower temperatures may be obtained. By
LOW TEMPERATURES

mechanical pumping a temperature of about 1°K may be obtained. Improved methods of pumping have resulted in temperatures of about 0·8°K, but this seems to be the lowest which can be obtained by this means.

The common isotope of helium has mass number 4. There is also an isotope of mass 3 which has a lower normal boiling point (3·2°K). Also its vapour pressure does not fall off so rapidly with decrease in temperature, and consequently it has been possible to produce temperatures of 0·3°K using this isotope.

The manipulation of liquefied gases requires very careful techniques which will not be discussed here, but are described in some of the books mentioned in the bibliography. In the case of helium, the gas itself is so expensive that care must be taken to collect all the evaporated gas and return it to the liquefier. Also care must be taken to exclude all air from contact with the liquid helium, as the air would solidify and block the apparatus.

8.8 Adiabatic Demagnetization

When a paramagnetic* salt is magnetized under adiabatic conditions its temperature rises and, conversely, when it is demagnetized adiabatically its temperature falls. This effect may be used to produce temperatures below 1°K.

The paramagnetic salt is suspended by fine threads in an enclosure immersed in a liquid helium bath. Initially the enclosure contains gaseous helium at low pressure, so that thermal contact is established between the salt and the liquid helium. The salt is subjected to a strong magnetic field, the heat produced being conducted away into the helium bath. When equilibrium has been established the enclosure is evacuated to isolate the salt from its surroundings and the magnetic field then removed. This adiabatic demagnetization cools the salt below its surroundings. Temperatures of 10^{-3°K} have been reached in this way.

Salts which have been used in this process include: gadolinium sulphate, iron ammonium alum, chromium potassium alum. These latter two have sufficiently high thermal capacity to enable other substances to be cooled to very low temperatures by thermal contact with the salt after the demagnetization has taken place.

Temperatures of the order of 10^{-5} to 10^{-6} °K have been obtained by applying the method of adiabatic demagnetization to the nuclei of atoms. Nuclei possess magnetic moments about 10^{-3} times the magnitude of the magnetic moments of the orbital electrons. These nuclear magnetic moments are aligned by a very powerful magnet, cooled to 10^{-2°K} by contact with a paramagnetic salt, and then the magnetic field is switched off.

* See Chapter 13 for the meaning of this term.
8.9 Measurement of Low Temperatures

(a) Constant-volume gas thermometers

To establish a standard scale of low temperatures the constant-volume gas thermometer is used. A standard hydrogen thermometer may be used down to about 20°K and a standard helium thermometer down to about 4°K. For temperatures below this the helium gas must be used at pressures lower than those used in the standard thermometer; it is then possible to use the thermometer at 1°K. Corrections have to be applied to convert the readings to the Kelvin scale.

(b) Vapour-pressure thermometers

We have mentioned in the previous section that a hydrogen gas thermometer cannot be used below 20°K; this is because the hydrogen liquefies. However, the liquid hydrogen exerts a saturated vapour pressure, and measurement of this pressure enables the temperature to be determined, if the variation of saturated vapour pressure with temperature is known. Such a relation may be found by thermodynamic methods (e.g. from the Clausius–Clapeyron equation; see Section 7.18).

The hydrogen vapour-pressure thermometer may be used over the range 20 to 10 °K and the helium vapour-pressure thermometer between 4°K and slightly less than 1°K.

(c) Resistance thermometers

The resistance thermometer is a secondary thermometer, and therefore must be calibrated against some standard thermometer (e.g. the gas thermometer).

Platinum is suitable for moderately low temperatures, but below 50°K its temperature coefficient of resistance becomes small, and hence the thermometer insensitive. Gold or lead may be used for temperatures down to about 10°K.

Carbon has proved to be one of the most suitable resistors to use in the low-temperature region; its resistance increases with decreasing temperature. The early carbon resistance thermometers used graphite, but more recently it has been discovered that certain commercial radio resistors are more suitable, and these have been used at temperatures lower than 0·1°K.

Thermistors (see Section 13.9) are also suitable as low-temperature resistance thermometers.

(d) Thermocouples

Copper–constantan and gold–silver thermocouples may be used down to about 15°K. For lower temperatures special alloys are needed.
LOW TEMPERATURES

(e) Magnetic thermometers

The magnetic susceptibility of a paramagnetic salt changes appreciably in the region below 1°C, so that its value gives an indication of the temperature. The susceptibility may be measured by winding two coils round the specimen and determining the charge which circulates in the secondary when a known current is reversed in the primary. For this purpose the ends of the secondary may be connected to a ballistic galvanometer. As temperatures below 1°C are usually obtained by adiabatic demagnetization, the paramagnetic salt used may serve as its own thermometer.

SOME LOW-TEMPERATURE PHENOMENA

8.10 Liquid Helium

Liquid helium is unique: it has some most striking properties quite unlike anything found amongst other liquids.

The first attempt to solidify helium followed the usual pattern of reducing the pressure above the liquid so that it boils under reduced pressure and its temperature falls. With all other liquids this results in the formation of the solid when the triple point is reached, but with helium this does not occur. It was later confirmed that helium has no triple point at which solid, liquid and vapour are in equilibrium.

![Diagram](a) Density vs Temperature at 2.19°C

![Diagram](b) Specific heat vs Temperature at 2.19°C

Fig. 8.4

Measurements of the physical properties of liquid helium indicated that at a temperature of 2.19°C the liquid undergoes a sudden change. The main pieces of evidence are:

(i) Density. As the temperature is lowered the density increases to a maximum at 2.19°C and then begins to fall. The maximum is not a smooth one (as is the case for water at 4°C), but the slope of the curve changes discontinuously at this point (Fig. 8.4 (a)).

(ii) Specific Heat. The discontinuity at 2.19°C is even more striking in
this case because there is a sudden change in the actual value of the specific heat at this temperature (Fig. 8.4 (b)).

(iii) Viscosity. Experiments on the flow of liquid helium through capillary tubes have indicated a sudden decrease in the coefficient of viscosity at 2.19⁰K.

The temperature of 2.19⁰K is referred to as the $\lambda$-point of liquid helium and marks the transition from liquid helium I (above the $\lambda$-point) to liquid helium II. It is the latter form which possesses the unique properties referred to above. Some of these will now be described.

The fountain effect. This was discovered by Allen and Jones in 1938, and may most convincingly be demonstrated by the apparatus shown in Fig. 8.5.

The beam from a torch is shone on to the emery powder, and immediately a fountain of liquid helium II is formed. The energy is derived from the heating effect of the light beam. Fountains as high as 1 ft have been observed.

The viscosity of liquid helium II is very low. The viscous properties of the liquid are different from those of other liquids and different results for the coefficient of viscosity are obtained by different methods. It has been suggested that liquid helium II may be considered as a mixture of two liquids: a normal one, and a superfluid one with special properties, including zero viscosity. An explanation of the fountain effect has been developed along these lines.

The thermal conductivity also shows anomalous behaviour, and it is not possible to define a coefficient in the normal way. Experiments have shown apparent coefficients as high as 2,000 times that of copper under certain conditions. These abnormal values have also been explained in terms of superfluid helium.

A thin film of liquid helium II forms on any surface placed in the liquid. This film can 'creep' over the surface and results in the startling effects of a vessel of the liquid emptying or filling itself as depicted in Fig. 8.6.
8.11 Solid Helium

Helium has not been obtained as a solid at atmospheric pressure even very close to the absolute zero, and it seems likely that it can exist only under increased pressure.

The solid was first obtained by Keesom in 1926 under high pressure. He detected its presence by the blocking of a capillary tube. The presence of solid helium in the liquid cannot be seen because the density and refractive indices of the two states are almost identical. In later experiments the solid was detected by a magnetically controlled 'hammer' in the containing vessel.

Although the critical temperature of helium is 5.2°K the solid has been obtained at temperatures as high as 50°K. The pressure is then more than 7,000 atmospheres. The liquid cannot, of course, exist at any temperature above 5.2°K so that if the pressure is reduced the solid helium sublimes.

Another interesting feature is that the latent heat of solid helium below 1°K is almost zero.

8.12 Superconductivity

As the temperature of a metal is lowered its resistance steadily decreases. Some metals behave in this way until a certain temperature is reached, and then, suddenly, their resistance decreases to zero. The temperature at which this occurs is called the transition temperature, and below it the metal is called a superconductor. A large number of metals show this effect, as do some alloys and compounds. For metals the transition occurs over a very small temperature range (less than 10⁻² deg K), but for compounds the resistance disappears more gradually and the transition may occupy an interval of 1 deg K.

A few examples of superconducting metals (with transition temperatures in brackets) are: lead (7.2°K), mercury (4.2°K), tin (3.7°K), aluminium (1.2°K), zinc (0.91°K). Copper, silver and gold do not become superconducting.

If an electric current is set up in a ring of a superconductor it continues to circulate long after the source of e.m.f. has been removed. Experiments by Onnes using currents of hundreds of amps in a ring of superconducting lead showed that there was no observable heating effect or diminution in the size of the current after several hours. It has been shown that any resistance present in superconducting lead is less than 10⁻¹⁵ times the resistance at 0°C!

The absence of any heating effect in a superconductor carrying a current opens up several exciting possibilities. For example, the problem which has always been associated with the production of very strong magnetic fields has been how to dissipate the heat produced by the very large currents flowing in the coils of the electromagnets. This has set a limit to the size of a continuous magnetic field obtainable by these means. But, if the electromagnet is wound with, say, lead wire and
cooled below the transition temperature, should it not be possible to use enormous currents without producing any heat at all?

Unfortunately, a further property of superconductors prevents the idea working; superconductivity is destroyed by strong magnetic fields. The field needed depends on the metal and the temperature, but for most pure metals is less than 1,000 oersted – not a very strong magnetic field. Recently (1963) reports have been made that alloys have been developed which remain superconducting at sufficiently high magnetic field strengths to make possible the production of very strong fields with a superconducting electromagnet.
FURTHER TOPICS IN ELECTRICITY

ELECTROSTATICS

9.1 Gauss’ Theorem

The lines of electric force due to a point charge $+Q$ e.s.u. in vacuo are directed radially outwards from it, and if a spherical surface of radius $r$ cm is drawn with $Q$ as centre, then all the lines cross this surface normally (Fig. 9.1). The intensity, $E$ e.s.u., of the electric field at the surface is given by $E = Q/r^2$. The number of lines of force crossing the surface of the sphere is $ES$, where $S$ is the surface area of the sphere, so that:

$$\text{number of lines of force} = \frac{Q}{r^2} \times 4\pi r^2 = 4\pi Q.$$

![Fig. 9.1](image1.png)
![Fig. 9.2](image2.png)

This result may be extended to a surface of any shape as follows. Fig. 9.2 shows a point charge $Q$ in vacuo with an imaginary surface enclosing it.

Consider a small element of the surface of area $\delta S$ distance $r$ from $Q$, $\theta$ is the angle between the direction of $r$ and the normal to $\delta S$. The effective area perpendicular to the field at $\delta S$ is $\delta S \cos \theta$. If the intensity of the field is denoted by $E$, then the number of lines of force passing through $\delta S$ is $E\delta S \cos \theta = (Q/r^2) \delta S \cos \theta$. But $\frac{\delta S \cos \theta}{r^2} = \delta \omega$, the solid angle subtended at $\theta$ by $\delta S$,

$$\therefore E \delta S \cos \theta = Q \delta \omega.$$

Integrating this over the whole surface gives

$$\int E \cos \theta \, dS = Q \int d\omega = 4\pi Q.$$
The left-hand side is the total number of lines of force through the surface, which is seen to be equal to $4\pi Q$, in agreement with the result obtained for the particular case of a spherical surface. The position of $Q$ inside $S$ does not affect the final result so that the above equation may be extended to include the case where $S$ encloses a number of charges, and it then becomes

$$\int E \cos \theta \, dS = 4\pi \Sigma Q \quad \quad \quad \quad \quad (9.1)$$

where the integral is evaluated over the whole of $S$, and the summation is for all charges enclosed by $S$. This result is Gauss' Theorem.

The left-hand side of equation (9.1) may be rewritten in terms of the component, $E_n$, of $E$ along the normal to $dS$, for $E_n = E \cos \theta$, and hence

$$\int E_n \, dS = 4\pi \Sigma Q.$$

If $Q$ is surrounded by dielectric the displacement vector, $D$, must replace $E$ in Gauss' theorem. $D$ and $E$ are related by $D = kE$, where $k$ is the dielectric constant of the medium.

### 9.2 Applications of Gauss' Theorem

Gauss' theorem may often be used to determine the intensity of the electric field in the neighbourhood of a certain distribution of charges. There follows one worked example; other examples are contained in the questions at the end of the chapter.

**Worked Example.** Find the intensity of the electric field just above the surface of: (i) a layer of charge of density $\sigma$ e.s.u. cm$^{-2}$; (ii) a flat charged conductor with surface density of charge $\sigma$ e.s.u. cm$^{-2}$.

(i) If the layer of charge is large ('infinite'), then, by symmetry, the field is normal to the surface and has the same intensity, $E$, on both sides. Fig. 9.3 shows a cylinder of area of cross-section 1 cm$^2$ drawn with its axis perpendicular to the surface. The number of lines of force crossing the ends of the cylinder is $E \times 2$ (total area = 2 cm$^2$), and there are none crossing the curved surface. The total charge enclosed by the cylinder is that on 1 cm$^2$ of the layer of charge, i.e. $\sigma$.
Hence, by Gauss' theorem,
\[ 2E = 4\pi \sigma, \]
\[ \therefore E = 2\pi \sigma. \]

(ii) In this case (Fig. 9.4) there is no field in the conductor, so that the total number of lines of force is \( E \times 1 \) and, by Gauss' theorem,
\[ E = 4\pi \sigma. \]

### 9.3 Capacitance of a Coaxial Cable

A coaxial cable consists of a wire of radius \( a \) cm surrounded by a hollow metal cylinder of radius \( b \) cm, the space between them being filled with an insulator of dielectric constant \( k \) (Fig. 9.5).

Consider length \( l \) cm of cable and suppose that the outer cylinder is earthed. Suppose that the inner one is given a charge \( q \) c.s.u. which raises its potential to \( V_1 \) c.s.u. Imagine a cylinder of radius \( r \) \((a < r < b)\) concentric with the wire, and denote the potential at its surface by \( V \). The surface area of this cylinder is \( 2\pi rl \), and the charge enclosed by it is \( Q \). By symmetry, the electric field at the surface of this cylinder is normal to the surface, ignoring end effects. If the intensity of this field is denoted by \( E \), then, by Gauss' theorem,
\[ kE \cdot 2\pi rl = 4\pi Q, \]
\[ \therefore E = \frac{2Q}{kl}. \]

The intensity of the field and the potential are related by
\[ E = \frac{dV}{dr}, \]
\[ \therefore \frac{dV}{dr} = \frac{2Q}{kl}. \]
\[ \therefore \int_{0}^{V_1} dV = -\frac{2Q}{kl} \int_{a}^{b} \frac{dr}{r}, \]
\[ 
\therefore V_1 = \frac{2Q}{kl} \ln \left(\frac{b}{a}\right). \]

Capacitance,
\[ C = \frac{Q}{V_1} = \frac{kl}{2 \ln \left(\frac{b}{a}\right)} \text{ c.s.u.} \]

### 9.4 Force on the Surface of a Charged Conductor

To obtain an expression for this force we must consider in more detail the way in which the field near the surface of a charged conductor
arises. We showed in Section 9.2 that a layer of charge of density \( \sigma \) e.s.u. cm\(^{-2} \) gives rise to a field of intensity \( 2\pi \sigma \) on either side of it (see Fig. 9.3). If this layer of charge is situated on a conductor we must assume that it still produces the same effect, so that to account for the actual zero field inside the conductor there must be another electric field present. This latter field arises from other charges on or near the conductor; these charges must be so distributed as to give a field of intensity \( 2\pi \sigma \), both inside and outside the conductor, in the direction of the outward normal (Fig. 9.6). Inside the conductor the two fields cancel to give zero field, and outside they add to give a field of intensity \( 4\pi \sigma \).

The force on the charge in the pill box in Fig. 9.6 is due to the field of intensity \( 2\pi \sigma \) arising from the other charges; the field due to the charge itself must not be taken into account in this case, because a charge cannot exert a force on itself. It follows that the force, \( F \), on unit area of charge is given by

\[
F = \sigma \cdot 2\pi \sigma = 2\pi \sigma^2.
\]

This force is communicated to the surface of the conductor, and so there is an outward force on each cm\(^2\) of the conductor, of magnitude \( 2\pi \sigma^2 \) dynes.

### 9.5 The Attracted Disc Electrometer

The formula obtained in the previous section is the basis of the attracted disc electrometer, an instrument for measuring electrostatic potential differences. Fig. 9.7 shows the principle of the construction of

![Fig. 9.7](image)

the instrument, which is essentially a circular parallel-plate capacitor with one plate (P) movable. The other plate (Q) is fixed and of larger area than P. G is a guard ring which is electrically connected to P and ensures that the lines of force from P to Q are parallel. P is supported
by the spring S, and weights may be added to or removed from P so that it is in the same plane as G.

When a potential difference, \( V \) e.s.u., is applied between P and Q, the force on P causes it to move downwards and weights are taken off it until it returns to its original position.

Let the area of P be \( A \) cm\(^2\), the distance between P and Q be \( d \) cm, and the mass which has to be removed from P when a potential difference \( V \) e.s.u. is applied be \( m \) g.

The intensity of the field, E, between the plates is \( V/d \) e.s.u. and, by Gauss' theorem, this is equal to \( 4\pi\sigma \), where \( \sigma \) e.s.u. cm\(^{-2}\) is the charge density on P.

\[
\therefore \sigma = \frac{V}{4\pi d}
\]

and the force on

\[
P = 2\pi\sigma^2 A
\]

\[
= 2\pi \left( \frac{V}{4\pi d} \right)^2 A
\]

\[
= \frac{V^2 A}{8\pi d^2} \text{ dynes.}
\]

But a weight of \( mg \) dynes has been removed from P to counterbalance this force, and hence

\[
mg = \frac{V^2 A}{8\pi d^2}
\]

\[
\therefore V = d \left( \frac{8\pi mg}{A} \right)^{1/2} \text{ e.s.u.}
\]

This instrument is an absolute instrument, because it does not need to be calibrated against any electrical standards: all the measurements necessary to calculate \( V \) are mechanical.

The forces involved are small and the instrument is suitable only for measuring potential differences of the order of kilovolts.

**9.6 The Quadrant Electrode**

This instrument is much more sensitive than the attracted disc electrometer and may be used to measure potential differences of \( \frac{1}{1000} \) volt. It consists of a light conducting vane, N, suspended by a torsion wire between the faces of hollow conducting quadrants, Q (Fig. 9.8). A small mirror, M, is used in conjunction with a lamp and scale to measure the angular deflection of N. Opposite pairs of quadrants (Q\(_1\), Q\(_3\); Q\(_2\), Q\(_4\)) are connected together electrically.

N is maintained at a potential \( V_1 \), Q\(_1\), Q\(_3\) at potential \( V_1 \) and Q\(_2\), Q\(_4\) at potential \( V_2 \). The needle and the quadrants form two parallel-plate capacitors (i.e. N with Q\(_1\), Q\(_3\) and N with Q\(_2\), Q\(_4\)) which have potential differences \( V - V_1 \) and \( V - V_2 \) respectively, between the plates. The electrometer thus stores electrostatic energy, the amount depending on the potential differences, and on the position of N. The instrument is
zeroed such that when \( V_1 = V_2 \), \( N \) is as shown in Fig. 9.8 and there is no torsion in the suspension. If \( V_1 \neq V_2 \), \( N \) rotates into a position of lower electrostatic energy, the angle \( \theta \) through which it rotates being such that the loss in electrostatic energy is equal to the gain in mechanical energy due to the twisting of the suspension fibre.

![Diagram](image)

*Fig. 9.8*

There are two ways of using the electrometer:

(i) **Heterostatic method.** \( V \) is kept constant at a high value which is much greater than \( V_1 \) or \( V_2 \). The potential difference to be measured is \( V_1 - V_2 \). \( \theta \) is approximately proportional to \( V_1 - V_2 \).

(ii) **Idiostatic method.** One pair of quadrants is connected to \( N \) so that \( V = V_2 \) and the potential difference to be measured is \( V_1 - V_2 \). \( \theta \) is approximately proportional to \( (V_1 - V_2)^2 \), and is thus independent of the sign of \( V_1 - V_2 \); alternating potential differences can be measured in this way.

### 9.7 Use of the Quadrant Electrometer to Measure a Very Small Current

In experiments on the ionization of gases it is necessary to measure currents of the order of \( 10^{-13} \) amp. This may be accomplished by connecting a quadrant electrometer in series with the ionization chamber and the source of e.m.f. (usually several kilovolts). The current, \( I \), flowing in the circuit charges the electrometer, so that

\[
I = \frac{dQ}{dt} = C \frac{dV}{dt},
\]

where \( Q \) is the charge on the electrometer when the potential difference between the quadrants is \( V \), and \( C \) is its capacitance. If the electrometer is connected heterostatically, then \( \theta = kV \), where \( k \) is a constant and hence \( I = kC \frac{d\theta}{dt} \). The current is determined, then, by observing the rate of change of the deflection of the needle with time. The capacitance
of the electrometer is very small (of the order of a few picofarads) so that a measurable \( \frac{d\theta}{dt} \) is obtained even with the very small current it is desired to measure.

If the experiment involves finding ratios of currents, then neither \( k \) nor \( C \) need be known. If actual values of current are required the instrument must be calibrated. \( k \) may be found by calibration against known potential differences. \( C \) is determined by using a capacitor of known capacitance, \( C' \) (which must be of the same order of magnitude as \( C \)). The electrometer is charged to a potential difference \( V \) and isolated; the capacitor is then connected across it. If the potential difference falls to \( V' \), then

\[
CV = (C + C')V',
\]

i.e.

\[
C\theta = (C + C')\theta',
\]

where \( \theta \) and \( \theta' \) are respectively the deflections before and after \( C' \) is connected. \( C \) may be determined from the above relation.

**ELECTROMAGNETISM**

**9.8 Magnetic Field on the Axis of a Circular Coil**

Consider a circular coil of \( N \) turns, of radius \( a \) cm carrying an electric current \( I \) c.m.u. We wish to find the strength of the magnetic field on the axis of the coil, distance \( x \) cm from the centre (Fig. 9.9).

![Fig. 9.9](image)

Use is made of the Biot–Savart formula for the field \( \delta H \) oersted at a point distance \( r \) cm from a current element of length \( \delta l \) cm, carrying a current \( I \) c.m.u.: 

\[
\delta H = \frac{I\delta l \sin \theta}{r^2},
\]

where \( \theta \) is the angle between the current element and the line joining it to the point.

In Fig. 9.9, \( \theta = 90^\circ \), so that \( \delta H = I\delta l/r^2 \) in the direction shown. The component of \( \delta H \) perpendicular to OP cancels out with the corresponding component from a similar current element at the opposite end of the diameter.
The component of $\delta H$ along $OP = \delta H \cos \phi$

$$= \frac{I \delta l}{r^2} \cos \phi.$$  

The total field at $P$ is thus along $OP$ and of magnitude

$$H = \sum \frac{I \delta l}{r^2} \cos \phi.$$  

But $\cos \phi = \frac{a}{r}$ and $r = (a^2 + x^2)^{\frac{1}{2}},$

$$\therefore \quad H = \frac{I}{(a^2 + x^2)} \cdot \frac{a}{(a^2 + x^2)^{\frac{1}{2}}} \sum \delta l$$  

$$= \frac{Ia}{(a^2 + x^2)^{\frac{1}{2}}} \cdot N \cdot 2\pi a$$  

$$\therefore \quad H = \frac{2\pi N a^2 I}{(a^2 + x^2)^{\frac{1}{2}}} \text{ oersted.}$$

If $P$ is at the centre of the coil this reduces to $H = \frac{2\pi NI}{a}$.

**9.9 Magnetic Field on the Axis of a Solenoid**

Suppose that the solenoid is of radius $a$ cm and has $n$ turns per cm. Consider a point, $P$, on the axis of the solenoid (Fig. 9.10). A small portion of the solenoid, of length $\delta x$, at distance $x$ cm from $P$ has $n \delta x$ turns and produces a field

$$\delta H = \frac{2\pi (n \delta x) a^2 I}{(a^2 + x^2)^{\frac{1}{2}}}$$

along the axis at $P$.

$$\therefore \text{total field at } P \text{ is } H = \int_{x_1}^{x_2} \frac{2\pi n a^2 I}{(a^2 + x^2)^{\frac{1}{2}}} \, dx.$$  

Now

$$x = a \cot \theta,$$

$$\therefore \quad dx = -a \csc^2 \theta \, d\theta,$$

and

$$x^2 + a^2 = r^2 = a^2 \csc^2 \theta.$$  

$$\therefore \quad H = \int_{\theta_1}^{\theta_2} \frac{2\pi n a^2 I(-a \csc^2 \theta)}{a^2 \csc^2 \theta} \, d\theta$$  

$$= -\int_{\theta_1}^{\theta_2} 2\pi n I \sin \theta \, d\theta,$$

$$\therefore \quad H = 2\pi n I (\cos \theta_2 - \cos \theta_1) \text{ oersted.}$$

Fig. 9.10
FURTHER TOPICS IN ELECTRICITY

If \( P \) is in the middle of a long solenoid \( \theta_2 = 0 \) and \( \theta_1 = \pi \), so that \( H = 2\pi n l (1 - (-1)) = 4\pi n l \). If \( P \) is at one end of a long solenoid \( \theta_2 = 0, \theta_1 = \pi/2 \), so that \( H = 2\pi n l (1 - 0) = 2\pi n l \).

9.10 Self-induction

Any circuit carrying a current produces magnetic flux, and, if the current changes, the flux linking the circuit also changes, so that an e.m.f. is induced. This e.m.f. is in a direction such that it opposes the change in current which produces it. This effect is known as self-induction.

If the induced e.m.f. is 1 volt when the current changes at a rate of 1 amp sec\(^{-1}\) the circuit is said to have a self-inductance, \( L \), of one henry. Thus

\[
E = -L \frac{dI}{dt},
\]

the minus sign indicating that the e.m.f. opposes the change in current.

Calculation of the self-inductance from the dimensions of a circuit is in general difficult. We consider only one idealized case: a long air-cored solenoid.

Suppose the solenoid has \( n \) turns per cm and is of length \( l \) cm and area of cross-section \( A \) cm\(^2\). If the current flowing is \( I \) e.m.u. the intensity of the field inside the solenoid is \( 4\pi n l \) oersted and the flux, \( \Phi \), threading the solenoid is \( 4\pi n l A \) maxwell. The flux linkage,

\[
(N\Phi) = nl \cdot 4\pi n l A.
\]

If the current changes, the induced e.m.f., \( E \) e.m.u., is given by

\[
E = -\frac{d(N\Phi)}{dt} = -4\pi n l A \frac{dI}{dt}.
\]

If the e.m.f. is expressed as \( V \) volts and the current as \( I' \) amps, then

\[
V = E \times 10^{-8} \text{ and } I' = I \times 10,
\]

\[\therefore V = -4\pi n l A \frac{dI'}{dt} \times 10^{-9}\]

\[\therefore L = 4\pi n l A \times 10^{-9} \text{ henry.}\]

9.11 Mutual Induction

When two circuits are close together the flux produced by one links the other, so that a change of current in the first circuit induces an e.m.f. in the second. This is mutual induction.

If a current changing at 1 amp sec\(^{-1}\) in one circuit induces an e.m.f. of 1 volt in the other the two circuits are said to have a mutual inductance of 1 henry. Thus

\[
E = -M \frac{dI}{dt}.
\]
We calculate a value for $M$ in one simple case: a short coil of $N$ turns wound on top of the solenoid discussed in the previous section.

As before, $\Phi = 4\pi nIA,$ so that the flux linkage with the short coil is $4\pi nNA$. Proceeding as before we obtain $M = 4\pi nNA \cdot 10^{-9}$ henry.

**9.12 Theory of the Ballistic Galvanometer**

In this section no units are mentioned, the formulae apply if the quantities are expressed either all in c.g.s. or all in m.k.s. units.

The galvanometer consists of a coil of area $A$, consisting of $N$ turns of wire, suspended in a radial magnetic field of flux density $B$ by a wire which provides a restoring torque, $\epsilon$, when twisted through 1 radian. The coil is wound on a non-conducting former so that the motion of the coil is not damped by eddy currents.

If a steady current, $i$, flows through the coil it experiences a torque $NBAi$ and comes to rest when this is balanced by the restoring torque in the suspension. The coil turns through an angle $\theta$ given by $NBAi = \epsilon \theta$. The current sensitivity, $k$, is defined as $\theta/i$, and thus

$$k = \frac{\theta}{i} = \frac{NBA}{\epsilon} \quad \ldots \quad (9.2)$$

If, however, the current lasts only for a short time, $\delta t$, then the coil receives an angular impulse of magnitude (couple $\times$ time), i.e. $NBAi \delta t$. But $i \delta t$ is the charge, $\delta Q$, which has passed through the coil in time $\delta t$. Hence, if a charge $Q$ is passed through the coil the angular impulse $= NBAQ$.

Suppose that the charge passes through the coil in a time which is short compared with the period of oscillation of the coil, so that the impulse is complete before the coil has moved appreciably. The effect is that the coil is given an initial angular velocity, $\omega_0$, and then moves only under the restoring torque of the suspension wire (assuming that damping may be neglected). The coil comes instantaneously to rest at angle $\theta$, such that the wire has potential energy equal to the original kinetic energy of the coil. The energy stored in the wire is equal to $\frac{1}{2} (\text{final torque}) \times (\text{angle turned through})$, so that $\theta$ is given by

$$\frac{1}{2} \epsilon \theta^2 = \frac{1}{2} I \omega_0^2,$$

where $I$ is the moment of inertia of the rotating parts.

$\omega_0$ may be calculated from the equation: angular impulse = increase in angular momentum,

i.e. $NBAQ = I \omega_0$,

$$\therefore \frac{1}{2} \epsilon \theta^2 = \frac{1}{2} I \left(\frac{NBAQ}{I}\right)^2,$$

$$\frac{NBAQ}{(I\epsilon)^{\frac{1}{2}}}.$$
This shows that, under these conditions, $\theta$ is proportional to $Q$. The quantity sensitivity, $\gamma$, is defined as $\theta/Q$ and is given by

$$\gamma = \frac{\theta}{Q} = \frac{NBA}{(Ic)^{\frac{1}{2}}}.$$

The relation between $k$ and $\gamma$ may be established as follows:

$$\frac{k}{\gamma} = \frac{NBA}{c} \sqrt{\frac{NBA}{(Ic)^{\frac{1}{2}}}} = \left(\frac{I}{c}\right)^{\frac{1}{2}} = \frac{T}{2\pi},$$

where $T$ is the period of free oscillations of the coil.

**Effect of damping**

The above treatment has ignored the presence of damping, which, in practice, arises from two causes: (i) air resistance; (ii) induced currents in the coil (*electromagnetic damping*).

The air damping is constant and usually small, but the electromagnetic damping depends on the resistance in the circuit and may be large.

When the coil rotates it cuts the lines of force of the magnetic field and the resultant induced e.m.f. causes a current to flow in the circuit in such a direction as to oppose the motion of the coil; this is the electromagnetic damping.

Suppose that the coil is rectangular, of height $h$ and width $b$. If it rotates through an angle $\delta\theta$ in time $\delta t$ the flux cut by one wire on one vertical side of the coil in this time is $B \cdot \frac{b}{2} \delta\theta \cdot l$ (Fig. 9.11), so that the total flux cut by the wires on both vertical sides of the coil is

$$NBAh \delta\theta = NBA \delta\theta.$$

![Fig. 9.11](image)

The induced e.m.f., $E$, is equal to the rate of cutting of the lines of flux,

$$\therefore E = -NBA \frac{\delta\theta}{\delta t}.$$

If $\delta t \rightarrow 0$, then $E = -NBA \frac{d\theta}{dt}$.
The induced current, $i$, is given by
\[ i = \frac{E}{R} = -\frac{NBA}{R} \frac{d\theta}{dt}, \]
where $R$ is the total resistance of the circuit. This induced current produces a torque,
\[ NBAi = -\frac{(NBA)^2}{R} \frac{d\theta}{dt}, \]
opposing the motion of the coil.

There is also a torque, $-c\theta$, produced by the suspension wire, and so the equation of motion of the coil is
\[ I \frac{d^2\theta}{dt^2} + \frac{(NBA)^2}{R} \frac{d\theta}{dt} + c\theta = 0 \]
where air damping has been neglected.

This equation is of the form of equation (3.3) with \( \frac{(NBA)^2}{IR} = k \) and \( c/I = \omega^2 \). It was shown in Section 3.2 that the solution is oscillatory only if \( k^2 < 4\omega^2 \),

i.e. if
\[ \left[ \frac{(NBA)^2}{IR} \right]^2 < \frac{4c}{I}. \]

i.e. if
\[ R > \frac{(NBA)^2}{2(4c)^2}. \]

In practice, so that the damping shall be small, it is necessary to have
\[ R \gg \frac{(NBA)^2}{2(4c)^2}. \]

Reference to Fig. 3.2 shows that the observed throw, $\theta_1$ (which corresponds to $a_1$ in this figure) is less than the true throw, $\theta$, in the absence of damping ($a$ in Fig. 3.2). To obtain an accurate correction for damping it is necessary to observe $\theta_1$, $\theta_3$, etc., from which $\theta$ may then be calculated; however, provided the damping is small, it is sufficient to add on to the observed first throw a correction equal to one quarter the difference between the first and second throws on the same side, i.e.
\[ \theta = \theta_1 + \frac{\theta_3 - \theta_1}{4}. \]

This allows for the damping during the $\frac{1}{4}$ oscillation which the coil has made before $\theta_1$ is observed.

ELECTRICAL CIRCUITS

9.13 Kirchhoff’s Laws

These laws apply to networks of conductors. The first expresses the fact that there cannot be an accumulation of charge at a junction, and the second is an extension of Ohm’s law.
**FURTHER TOPICS IN ELECTRICITY**

*First law.* At any junction of conductors the algebraic sum of the currents flowing into the junction is zero.

*Second law.* In a complete circuit the algebraic sum of the (current × resistance) products is equal to the algebraic sum of the sources of e.m.f.

In applying these laws it is usually best to mark the currents on the circuit diagram, making use of the first law to reduce the number of unknowns to a minimum. The second law is then applied to various branches of the network to give the number of equations needed to determine the currents.

**Worked Example.** A metre bridge is used to determine the value of a resistance. When a known resistance of 5 ohms is used the slider is at 50.0 cm when the bridge is balanced. What current will flow through the galvanometer when the slider is moved 1 cm from balance, if the meter has a resistance of 50 ohms and the slide wire a resistance of 10 ohms? The driver cell has an e.m.f. of 2 volts.

The 'unknown' resistance must be 5 ohms because the bridge balances in the middle. Fig. 9.12 shows the network and the currents flowing. Use has been made of Kirchhoff’s first law.

![Circuit Diagram](image)

**Fig. 9.12**

Using Kirchhoff’s second law:

in circuit BCEB,

\[ 5i_2 + 50i_3 - 4.9(i_1 - i_2) = 0, \]

\[ \therefore -4.9i_1 + 9.9i_2 + 50i_3 = 0 \quad \ldots \quad (i) \]

in circuit DCED,

\[ -5(i_2 - i_3) + 50i_3 + 5.1(i_1 - i_2 + i_3) = 0, \]

\[ \therefore 5.1i_1 - 10.1i_2 + 60.1i_3 = 0 \quad \ldots \quad (ii) \]

in circuit ABCDA,

\[ 5i_2 + 5(i_3 - i_3) = 2, \]

\[ \therefore 10i_2 - 5i_3 = 2 \quad \ldots \quad (iii) \]

Substituting for \(i_3\) from (iii) in (i) and (ii) gives two equations containing \(i_1\) and \(i_2\).
Eliminating \( i_1 \) from these equations gives

\[ i_3 = -3.6 \times 10^{-4} \text{ amp.} \]

Thus the current through the galvanometer is in the opposite direction to that indicated in Fig. 9.12 and is of magnitude 0.36 mA.

### 9.14 D.C. Circuit containing Capacitance and Resistance

(a) Capacitor charging

![Fig. 9.13](image)

Suppose that at time \( t \) the charge on \( C \) is \( Q \) and the current is \( I \). The potential difference, \( V \), across \( C \) is then \( Q/C \). Applying Ohm’s law:

\[ E = IR + V \quad \cdot \quad (9.3) \]

\[ \therefore E = IR + \frac{Q}{C} \]

Differentiating with respect to \( t \) gives

\[ 0 = R \frac{dI}{dt} + \frac{1}{C} \frac{dQ}{dt}. \]

But

\[ \frac{dQ}{dt} = I, \quad \therefore CR \frac{dI}{dt} = -I. \]

Separating the variables and integrating gives

\[ CR \ln I = -t + c, \]

where \( c \) is a constant of integration. If at \( t = 0 \) the capacitor is uncharged, then \( I \) (at \( t = 0 \)) = \( E/R \),

\[ \therefore CR \ln \frac{E}{R} = c. \]

Hence

\[ CR \ln \left( \frac{I}{E/R} \right) = -t, \]

\[ \therefore I = \frac{E}{R} e^{-\frac{t}{CR}}. \]

Substituting in equation (9.3) gives

\[ V = E(1 - e^{-\frac{t}{CR}}), \]

and hence

\[ Q = CE(1 - e^{-\frac{t}{CR}}). \]

The quantity \( CR \) is called the time constant of the circuit. When \( t = CR, \) \( e^{-\frac{t}{CR}} = e^{-1} = 0.37 \). Hence the time constant is the time taken for \( I \) to fall to 0.37 of its maximum value, and for \( V \) or \( Q \) to rise to \( 1 - 0.37 = 0.63 \) of their final value. Fig. 9.14 shows graphs of \( I, V, Q \) against \( t \).
(b) Capacitor discharging

If $C$ is initially charged to a potential difference $V_0$ and then allowed to discharge through $R$ (there being no e.m.f. in the circuit), equation (9.3) becomes

$$O = IR + V$$

which may be solved as before to give

$$I = -\frac{V_0}{R} e^{-\frac{t}{CR}}.$$ 

Substituting in equation (9.4), gives

$$V = V_0 e^{-\frac{t}{CR}},$$

and hence

$$Q = CV_0 e^{-\frac{t}{CR}} = Q_0 e^{-\frac{t}{CR}}.$$

This last equation is the basis of a method for determining a very high resistance. A capacitor of known capacitance, $C$, is charged to a potential difference $V_0$ and then discharged through a ballistic galvanometer to determine $Q_0$. $C$ is then recharged to the same potential difference, allowed to discharge through the resistance $R$ for $t$ seconds, and then discharged through the ballistic galvanometer to determine $Q$. All the quantities in the equation are now known except $R$, which may therefore be calculated. The values of $C$ and $R$ must be such that the time constant $CR$ is several seconds,

\[ R = 10 \text{ megohm}, \ C = 5 \mu F, \]

whence the time constant is 50 seconds. Allowance must be made for the leakage resistance of the capacitor; this may be determined by repeating the experiment without the resistance $R$. 
9.15 D.C. Circuit containing Resistance and Inductance

An inductance \( L \) and a resistance \( R \) are connected in series with a cell of e.m.f. \( E \). The e.m.f. induced across the inductance when the current is \( I \) is \( -L \frac{dI}{dt} \), so that Ohm's law applied to the circuit gives

\[
E - L \frac{dI}{dt} = IR.
\]

(9.5)

\[
\therefore \int \frac{dI}{E/R - I} = \frac{R}{L} \int dt,
\]

\[
\therefore -\ln \left( \frac{E}{R} - I \right) = \frac{R}{L} t + c.
\]

When \( t = 0, I = 0, \)

\[
\therefore c = -\ln \frac{E}{R}
\]

\[
\therefore \ln \left( \frac{E/R - I}{E/R} \right) = -\frac{Rt}{L}
\]

\[
\therefore I = \frac{E}{R} \left( 1 - e^{-\frac{Rt}{L}} \right).
\]

The time constant of this circuit is \( L/R \) sec.

If, when the current has reached the steady value \( E/R, E \) is shorted out, then equation (9.5) becomes

\[
-L \frac{dI}{dt} = IR
\]

which, when solved, gives

\[
I = \frac{E}{R} e^{-\frac{Rt}{L}}.
\]

Fig. 9.16 illustrates the two cases.

\[
I = \frac{E}{R}
\]

\[
I
\]

\[
t
\]

Fig. 9.16
FURTHER TOPICS IN ELECTRICITY

When a current is flowing in an inductance the magnetic field stores energy. The magnitude of this energy may be deduced from equation (9.5), i.e.

\[ E = L \frac{dI}{dt} + IR. \]

The energy supplied by the source of e.m.f. in time \( \delta t \) is \( EI \delta t \), and

\[ EI \delta t = L \frac{dI}{dt} \cdot I \delta t + I^2 R \delta t. \]

The second term on the right-hand side of this equation is the energy dissipated as heat in the resistance; we must interpret the first term on that side as the energy stored in the inductance. If the current increases from zero to \( I \), the energy stored in the inductance is thus

\[ \int_0^I LI \, dI = \frac{1}{2} LI^2. \]

If \( L \) is in henries and \( I \) in amps, \( \frac{1}{2} LI^2 \) is the stored energy in joules.

9.16 Discharge of a Capacitor through an Inductance

Assume first that there is no resistance in the circuit (Fig. 9.17), and that \( I = 0 \) at \( t = 0 \).

As there is no source of e.m.f. in the circuit

\[ L \frac{dI}{dt} + \frac{Q}{C} = 0, \]

\[ \therefore \frac{d^2I}{dt^2} + \frac{1}{LC} \cdot I = 0. \]

This is an equation of the type discussed in Section 3.1 (cf. equation (3.1) with \( \omega^2 = \frac{1}{LC} \)), and hence the solution is \( I = I_0 \sin \left( \left( \frac{1}{LC} \right)^\frac{1}{2} t \right) \).

The current is oscillatory, the period of the oscillation being \( 2\pi(\frac{LC})^{\frac{1}{2}} \), i.e. a frequency of \( \frac{1}{2\pi(\frac{LC})^{\frac{1}{2}}} \).

If the circuit contains resistance \( R \), then

\[ L \frac{dI}{dt} + IR + \frac{Q}{C} = 0, \]

\[ \therefore \frac{d^2I}{dt^2} + R \frac{dI}{dt} + \frac{I}{LC} = 0. \]

This is like equation (3.3) with \( k = \frac{R}{L} \) and \( \omega^2 = \frac{1}{LC} \). As shown in Section 3.2, the discharge is oscillatory only if

\[ k^2 < 4\omega^2, \]
\[ \frac{R^2}{L^2} < \frac{4}{LC} \]

i.e. if

\[ R^2 < \frac{4L}{C} \]

The solution is given in equation (3.4) and a plot of this in Fig. 3.2, the current \( I \) replacing the displacement \( x \). The solutions for

\[ R^2 \geq \frac{4L}{C} \quad (k^2 \geq 4\omega^2) \]

are given in equations (3.5) and (3.6).

9.17 A.C. Circuit containing Resistance only

If an alternating e.m.f., \( E = E_0 \sin \omega t \), is applied to a resistance \( R \), then the current \( I \) at any instant is

\[ I = \frac{E}{R} = \frac{E_0}{R} \sin \omega t \]

\[ = I_0 \sin \omega t. \]

The current and voltage are in phase and the peak value of the current, \( I_0 = \frac{E_0}{R} \).

9.18 A.C. Circuit containing Capacitance only

If an alternating e.m.f., \( E = E_0 \sin \omega t \), is applied to a capacitance \( C \), then the charge, \( Q \), on \( C \) at any instant is given by

\[ Q = CE = CE_0 \sin \omega t. \]

The current,

\[ I = \frac{dQ}{dt} \]

\[ = CE_0 \frac{d}{dt} (\sin \omega t) \]

\[ = CE_0 \omega \cos \omega t. \]

\[ \therefore I = I_0 \sin \left( \omega t + \frac{\pi}{2} \right). \]

The phase of the current leads that of the voltage by \( \pi/2 \). The peak value of the current, \( I_0 = \frac{E_0}{1/\omega C} \). By comparison with Ohm’s law, \( 1/\omega C \) is seen to be analogous to resistance, and is called the reactance of \( C \), denoted by \( X_c \).

Hence

\[ X_c = \frac{1}{\omega C} = \frac{1}{2\pi f C} \]

where \( f \) is the frequency of the e.m.f. in cycles sec\(^{-1}\). If \( C \) is in farads, \( X_c \) is in ohms. It should be noted that the reactance decreases as \( f \) increases.
9.19 A.C. Circuit containing Inductance only

If an alternating current, \( I = I_0 \sin \omega t \), is flowing through an inductance \( L \), then the e.m.f. induced across it is \(-L \frac{dI}{dt}\). This must be balanced by the applied e.m.f. \( E \),

\[
E = L \frac{dI}{dt} = 0
\]

\[
\therefore E = L I_0 \omega \cos \omega t
\]

\[
= E_0 \sin \left( \omega t + \frac{\pi}{2} \right).
\]

The phase of the voltage leads that of the current by \( \pi/2 \). The peak value of the e.m.f., \( E_0 = \omega L I_0 \). The reactance, \( X_L \), is then

\[
X_L = \frac{E_0}{I_0} = \omega L = \frac{2L}{\pi} fL.
\]

If \( L \) is in henries, \( X_L \) is in ohms.

The reactance increases as \( f \) increases.

9.20 A.C. Circuits containing either \( L \) or \( C \) with \( R \)

The problems are solved most simply by vector methods.

(a) Series circuit

The same current flows through both components, so a vector (usually drawn in the \( x \)-direction) is used to represent current. Vectors

\[
Z = \left[ (\omega L)^2 + R^2 \right]^{1/2}
\]

is the impedance of the circuit,

\[
\phi = \tan^{-1} \left( \frac{\omega L}{R} \right)
\]

is the phase difference between \( V \) and \( I \) (\( V \) leads).

Fig. 9.18

\[
Z = \left[ \frac{1}{(\omega C)^2} + R^2 \right]^{1/2}
\]

is the impedance of the circuit,

\[
\phi = \tan^{-1} \left( \frac{1}{\omega CR} \right)
\]

is the phase difference between \( I \) and \( V \) (\( I \) leads).

Fig. 9.19
representing the potential differences are then drawn, due account being taken of phase differences. The total potential difference across the circuit is found by vector addition (Figs. 9.18 and 9.19).

(b) Parallel circuit

There is the same potential difference across each component. The vector diagrams are shown in Figs. 9.20 and 9.21.

\[ I = \frac{V}{\omega L} \]
\[ I_R = \frac{V}{R} \]
\[ I_L = \sqrt{V^2 - \left(\frac{V}{\omega L}\right)^2} \]
\[ I_C = \frac{V}{\sqrt{\omega C}} \]
\[ 1 = \left(\frac{1}{R^2} + \left(\frac{1}{\omega L}\right)^2\right)^{\frac{1}{2}} \]
\[ \phi = \tan^{-1} \frac{R}{\omega L} \quad (V \text{ leads}) \]

Fig. 9.20

\[ 1 = \left(\frac{1}{R^2} + (\omega C)^2\right)^{\frac{1}{2}} \]
\[ \phi = \tan^{-1} \omega CR \quad (I \text{ leads}) \]

Fig. 9.21

9.21 A.C. Circuit containing L, C and R in Series

The vector diagram is shown in Fig. 9.22.
FURTHER TOPICS IN ELECTRICITY

\[ Z = \left( R^2 + \left( \omega L - \frac{1}{\omega C} \right) \right)^{\frac{1}{2}}; \quad \phi = \tan^{-1} \left( \frac{\omega L - \frac{1}{\omega C}}{R} \right). \]

A special case occurs when \( \omega L = 1/\omega C \). The current and voltage are then in phase and the impedance has the value \( R \), which is the least value it can have. The current, for a given \( V \), is thus a maximum. This is an example of resonance. The resonance occurs when \( \omega L = 1/\omega C \), i.e.

\[ \omega = \frac{1}{(LC)^{\frac{1}{2}}}; \quad f = \frac{1}{2\pi(LC)^{\frac{1}{2}}}. \]

The circuit is sometimes called an acceptor circuit because if a mixture of frequencies is applied to the circuit it accepts (has the minimum impedance for) the frequency \( f = \frac{1}{2\pi(LC)^{\frac{1}{2}}} \).

9.22 Algebraic Treatment of the \( L, C, R \) Series Circuit

If the applied e.m.f. is \( E = E_0 \sin \omega t \), then the current is given by:

\[
L \frac{dI}{dt} + IR + \frac{Q}{C} = E_0 \sin \omega t.
\]

The solution of this equation has two parts: a complementary function and a particular integral. The solutions for the complementary function are given in Section 3.2, and it will be noticed that they all contain an exponential term, which causes the amplitude of the solution to decrease with time. After a certain time has elapsed these transients have become negligible. The steady condition is obtained by determining the particular integral as follows.

The equation has a solution of the form

\[ I = A \cos \omega t + B \sin \omega t. \]  \hspace{1cm} (9.7)

Differentiating twice and substituting for \( dI/dt \) and \( d^2I/dt^2 \) in equation (9.6) gives

\[
L(-A\omega^2 \cos \omega t - B\omega^2 \sin \omega t) + R(-A\omega \sin \omega t + B\omega \cos \omega t) + \frac{1}{C} (A \cos \omega t + B \sin \omega t) = E_0 \omega \cos \omega t.
\]

Equating coefficients of:

\[
\begin{align*}
\cos \omega t: & \quad -LA\omega^2 + RB\omega + \frac{A}{C} = E_0 \omega; \\
\sin \omega t: & \quad -LB\omega^2 - RA\omega + \frac{R}{C} = 0.
\end{align*}
\]
Solving for \( A \) and \( B \) gives

\[
A = \frac{E_0 \left( \frac{1}{\omega C} - \omega L \right)}{R^2 + \left( \frac{1}{\omega C} - \omega L \right)^2},
\]

\[
B = \frac{E_0 R}{R^2 + \left( \frac{1}{\omega C} - \omega L \right)^2}.
\]

Equation (9.7) may also be written as

\[
I = \frac{E_0}{Z} \sin(\omega t - \phi),
\]

where \( Z \) is the impedance and \( \phi \) is the phase angle by which the voltage leads the current. Expanding this equation gives

\[
I = \frac{E_0}{Z} \cos \phi \sin \omega t - \frac{E_0}{Z} \sin \phi \cos \omega t,
\]

and equating coefficients with those of equation (9.7)

\[
A = -\frac{E_0}{Z} \sin \phi,
\]

\[
B = \frac{E_0}{Z} \cos \phi.
\]

Squaring and adding these two equations gives

\[
A^2 + B^2 = \frac{E_0^2}{Z^2},
\]

i.e.

\[
Z = \frac{E_0}{(A^2 + B^2)^{1/4}},
\]

and dividing,

\[
\tan \phi = \frac{A}{B}.
\]

The reader should verify, by substituting the values of \( A \) and \( B \), that these are the same values of \( Z \) and \( \phi \) that were obtained by the vector method.

9.23 Use of Complex Numbers in A.C. Circuit Problems

It will be noticed that the vector diagrams (Figs. 9.19 to 9.22) are similar to the Argand diagrams used in the treatment of complex numbers. This suggests that A.C. circuit problems may be treated by the use of complex numbers if we represent the reactance of an inductance as \( j \omega L \), and that of a capacitance by \(- \frac{j}{\omega C}\) which is \( \frac{1}{j \omega C} \), where \( j = \sqrt{-1})\).
The total impedance, \( Z \), of a circuit is then represented by a complex number \( (Z = X + jY) \), which contains information about the real impedance \( |Z| = (X^2 + Y^2)^{\frac{1}{2}} \) and the phase angle \( \phi \) (\( \tan \phi = Y/X \)) by which the voltage leads the current (Fig. 9.23).

![Fig. 9.23](image1)

![Fig. 9.24](image2)

An example will make this clear. Fig. 9.24 shows a circuit which may be met with in practice, the resistance \( R \) being that which is introduced by the inductance \( L \).

To find the impedance of the circuit we first combine the complex impedances of \( L \) and \( R \) in series to give \( R + j\omega L \) and then add this to the complex impedance of \( C \) in parallel, the total complex impedance, \( Z \), being given by

\[
\frac{1}{Z} = \frac{1}{R + j\omega L} = \frac{j\omega C(R + j\omega L) + 1}{R + j\omega L} = \frac{1 - \omega^2 LC + j\omega CR}{1 - \omega^2 LC + j\omega CR} \quad \therefore Z = \frac{[R + j\omega L][1 - \omega^2 LC - j\omega CR]}{[1 - \omega^2 LC + j\omega CR][1 - \omega^2 LC - j\omega CR]} = \frac{R + j(\omega CR^2 + \omega L - \omega^3 L^2 C)}{(1 - \omega^2 LC)^2 + \omega^2 C^2 R^2}.
\]

The real impedance is thus

\[
|Z| = \frac{[R^2 + (\omega CR^2 + \omega L - \omega^3 L^2 C)^2]^\frac{1}{2}}{(1 - \omega^2 LC)^2 + \omega^2 C^2 R^2}
\]

and

\[
\phi = \tan^{-1}\left[\frac{-\omega CR^2 + \omega L - \omega^3 L^2 C}{R}\right].
\]

**9.24 Power in A.C. Circuits**

We have seen that, in general, there is a phase difference between the voltage and the current in an A.C. circuit. We may represent the voltage and current by

\[
E = E_0 \sin \omega t,
\]

and

\[
I = I_0 \sin (\omega t - \phi),
\]
The instantaneous value of the power is \( EI \), which is found by multiplying the two equations together:

\[
EI = E_0I_0 \sin \omega t \sin (\omega t - \phi) \\
= \frac{1}{2}E_0I_0[\cos \phi - \cos (2\omega t - \phi)],
\]

using the relation, \( \sin A \sin B = \frac{1}{2}[\cos (A - B) - \cos (A + B)] \).

The average value of the power is found by averaging \( EI \) over a complete cycle:

\[
\bar{EI} = \frac{1}{2}E_0I_0[\cos \phi - \cos (2\omega t - \phi)]
\]

The second term in the brackets averages to zero in a cycle, and hence

\[
\bar{EI} = \frac{1}{2}E_0I_0 \cos \phi.
\]

But \( E_0 = \sqrt{2} \cdot E_{R.M.S.} \) and \( I_0 = \sqrt{2} \cdot I_{R.M.S.} \).

\[
\therefore \bar{EI} = E_{R.M.S.} \cdot I_{R.M.S.} \cos \phi.
\]

\( \cos \phi \) is the power factor of the circuit and is the ratio between the true power dissipated (\( \bar{EI} \)) and the apparent power. The apparent power may be determined by measuring \( E_{R.M.S.} \) and \( I_{R.M.S.} \) separately by meters, and finding their product.

### 9.25 The Transformer

We consider the case of an ideal transformer, i.e. one in which the following conditions are fulfilled:

(i) the primary has zero resistance;

(ii) all the flux generated by the primary passes through the secondary;

(iii) the secondary is on open circuit.

Suppose that the primary has a total of \( N_p \) turns and the secondary a total of \( N_s \) turns. Denote the inductances of the primary and secondary by \( L_p \) and \( L_s \) respectively and their mutual inductance by \( M \).

If the flux generated by the primary changes at a rate \( \frac{d\Phi}{dt} \), the e.m.f., \( E_s \), generated in the secondary is \( -N_s \frac{d\Phi}{dt} \) and the back e.m.f. in the primary is \( -N_p \frac{d\Phi}{dt} \). As the resistance of the primary is negligible, the e.m.f. applied, \( E_p \), must be equal and opposite to the back e.m.f. in the primary, i.e.

\[
E_p = N_p \frac{d\Phi}{dt}
\]

\[
\therefore \frac{E_p}{E_s} = \frac{N_p}{N_s}.
\]

The negative sign indicates that there is a phase difference of \( \pi \) between the two e.m.f.s.
FURTHER TOPICS IN ELECTRICITY

If the secondary is not on open circuit the induced current flowing in it produces a flux which opposes the change in flux produced by the changing current in the primary. This would result in a decrease in the back e.m.f. in the primary, but this is not possible, as this e.m.f. must remain equal and opposite to the applied e.m.f.; the current in the primary must therefore increase to offset the reduction in flux.

QUESTIONS

1. A charge is distributed uniformly through the volume of a sphere. Use Gauss' theorem to find expressions for the intensity of the field, both inside and outside the sphere.

2. Charge is distributed uniformly along a long line such that there is Q e.s.u. per cm. Use Gauss' theorem to show that the intensity of the field at distance r cm from the line is Q/2r e.s.u.

3. A soap bubble of radius r cm carries a charge Q e.s.u. Show that the excess pressure inside it is \( \frac{4\gamma}{r} - \frac{Q^2}{8\pi r^4} \), where \( \gamma \) is the surface tension of the soap solution.

4. Show that for a ballistic galvanometer \( \ln(\theta_1/\theta_2) = \ln(\theta_2/\theta_3) = \text{constant} \) (see Chapter 3, Question 4). This constant is the logarithmic decrement (\( \delta \)). Show that the corrected throw \( \theta \) may be found from \( \theta = \theta_1 e^{\delta/4} \), and hence describe how you would make an accurate correction for damping when using the ballistic galvanometer.

5. The secondary of a mutual inductance is connected to a ballistic galvanometer through a high resistance, and a current I amps is reversed in the primary. Show that the charge through the ballistic galvanometer is \( 2MI/R \) coulombs, where \( M \) is the mutual inductance in henries and \( R \) is the total resistance of the secondary circuit, in ohms. Describe a way of calibrating a ballistic galvanometer, based on the above.

6. Use the \( j \) method to find the impedance of the following circuits:
   (i) \( C \) and \( R \) in parallel connected in series with \( L \);
   (ii) \( L, C, R \) in parallel.

7. Show that for an ideal transformer the self-inductance of the primary, \( L_p \), the self-inductance of the secondary, \( L_s \), and the mutual inductance, \( M \), are related by \( M^2 = L_p L_s \).
ELECTRICITY IN RATIONALIZED M.K.S. UNITS

10.1 Introduction

There are in use in physics two main systems of units: the centimetre-gram-second (c.g.s.), and the metre-kilogram-second (m.k.s.) systems. In the former the three primary quantities in mechanics, length, mass and time, are measured respectively in centimetres, grams and seconds; in the latter in metres, kilograms and seconds. In electricity it is convenient to fix the units of a fourth primary quantity, and the way in which this is done leads to a variety of systems of electrical units. There are the c.g.s. electrostatic and electromagnetic systems, and the rationalized and unrationlized m.k.s. systems of electrical units. It is assumed that the reader is already familiar with the first two and has a grasp of the fundamental concepts involved, so that any material which is common to both the c.g.s. and m.k.s. systems will not be repeated in this chapter. The m.k.s. system of units is usually associated with an approach to the subject which is different from that adopted in either the electromagnetic or electrostatic systems; such an approach is followed in this chapter.

10.2 Mechanical Units

The units of the derived, or secondary, quantities in mechanics

<table>
<thead>
<tr>
<th></th>
<th>M.K.S.</th>
<th>C.G.S.</th>
<th>Relation between two sets of units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Force</td>
<td>1 newton produces an acceleration of 1 m sec(^{-2}) when it acts on a mass of 1 kg</td>
<td>1 dyne produces an acceleration of 1 cm sec(^{-2}) when it acts on a mass of 1 g</td>
<td>1 newton = (10^5) dynes</td>
</tr>
<tr>
<td>Work or Energy</td>
<td>1 joule of work is done when a force of 1 newton moves its point of application 1 m</td>
<td>1 erg of work is done when a force of 1 dyne moves its point of application 1 cm</td>
<td>1 joule = (10^7) ergs</td>
</tr>
<tr>
<td>Power</td>
<td>1 watt of power is delivered when work is done at the rate of 1 joule per sec</td>
<td>1 erg sec(^{-1}) of power is delivered when work is done at the rate of 1 erg per sec</td>
<td>1 watt = (10^7) erg sec(^{-1})</td>
</tr>
</tbody>
</table>

Table 10.1 Definitions of the derived units in mechanics
10.3 Rationalized M.K.S. in Electricity

Many of the formulae encountered in the c.g.s. development of the subject contain a factor of \( \pi \), and these formulae happen to be the ones most used, particularly by electrical engineers. The rationalized m.k.s. development arranges that \( \pi \) is missing from these formulae, but this results in its reappearing in formulae where it is not to be found in c.g.s. The student will notice that most formulae are different by a factor of \( 1/4\pi \) from those to which he is accustomed.

The electrostatic system of units (e.s.u.) has as its basis the repulsion between like electric charges. The constant in the inverse square law of force is made equal to unity, thus defining the unit of charge, and from this follow the definitions of field strength, potential, capacitance, etc. The electromagnetic system of units (e.m.u.) starts with the inverse square law of repulsion between like magnetic poles, and again the constant is made equal to unity. This defines the unit magnetic pole, and from this follow the definitions of the units of magnetic field, current, etc. The starting point in the rationalized m.k.s. system is the force which acts between conductors carrying electric currents. The experimental work on which this approach is based was first carried out by Ampère in 1820–5.
statement of his conclusions, and no attempt is made to show how these are related to the experimental results.

Ampère's approach was to consider each circuit to consist of a very large number of small lengths of wire, each carrying the same current as the circuit of which it forms part; these are referred to as current elements. Each current element in one circuit exerts a force on every current element in the other circuit; the total force experienced by the latter circuit is the sum of the forces on each of its current elements. What is required, then, is a formula for the force exerted by one current element on another, and it is this formula which may be deduced from the results of Ampère's experiments. We start with a simplified version of the full formula, confining our attention to the special case of two current elements in the same plane, placed so that the line joining their centres is perpendicular to both elements (Fig. 10.1). The two current elements are of length \( \delta l_1 \) and \( \delta l_2 \) and carry currents \( I_1 \) and \( I_2 \) respectively; the distance between their centres is \( r \). The force, \( \delta F \), on \( \delta l_2 \) is given by

\[
\delta F = C \frac{(I_1 \delta l_1)(I_2 \delta l_2)}{r^2}
\]

where \( C \) is a constant which depends on the system of units used. The direction of the force is indicated in Fig. 10.1; if the direction of one current is reversed the direction of the force is reversed. (In this particular case there is, of course, an equal and opposite force on \( \delta l_1 \), but we shall confine our discussion to the force on \( \delta l_2 \).)

The value of the constant in equation (10.1) depends on the system of units used. In the m.k.s. system we have not yet defined a unit of current, so that we may assign to \( C \) any value we choose.

There is an obvious similarity between the fixing of \( C \) in the above law of force between two current elements and the corresponding procedure in electrostatics and magnetism in the c.g.s. system, but there are also two important differences. First, a current element is a vector, whereas electric charge and magnetic pole are scalars, so that the above equation is a special case rather than a general form; this does not make the definition of the units any less general. Secondly, in this case \( C \) is not put equal to unity. The reason for this is purely historical. If \( C \) were put equal to unity this would fix a unit of current which would be different from both the electromagnetic unit of current and the ampere, and this would create an unnecessary multiplicity of units. Accordingly, \( C \) is chosen so that the unit of current on the m.k.s. system is the ampere (the system is often referred to as the m.k.s.A. system). On the rationalized system, \( C \) is written as \( \mu_0 / 4\pi \) in a vacuum, so that the \( 4\pi \) is included in the formula. \( \mu_0 \) is called the permeability of free space and is given the value \( 4\pi \times 10^{-7} \).
ELECTRICITY IN M.K.S. UNITS

Equation (10.1) then becomes

\[ \delta F = \frac{\mu_0 (I_1 \delta l_1)(I_2 \delta l_2)}{4\pi r^2} \]  \hspace{1cm} (10.2)

and applies only in a vacuum.

10.5 Flux Density (Magnetic Induction)

It is convenient to picture the force on \( \delta l_2 \) arising from a two-stage process. The current element \( I_1 \delta l_1 \) is said to produce a flux density or magnetic induction (denoted by \( B \)) in the region round it, and in particular at the position of \( \delta l_2 \). The force on \( \delta l_2 \) is then said to arise from the interaction of \( B \) and the current element. If we denote the flux density at \( \delta l_2 \) by \( \delta B \) we may write the force on \( \delta l_2 \) as

\[ \delta F = \delta B(i_2 \delta l_2) \]  \hspace{1cm} (10.3)

\( B \) is a vector quantity, and in the particular case we are considering it is taken to be at right angles both to \( I_2 \) and \( \delta F \).

Experiments to determine how the force, \( F \), on a wire of length \( \delta l \), carrying a current \( I \), varies as the angle between \( B \) and \( \delta l \) is changed show that the force is always perpendicular both to \( B \) and \( I \), and that its magnitude is

\[ F = BI \delta l \sin \phi \]  \hspace{1cm} (10.4)

where \( \phi \) is the angle between \( B \) and \( \delta l \) (Fig. 10.2). The maximum force is exerted when \( B \) and \( \delta l \) are perpendicular and is \( F = BI \delta l \).

We may thus phrase the definition of \( B \) as follows:

The magnitude of the flux density at a point is the maximum force which it can exert per metre length on a current element carrying a current of 1 amp, placed at that point.

When the current element is orientated so that it is experiencing its maximum force the direction of \( B \) is perpendicular both to the force and the current, and its positive direction is given by Fleming's left-hand rule. Note that we are here using the rule to define the direction of \( B \) in terms of the directions of \( F \) and \( I \) (Fig. 10.3).

From equation (10.4) it is seen that \( B \) could be expressed in newton amp\(^{-1}\) m\(^{-1} \), but a more usual unit is weber m\(^{-2} \).

By comparing equations (10.2) and (10.3) we obtain an expression for the flux density, \( \delta B \), produced by \( I_1 \delta l_1 \), at \( \delta l_2 \),

i.e. \[ \delta B = \frac{\mu_0 I_1 \delta l_1}{4\pi r^2} \].

![Fig. 10.2](image-url)

![Fig. 10.3](image-url)
This expression applies to any point which is distance $r$ metres from
the current element and lies on a line drawn perpendicular to it.
For a point, $P$, which lies on a line making an angle $\theta$ with $\delta l_1$ (Fig.
10.4) the expression becomes

$$\delta B = \frac{\mu_0 I_1 \delta l_1 \sin \theta}{4\pi r^2}$$

(10.5)

$\delta B$ is in a direction perpendicular to the plane containing $\delta l_1$ and $P$, its
positive direction is indicated by the direction of rotation of a right-
hand screw advancing in the direction of $I$.

![Diagram](image)

$B$ at $P$ is perpendicular to the plane
of the paper and directed into it

Fig. 10.4

For a complete circuit in a vacuum (dropping the suffixes)

$$B = \mu_0 \sum I \delta l \sin \theta \frac{1}{4\pi r^2}$$

(10.6)

By combining equations (10.4) and (10.5) the equation for the force
exerted by $I_1 \delta l_1$ on $I_2 \delta l_2$ in a vacuum when there is no restriction on
their directions, may be written

$$\delta F = \frac{\mu_0 (I_1 \delta l_1 \sin \theta)(I_2 \delta l_2 \sin \phi)}{4\pi r^2}$$

(10.7)

This formula cannot be tested experimentally, but the theoretical
results for complete circuits to which it leads have been well verified in
the experiments of Ampère and later workers. This verification may
also be taken as verification of equation (10.5). It will be seen from
equation (10.7) that $\mu_0$ has dimensions and that it could be expressed as
newton amp$^{-2}$; this may be shown to be equivalent to henry m$^{-1}$ and this
is the unit in which $\mu_0$ is usually expressed.

10.6 Definition of the Ampere

Now that $\mu_0$ has been given the value $4\pi \cdot 10^{-7}$ henry m$^{-1}$, equation
(10.2) could be used to define the ampere in terms of the force between
two current elements.

It is usual, however, to phrase the definition of the ampere in terms
of the force between two long parallel wires carrying currents in the
same direction. This force may be calculated in terms of the currents $I_1$, $I_2$ and the distance apart, $a$, of the wires, in the following way.

Flux density $B$ at distance $a$ from a long straight wire carrying a current $I_1$ is given by (Fig. 10.5):

$$B = \frac{\mu_0 I_1 \delta l_1 \sin \theta}{4\pi a^2}$$

$$= \frac{\mu_0 I_1}{2\pi a} \text{ (see Table 10.3).}$$

Substituting this value for $B$ in equation (10.4) and putting $\phi = 90^\circ$, and $\delta l = 1$ metre, gives for the force per metre on the second wire

$$F = \frac{\mu_0 I_1 I_2}{2\pi a} \text{ newtons per metre length of wire.}$$

Putting $I_1 = I_2 = 1$ amp, $a = 1$ metre, and $\mu_0 = 4\pi \times 10^{-7}$ gives

$$F = \frac{4\pi \cdot 10^{-7} \cdot 1 \cdot 1}{2 \cdot \pi \cdot 1} = 2 \times 10^{-7} \text{ newtons.}$$

This result is used to define the ampere as follows:

*The ampere is that current which, flowing in two long thin parallel conductors 1 metre apart in a vacuum, gives a force of $2 \times 10^{-7}$ newtons per metre length of each conductor.*

It is possible to regard the above definition of the ampere as the starting point of the m.k.s.A. system of units in electricity, and to use the above argument to show that this leads to the value $\mu_0 = 4\pi \times 10^{-7}$ henry m$^{-1}$. It is obvious that either approach will lead to the same set of units.

To show that the ampere as defined above is the same as the ampere previously defined in the c.g.s. system as $10^{-8}$ e.m.u. of current, the student should calculate the force per metre on each of two long conductors 1 metre apart and carrying currents of 1 amp, using the c.g.s. formulae, and check that the result obtained is $2 \times 10^{-2}$ dynes (i.e. $2 \times 10^{-7}$ newtons).

**10.7 Definition of the Coulomb**

In the m.k.s. system the unit of charge is the coulomb, and this is defined in terms of the ampere as follows:

*If a current of 1 amp flows through a conductor for 1 second then a charge of 1 coulomb passes through any cross-section of the conductor.*

**10.8 Magnetic Flux**

Lines of flux may be used to represent a magnetic field. This is done by drawing lines in the direction of $B$, and making the number of lines
per unit area normal to $B$ numerically equal to $B$. An important property of lines of flux is that they are continuous closed curves, whether in a vacuum or a magnetic material, or both.

A number of lines of flux are said to represent a certain quantity of magnetic flux. Quantitatively, magnetic flux is defined in terms of the flux density $B$. Suppose that in a particular magnetic field the flux density is constant over a region (uniform magnetic field). In this case the lines of flux are parallel straight lines. Imagine now an area, $A$, drawn perpendicular to the direction of $B$, so that the lines of flux cross this area at right angles to it. The quantity of magnetic flux, $\Phi$, crossing this area is defined as the product of the flux density and the area, i.e.

$$\Phi = BA \text{ webers.}$$

If $B$ is not constant over the area in question, then the flux must be found by integration, i.e.

$$\Phi = \int B \, dA \text{ webers},$$

where the elementary areas $dA$ are always drawn perpendicular to $B$.

10.9 Electromagnetic Induction

Faraday's laws of electromagnetic induction in their general form (i.e. not mentioning any units) must, of course, still hold good in the m.k.s. system. Thus, the induced e.m.f. is proportional to the rate of change of flux linkage with a circuit, i.e.

$$E \propto -\frac{d(N\Phi)}{dt},$$

where $E$ is measured in volts (defined in the usual way, see Section 10.15), and $N$ is the number of conductors linked by the flux, $\Phi$ webers.

![Fig. 10.6](image)

All the units involved are therefore defined, and it now remains to determine the constant of proportionality in the expression. The following argument shows that this constant is unity.

The wire XY (Fig. 10.6), of length $l$ metres is free to slide on horizontal rails with which it makes electrical contact. A current is maintained in the circuit by a source of e.m.f., and there is a magnetic field
of flux density $B$ weber m$^{-2}$ directed perpendicularly into the paper. As a result XY is acted on by a force $F$ newtons in the direction shown, and hence moves.

Suppose that at some instant the wire has velocity $v$ m sec$^{-1}$, and that the current in the circuit is then $I$ amp. XY is cutting across lines of flux, so that an induced e.m.f., $E$, is developed across its ends, given by

$$ E = -K \frac{d(N\Phi)}{dt} \text{ volts}, $$

where $K$ is a constant to be determined.

The area swept out by XY per second is $lv$ m$^2$ sec$^{-1}$, and hence the rate of change of flux linkage is $Blv$ weber sec$^{-1}$,

$$ \therefore E = -KBlv \text{ volts}. $$

To maintain the current at the same value, $I$, which it had when the wire was at rest the source of e.m.f. must increase by $E$ volts and hence deliver extra energy at the rate of $EI$ watts

$$ = KBlvI \text{ watts}. $$

Now the force, $F$, exerted on XY, which is given by

$$ F = BlI \text{ newtons}, $$

does work at the rate of $FV$ watts,

i.e. $Blv$ watts.

But, according to the law of conservation of energy, the rate at which energy is expended to overcome the back e.m.f. must be the same as the rate at which energy is acquired by the wire, which means that

$$ KBlvI = Blv $$

$$ \therefore K = 1. $$

The mathematical statement of Faraday's law then becomes

$$ E = -\frac{d(N\Phi)}{dt} \text{ volts} \quad \quad \quad \ldots (10.8) $$

where $\Phi$ is the flux measured in webers.

From equation (10.8) it can be seen that flux might also be expressed in the units volt sec, i.e. weber = volt sec, a relation which is often useful in checking the dimensions of equations.

10.10 Magnetizing Force

Magnetizing force, $H$, is defined such that in a vacuum, $B = \mu_0 H$. Thus $H$ is a vector quantity, and in a vacuum is proportional to $B$ and in the same direction.

From equation (10.6)

$$ H = \sum \frac{I \delta l \sin \theta}{4\pi r^2} \quad \quad \quad \ldots (10.9) $$
From this it is seen that the dimensions of $H$ are amp m$^{-1}$, and this is the unit in which it is expressed.

We have defined $H$ in a vacuum, but the way in which we have done this has led to an equation in which $H$ is seen to depend only on a current and on certain factors relating to the geometry of the circuit in which that current is flowing. This enables us to say that, even if the circuit is not in a vacuum, we will still calculate $H$ from equation (10.9), and hence obtain the same result as before. What we are doing is to regard $H$ purely as a property of an electric current in a circuit; this is emphasized by the choice of the unit amp m$^{-1}$ for the measurement of $H$. Some authors, in fact, use equation (10.9) to define $H$.

### 10.11 Magnetic Materials

If a magnetic material is placed in a magnetic field it becomes magnetized by induction and alters the flux density of the field in which it is situated. The flux density produced in the magnetic material by induced magnetism is called its intensity of magnetization, denoted by $J$. It is measured in the same units as flux density, i.e. weber m$^{-2}$.

If the magnetizing force is $H$, then the total flux density, $B$, in the material is given by

$$B = \mu_0 H + J.$$ 

This may also be written

$$B = \mu_0 \mu_r H,$$

and $\mu_r$ is called the relative permeability of the medium.* It has no units, and its value is the same as the value of the c.g.s. permeability.

### 10.12 Magnetic Circuits

Consider a solenoid of $N$ turns wound on an iron toroid of circumference $l$ metres. The magnetizing force inside the solenoid is given by

$$H = \frac{NI}{l} \text{ (see Table 10.3)},$$

as $N/l$ is the number of turns per metre.

Rearranging this equation gives

$$Hl = NI \text{ ampere-turns} \quad . \quad (10.10)$$

The quantity $Hl$ on the left of this equation is called the magnetomotive force (m.m.f.), and this is regarded as the cause of the magnetic flux in the toroid. There is an analogy here with the concept in electrical circuitry, in which an e.m.f. causes an electric current in the circuit. We

* Some authors use $\mu$ to denote relative permeability, whilst others use it to denote absolute permeability ($= \mu_0 \mu_r$); we use the suffix throughout to avoid possibility of confusion.
now proceed to find the relation between the m.m.f. and the flux it produces.

Suppose that the area of cross-section of the toroid is $A$, that the flux density in it is $B$, and that the relative permeability of the iron is $\mu_r$.

Then the flux, $\Phi$, in the toroid is given by

$$
\Phi = BA \\
= \mu_0 \mu_r H \\
= \mu_0 \mu_r \frac{NI}{l} \\
= \frac{NI}{l/\mu_0 \mu_r A}.
$$

The numerator in this expression is the magnetomotive force (see equation (10.10)), and the denominator, $l/\mu_0 \mu_r A$, is called the reluctance of the magnetic circuit. This is analogous to resistance in the electrical case, which is $\rho l/A$, where $\rho$ is the resistivity, or $l/\sigma A$, where $\sigma$ is the conductivity.

In general terms we may state a law for the magnetic circuit which is analogous to Ohm's law in electricity, viz.:

$$
\text{flux} = \frac{\text{m.m.f.}}{\text{reluctance}}. \tag{10.11}
$$

This analogy must not be taken too far, one important difference between the two cases being that flux (despite its name) does not flow whereas electric current does.

If the magnetic circuit is such that $H$ has several different values, then the m.m.f. is the sum of the $HI$ products taken round the circuit. If $H$ changes continuously integration must replace summation, i.e.

$$
\text{m.m.f.} = \int H \, dl. \tag{10.12}
$$

where the distances $dl$ are measured in the direction of $H$. If the path of integration is not in the direction of $H$, then one takes the component of $H$ in the direction of $dl$.

In the special case of the toroid it was shown that the m.m.f. is equal to $NI$; this result may be shown to be a general one (if one excludes permanent magnets) so that, using equation (10.12),

$$
\int H \, dl = NI. \tag{10.13}
$$

$NI$ represents the number of ampere-turns enclosed by the path round which the integration is performed. Equation (10.13) enables the magnetizing force of certain simple electrical circuits to be calculated very easily.
200 ADVANCED THEORETICAL PHYSICS

Worked Example; Magnetic Circuit. It is desired to make an electromagnet which shall have a flux density of 0·1 weber m⁻² in the air gap. The core of the magnet is made of 50 cm of material of relative permeability 2,000 and is of constant cross-section. The air gap is to be 1 cm. If the windings can carry a current of 5 amp, how many turns will be needed to produce the required flux density?

It will be assumed that there is no leakage of flux and that the lines of flux across the air gap are straight so that the area of cross-section of the air gap is the same as that of the core.

Let the area of cross-section of the core be \( A \). The reluctance of the core is

\[
\frac{L_{\text{core}}}{\mu_0 \mu_r A} = \frac{0.5}{4 \cdot \pi \cdot 10^{-7} \cdot 2,000 A}
\]

\( (N.B. \ l \text{ is expressed in metres.}) \)

Reluctance of air gap is

\[
\frac{L_{\text{air}}}{\mu_0 A} = \frac{0.01}{4 \cdot \pi \cdot 10^{-7} A}
\]

In the magnetic circuit the core and the air gap are in series and the total reluctance is the sum of their reluctances (cf. addition of resistances in series).

\[
\therefore \text{ total reluctance} = \frac{0.5 + 0.01 \times 2,000}{4 \cdot \pi \cdot 10^{-7} \cdot 2,000 A} = \frac{20.5}{8 \pi A} \times 10^4
\]

The m.m.f. in the circuit = \( NI = 5N \), where \( N \) is the number of turns.
The flux in the circuit = \( BA = 0.01A \).

Substituting these in equation (10.11) gives

\[
0.1A = \frac{5N}{20.5 \times 10^{-4}}
\]

\[
\therefore N = \frac{0.1A \cdot 20.5}{8 \pi A \cdot 5} \times 10^4 = 163 \text{ turns.}
\]

In the above example we assumed that there was no leakage of flux from the core and no fringing of the flux in the air gap; this means that the flux density, \( B \), is the same in the core and in the air gap. The magnetizing force, \( H \), is then discontinuous at the air gap. In the magnetic material it has the value \( B/\mu_0 \mu_r \), and in the air gap the larger value \( B/\mu_0 \).

10.13 Magnetic Moment

No mention has yet been made of permanent magnetism, the starting-point in the electromagnetic system of units. In the approach indicated in this chapter no reference to magnetic pole strength is made, this being an unnecessary concept. In practice, one can never deal with a magnetic pole, but only with a pair of equal and opposite poles, known as a magnetic dipole. The strength of a magnetic dipole is indicated by its magnetic dipole moment.

The concept of magnetic moment is most easily introduced by an electromagnetic example.
Fig. 10.7 shows an end-on view of a rectangular coil of dimensions \( l \) metres by \( b \) metres wound with \( N \) turns of wire carrying a current \( I \) amp. A uniform field of flux density \( B \) weber \( \text{m}^{-2} \) makes an angle of \( \theta \) with the plane of the coil.

There are forces as shown on the sides of the coil perpendicular to the paper, given by

\[
F = NBli \text{ newtons.}
\]

These produce a torque \( T \) newton metres tending to turn the coil, where

\[
T = nBlib \cos \theta = NBAI \cos \theta,
\]

where \( A \) metre\(^2\) is the area of the plane of the coil.

This torque has a maximum value, \( T_{\text{max.}} \), when \( \theta = 0 \).

\[
\therefore T_{\text{max.}} = NBAI.
\]

The magnetic moment, \( M \), of the coil is defined as this maximum torque divided by the flux density* of the field in which it is situated, i.e.

\[
M = \frac{T_{\text{max.}}}{B}. \quad \ldots \ldots \quad (10.14)
\]

In words the magnetic moment may be defined thus:

The magnetic moment of a coil carrying an electric current is the torque required to hold it with its axis at right angles to a field of unit flux density.

In terms of the magnetic moment, \( M \), the torque on the coil may be written

\[
T = MB \cos \theta.
\]

Suppose now that the circuit takes the form of a long solenoid. This may be considered as a large number of coils laid end to end, and as the above argument applies to each of these coils separately, it also applies to the solenoid as a whole. We now show how the magnetic moment of this solenoid is related to the flux density it produces, and its length.

If the solenoid has area of cross-section \( A \), length \( l \) and a total of \( N \) turns carrying a current \( I \), its magnetic moment is given by

\[
M = \frac{T_{\text{max.}}}{B} = \frac{NBAI}{B} = NAI.
\]

But the flux density produced by the solenoid along its axis is

\[
\frac{NI}{\mu_0 l} \quad (\text{see Table 10.3})
\]

* Magnetic moment may also be defined as \( T_m/H \); the student must take care to establish which definition is used in any literature he consults.
and hence the flux, \( \Phi \), inside it (assuming that it is long enough for the field to be uniform inside) is given by

\[
\Phi = \mu_0 \frac{NIA}{l} = \frac{\mu_0 M}{l}
\]

\[
\therefore \mu_0 M = \Phi l
\]

(10.15)

The lines of flux of a bar magnet are the same shape as those of a solenoid, and so it behaves in the same way in a magnetic field. Equations (10.14) and (10.15) are also applied in this case to define magnetic moment, and to show its relation to the flux and length of the magnet. In the latter case \( \Phi \) is the total flux through the centre of the magnet and \( l \) is the distance between its poles.

ELECTROSTATICS

10.14 Attraction between Point Charges

Coulomb's law for the force between two point electric charges \( Q_1, Q_2 \) a distance \( r \) apart is,

\[
F = K \frac{Q_1 Q_2}{r^2}
\]

In the electrostatic system of units \( K \) is made equal to unity and the definition of unit charge follows. In the rationalized m.k.s. system all the units in the equation are already defined: \( F \) in newtons, \( r \) in metres and \( Q \) in coulombs, so that \( K \) is a constant to be determined by experiment and not one that can arbitrarily be chosen.

In line with the practice in electromagnetism, \( 4\pi \) is introduced into the law of force equation, and it is written (to apply in a vacuum)

\[
F = \frac{Q_1 Q_2}{4\pi \varepsilon_0 r^2}
\]

(10.16)

where \( \varepsilon_0 \) is a constant known as the permittivity of free space. It will be shown later that \( \varepsilon_0 \) can be given the units farad m\(^{-1}\).

The value of \( \varepsilon_0 \) could be arrived at by a direct experimental determination, but it can be shown theoretically that \( \varepsilon_0, \mu_0 \) and the velocity of light, \( c \), are related by the expression

\[
\varepsilon_0 \mu_0 = \frac{1}{c^2}
\]

The velocity of light is very nearly \( 3 \times 10^8 \) m sec\(^{-1} \) (actually \( 2.998 \times 10^8 \)) and \( \mu_0 \) is \( 4\pi \times 10^{-7} \) henry m\(^{-1}\).

\[
\therefore \varepsilon_0 = \frac{1}{9 \cdot 10^{16} \cdot 4\pi \cdot 10^{-7}} = \frac{1}{36\pi} \times 10^{-9} \text{ farad m}^{-1}.
\]
10.15 Intensity of Electric Field and Potential Difference

The electric intensity at a point in an electric field is defined as the force exerted on a unit charge placed at that point.

There is unit electric intensity at a point if a charge of 1 coulomb placed at that point experiences a force of 1 newton. It follows that

\[ F = QE \quad (10.17) \]

where \( F \) is the force in newtons experienced by a charge \( Q \) coulombs in a field of intensity of \( E \) volt m\(^{-1}\). It will be shown shortly how the units quoted for \( E \) are arrived at.

Comparison of equations (10.16) and (10.17) shows that the field, \( E \), at distance \( r \) from a charge \( Q \) is given by

\[ E = \frac{Q}{4\pi\varepsilon_0 r^2} \quad (10.18) \]

The potential difference, \( V \) volts, between two points in an electric field is the work needed, in joules, to move a charge of 1 coulomb between the two points.

As in the c.g.s. development these definitions of intensity and potential lead to the relation

\[ E = -\frac{dV}{dx} \quad (10.19) \]

and this shows that \( E \) is measured in volt m\(^{-1}\).

Use of equation (10.19) leads to the expression for the potential, \( V \), at a distance \( r \) from a charge \( Q \):

\[ V = \frac{Q}{4\pi\varepsilon_0 r} \quad (10.20) \]

10.16 Electric Flux and Flux Density

An electric charge \( +Q \) coulombs is considered as the origin of lines or tubes of electric flux in the space surrounding it which end on a charge or charges of total \( -Q \) coulombs. Flux is measured in the same units as charge so that a total flux of \( Q \) coulombs originates from a charge of \( +Q \) coulombs. The electric flux density, \( D \), at a point in an electric field is defined as the amount of flux crossing unit area perpendicular to the lines of flux at that point. It is a vector quantity whose direction is the same as that of the lines of flux, and its units are coulomb m\(^{-2}\).

In a region where the flux density is constant

\[ D = \frac{Q}{A} \]

If the flux density is not constant the differential form is used, i.e.

\[ D = \frac{dQ}{dA} \]
10.17 Relation between $D$ and $E$ in a vacuum

It will be shown that in a vacuum $D$ and $E$ are proportional, and that the ratio $D/E$ is $\varepsilon_0$, the permittivity of free space.

Consider a point charge $+Q$ in a vacuum and imagine a sphere of radius $r$ drawn round it. The lines of flux from $Q$ are radially outwards, and hence cross the surface of the sphere at right angles. At the surface, the intensity of the electric field is given by

$$E = \frac{Q}{4\pi \varepsilon_0 r^2}.$$ 

The total flux crossing the surface of the sphere is $Q$ and the area is $4\pi r^2$, so that the flux density, $D$, is given by

$$D = \frac{Q}{4\pi r^2}.$$ 

$$\therefore \frac{D}{E} = \varepsilon_0.$$ 

The relation enables us to check the units given for $\varepsilon_0$: the units of $\varepsilon_0$ are

$$\frac{\text{coulomb m}^{-2}}{\text{volt m}^{-1}} = \text{farad m}^{-1}$$ 

as

$$\frac{\text{coulombs}}{\text{volts}} = \text{farads}.$$ 

10.18 Dielectrics

The presence of a dielectric has no effect on the value of $D$; this is because the dielectric introduces no free charges, and so makes no contribution to the electric flux. It is found, however, that the intensity of an electric field, $E$, produced by a certain distribution of charges is less in a dielectric medium than in a vacuum. This is explained in terms of the polarization of the medium (see Section 13.14).

It follows that the ratio $D/E$ in a dielectric medium is greater than $\varepsilon_0$. We write $D/E = \varepsilon_0 \varepsilon_r$, where $\varepsilon_r$ is a constant called the relative permittivity of the medium; its value is the same as that of the dielectric constant.

The expressions for the force between two charges, and electric intensity and potential due to a point charge (equations (10.16), (10.18), (10.20)), can be applied in a dielectric medium, if $\varepsilon_0$ is replaced in these equations by $\varepsilon_0 \varepsilon_r$.

10.19 Capacitance

The capacitance of a conductor is defined as the charge needed to raise its potential by unity.

If charge is in coulombs and potential in volts then the capacitance is in farads.

The method used in calculating the capacitance of a capacitor from its dimensions is indicated in the following example.
**Worked Example.** Find the capacitance of a parallel-plate capacitor which has plates of area \( A \) metre\(^2\) separated by a distance \( d \) metre, if the space between the plates contains a parallel-sided slab of dielectric of thickness \( t \) metre (\( t < d \)). The relative permittivity of the dielectric is \( \varepsilon_r \) (see Fig. 10.8). Edge corrections may be ignored.

![Fig. 10.8](image)

Suppose that the plates have charges \( +Q \) and \( -Q \) coulombs. Then the flux density, \( D \), in the space between the plates is given by

\[
D = \frac{Q}{A} \text{ coulomb m}^{-2}.
\]

The electric intensity, \( E_1 \), in the dielectric = \( D/\varepsilon_0 \varepsilon_r \) volt m\(^{-1}\).

The electric intensity, \( E_2 \), in the air space = \( D/\varepsilon_0 \) volt m\(^{-1}\).

Now the potential difference between two points in a field is (field \( \propto \) distance), so that the potential difference, \( V \), between the plates of the capacitor is

\[
V = E_2(d - t) + E_1t \text{ volts}
\]

\[
= \frac{D}{\varepsilon_0}[(d - t) + t/\varepsilon_r]
\]

\[
= \frac{Q}{\varepsilon_0 d}[(d - t) + t/\varepsilon_r]
\]

\[
\therefore \frac{Q}{V} = \frac{\varepsilon_0 A}{d - t(\varepsilon_r - 1)/\varepsilon_r}
\]

which is the required capacitance in farads.
<table>
<thead>
<tr>
<th>Symbol</th>
<th>C.G.S.</th>
<th>M.K.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r$, $l$, $d$, etc.</td>
<td>Lengths, etc. in centimetres</td>
<td>Lengths, etc., in metres</td>
</tr>
<tr>
<td>$A$</td>
<td>Area in cm$^2$</td>
<td>Area in m$^2$</td>
</tr>
<tr>
<td>$F$</td>
<td>Force in dynes</td>
<td>Force in newtons</td>
</tr>
<tr>
<td>Electro-</td>
<td>Flux density in gauss</td>
<td>Flux density (magnetic induction) in weber m$^{-2}$</td>
</tr>
<tr>
<td>magnetism</td>
<td>Magnetic intensity in oersteds</td>
<td>Magnetizing force in amp m$^{-1}$</td>
</tr>
<tr>
<td>$B$</td>
<td>Permeability</td>
<td>Relative permeability ($\mu_0 = 4\pi \times 10^{-7}$ henry m$^{-1}$)</td>
</tr>
<tr>
<td>$H$</td>
<td>Magnetic flux in maxwells</td>
<td>Magnetic flux in webers</td>
</tr>
<tr>
<td>$\mu$ ($= B/H$)</td>
<td>Current in e.m.u.</td>
<td>Current in amperes</td>
</tr>
<tr>
<td>$\mu_r = \frac{B}{\mu_0 H}$</td>
<td>E.M.F. or p.d. in e.m.u.</td>
<td>E.M.F. or p.d. in volts</td>
</tr>
<tr>
<td>$\Phi$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$I$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$E$, $V$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electro-</td>
<td>Charge in c.s.u.</td>
<td>Charge in coulombs</td>
</tr>
<tr>
<td>statics</td>
<td>Electric intensity in c.s.u.</td>
<td>Electric intensity in volt m$^{-1}$</td>
</tr>
<tr>
<td>$Q$</td>
<td>Electric displacement in e.s.u.</td>
<td>Electric displacement (flux density) in coulomb m$^{-2}$</td>
</tr>
<tr>
<td>$E$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$D$</td>
<td>Dielectric constant</td>
<td>Relative permittivity ($\epsilon_0 = \frac{1}{96\pi} \times 10^{-9}$ farad m$^{-1}$)</td>
</tr>
<tr>
<td>$\epsilon$ ($= D/E$)</td>
<td>Potential in e.s.u.</td>
<td>Potential in volts</td>
</tr>
<tr>
<td>$\epsilon_r = \frac{D}{\epsilon_0 E}$</td>
<td>Capacitance in e.s.u.</td>
<td>Capacitance in farads</td>
</tr>
<tr>
<td>$V$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 10.2 Symbols and units
### ELECTRICITY IN M.K.S. UNITS

<table>
<thead>
<tr>
<th>Electromagnetism</th>
<th>C.G.S.</th>
<th>M.K.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biot–Savart</td>
<td>$H = \frac{\Sigma l B \sin \theta}{r^2}$</td>
<td>$H = \frac{\Sigma l B \sin \theta}{4\pi r^2}$</td>
</tr>
<tr>
<td>Circuital theorem for a path enclosing a current $I$ at distance $a$ from a long straight wire at centre of a circular coil of radius $r$ with $N$ turns at distance $x$ along axis of coil on the axis of a solenoid of $n$ turns per unit length inside an infinitely long solenoid on a straight wire of length $l$ carrying a current $I$ at right angles to a magnetic field of flux density $B$ electromagnetic induction</td>
<td>Work done by unit pole $= 4\pi I$</td>
<td>$dHdl = I$</td>
</tr>
<tr>
<td></td>
<td>$H = \frac{2I}{a}$</td>
<td>$H = \frac{I}{2\pi a}$</td>
</tr>
<tr>
<td></td>
<td>$H = \frac{2\pi NI}{r}$</td>
<td>$H = \frac{NI}{r}$</td>
</tr>
<tr>
<td></td>
<td>$H = \frac{2\pi NI r^2}{(r^2 + x^2)^{3/2}}$</td>
<td>$H = \frac{nI}{2}(\cos \theta_1 - \cos \theta_2)$</td>
</tr>
<tr>
<td></td>
<td>$H = 2\pi l (\cos \theta_1 - \cos \theta_2)$</td>
<td>$H = \frac{nI}{2}(\cos \theta_1 - \cos \theta_2)$</td>
</tr>
<tr>
<td></td>
<td>$F = B l I$</td>
<td>$F = BI l$</td>
</tr>
<tr>
<td>$E = -\frac{d(N\Phi)}{dt}$</td>
<td>$E = -\frac{d(N\Phi)}{dt}$</td>
<td></td>
</tr>
</tbody>
</table>

### Table 10.3 Some formulae in electromagnetism and electrostatics

<table>
<thead>
<tr>
<th></th>
<th>C.G.S.</th>
<th>M.K.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrostatics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coulomb’s law (in a vacuum)</td>
<td>$F = \frac{Q_1 Q_2}{r^2}$</td>
<td>$F = \frac{Q_1 Q_2}{4\pi \varepsilon_0 r^2}$</td>
</tr>
<tr>
<td>At distance $r$ from a point charge $Q$ in a vacuum</td>
<td>$E = \frac{Q}{r^2}$</td>
<td>$E = \frac{Q}{4\pi \varepsilon_0 r}$</td>
</tr>
<tr>
<td>Relation between $E$ and $V$</td>
<td>$V = \frac{Q}{r}$</td>
<td>$V = \frac{Q}{4\pi \varepsilon_0}$</td>
</tr>
<tr>
<td>Gauss’ theorem</td>
<td>$E = -\frac{dV}{dx}$</td>
<td>$E = -\frac{dV}{dx}$</td>
</tr>
<tr>
<td>Parallel-plate capacitor</td>
<td>$C = \frac{\varepsilon A}{4\pi d}$</td>
<td>$C = \frac{\varepsilon_0 A}{d}$</td>
</tr>
<tr>
<td>Isolated sphere of radius $a$</td>
<td>$C = a$</td>
<td>$C = \frac{4\pi \varepsilon_0 a}{d}$</td>
</tr>
<tr>
<td>Coaxial cable</td>
<td>$C = \frac{e l}{\ln b/a}$</td>
<td>$C = \frac{2\pi \varepsilon_0 l}{\ln b/a}$</td>
</tr>
<tr>
<td>Force on unit area of the surface of a charged conductor</td>
<td>$F = 2\pi \sigma^2$</td>
<td>$F = \frac{\sigma^2}{2\pi l}$</td>
</tr>
<tr>
<td>Attracted-disc electrometer</td>
<td>$F = \frac{V^2 A}{8\pi d^2}$</td>
<td>$F = \frac{V^2 A}{8\pi d^2}$</td>
</tr>
</tbody>
</table>

### Table 10.4 Relations between units in c.g.s. and m.k.s. systems

<table>
<thead>
<tr>
<th></th>
<th>1 amp = $10^{-4}$ e.m.u.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>1 coulomb = $10^{-4}$ e.m.u.</td>
</tr>
<tr>
<td>Charge</td>
<td>1 weber $m^{-2}$ = $10^4$ gauss</td>
</tr>
<tr>
<td>Magnetic flux density</td>
<td>1 weber = $10^8$ maxwell</td>
</tr>
<tr>
<td>Magnetic flux</td>
<td>1 amp $m^{-1}$ = $4\pi \times 10^{-3}$ oersted</td>
</tr>
<tr>
<td>Magnetic intensity</td>
<td>1 ohm = $10^9$ e.m.u.</td>
</tr>
<tr>
<td>Resistance</td>
<td>1 volt = $\frac{1}{3} \times 10^4$ c.s.u. = $10^8$ e.m.u.</td>
</tr>
<tr>
<td>Potential</td>
<td>1 volt $m^{-1}$ = $\frac{1}{3} \times 10^4$ c.s.u.</td>
</tr>
<tr>
<td>Electric intensity</td>
<td>1 farad = $9 \times 10^{11}$ c.s.u.</td>
</tr>
</tbody>
</table>
THE QUANTUM THEORY

11.1 Black-body Radiation

It was mentioned in Section 5.18 that the first attempts to obtain a formula to describe the intensity distribution in the spectrum of the radiation from a black body were unsuccessful. The way in which the problem had been treated was to assume that the radiation was emitted by a large number of harmonic oscillators in the black body, each one of which could oscillate with any energy. This treatment led to the unsuccessful Rayleigh–Jeans law.

In 1901 Planck derived a formula which was a very good fit with experimental results by making the assumption that an oscillator of frequency \( \nu \) can have only certain discrete amounts of energy; and that, when the oscillator emits, the energy of the radiation can have only one particular value, dependent on \( \nu \). He showed that this energy is equal to \( h\nu \), where \( h \) is a constant (now known as Planck’s constant) having the value \( 6.63 \times 10^{-27} \) erg sec. This suggestion was a very revolutionary one and met with much opposition, but it certainly solved the black-body problem very neatly – of this there was no doubt. The formula deduced by Planck for the emissive power, \( E_\lambda \), of a black body at wavelength \( \lambda \) is

\[
E_\lambda = \frac{8\pi hc}{\lambda^5} \cdot \frac{1}{e^{hc/\lambda kT} - 1},
\]

where \( c \) is the velocity of light and \( k \) is Boltzmann’s constant.

What Planck’s hypothesis means is that any oscillator (even a simple pendulum) of frequency \( \nu \) can possess energy \( E = 0, h\nu, 2h\nu, 3h\nu, \ldots nh\nu, \ldots \), but no intermediate values. If it radiates or absorbs energy it must do so in units of \( h\nu \). The very small value of \( h \) means that in any large-scale system the spacing between the allowed energies is so small compared with the actual value of the energy that, in practice, one gets the effect of a continuous range of energies. In smaller systems, such as the bound electrons in an atom, the effects of discrete energy values are of paramount importance.

11.2 The Photo-electric Effect

Planck’s hypothesis was that radiation is emitted or absorbed in units (quanta) of \( h\nu \). It was later suggested by Einstein that the same considerations apply to the radiation itself, i.e. that it consists of small packets of energy (often referred to as photons) travelling through space with the velocity of light. Each photon has energy \( h\nu \), where \( \nu \) is the fre-
frequency of the radiation concerned; it also has momentum which may be shown to be equal to $h\nu/c$.

In support of this theory may be quoted evidence obtained from a study of the photo-electric effect. If the surface of a metal is irradiated by ultra-violet radiation, then electrons may be liberated from it. If the experiment is performed in an evacuated vessel the number of electrons ejected per second may be measured by collecting them at a positive electrode and measuring the current (Fig. 11.1). The energy of the electrons may be determined by making the electrode A in Fig. 11.1 negative with respect to the metal and finding the potential difference between the two which just prevents electrons from reaching A.

![Diagram of photo-electric effect](image)

**Fig. 11.1**

The results of such experiments may be summarized as follows:

(i) below a certain frequency of the incident radiation no electrons are emitted;

(ii) when electrons are emitted they have a range of energies up to maximum, $E$, which depends on the frequency of the radiation according to the equation $E = h\nu - \phi$, where $\phi$ is a constant for a given metal;

(iii) the intensity of the beam of radiation determines the number of electrons emitted per second but does not affect their maximum energy.

These observations cannot be explained by classical theory, but they can easily be explained in terms of the quantum theory. A photon of frequency $\nu$ has energy $h\nu$, and when it collides with an electron in the surface of the metal this energy may be given to the electron. For an electron to leave the surface it must have a certain minimum energy, known as the work function ($\phi$) of the metal. Thus if $h\nu < \phi$ no electrons are emitted, in agreement with (i) above.

When $h\nu > \phi$ the maximum energy which an electron leaving the surface may have is $h\nu - \phi$, as in (ii). This energy depends on the energy of each photon, but not on the intensity of the beam. An intense beam of radiation means a large number of photons per second, and
hence a large number of electrons emitted per second, but the energy of each photon, and hence the maximum energy of the emitted electrons, is unaltered.

11.3 Waves or Photons?

The quantum theory of radiation provides a convincing explanation of several phenomena which involve the interaction of radiation with matter (e.g. the photo-electric effect, and also the change in frequency which occurs when X-rays are scattered by matter – the Compton effect). On the other hand, the wave theory accounts for interference and diffraction effects and for polarization; these phenomena do not make sense in terms of photons. Any explanation in terms of an interaction of photons with each other has been shown to be incorrect by an experiment in which interference fringes were photographed using light so weak that the average distance between the photons was several metres.

It is impossible to devise a crucial experiment to decide whether light is a wave or is a stream of photons; any such attempt is foiled by interactions between the measuring apparatus and the radiation (see Section 11.7).

The view taken is that the two pictures are complementary. We may say that in certain respects light resembles waves and in other respects a stream of particles. To answer the question 'is light a wave motion or a stream of particles?' remains impossible; the nature of light cannot be explained by a single picture.

11.4 Matter Waves

If light, which was firmly held to be a wave motion, can exhibit particle properties, is it possible that particles may sometimes behave like waves? The first demonstration that this is indeed so was made by Davisson and Germer in 1927. They showed that, under suitable conditions, the reflection of beams of electrons from the surface of metals is not regular but that a diffraction pattern is formed, the reflected beam being more intense in some directions than others. In the same year G. P. Thomson was able to obtain sufficiently thin metal foils to observe the passage of electrons through them, and again characteristic diffraction patterns were formed. It looks as if the electrons are behaving, not like particles, but like waves, which form a diffraction pattern when reflected from the metallic lattice, similar to that formed by reflecting X-rays from a crystal lattice.

It was shown theoretically by de Broglie that the wave associated with a particle of momentum \( p \) has a wavelength \( \lambda = \frac{h}{p} \). This is known as the de Broglie wavelength. By accelerating electrons through very high potential differences it is possible to obtain values of \( \lambda \) much less than 1 Å, and this has led to the development of the electron microscope. The de Broglie wavelength may easily be calculated as follows.

Suppose that an electron of charge e e.s.u. and mass \( m \) g is accelerated
from rest by a potential difference of $V$ volts (\(= V/300\) e.s.u.). The loss in potential energy of the electron is $eV/300$ ergs. The gain in kinetic energy of the electron is $\frac{1}{2}mv^2$, where $v$ is the velocity it acquires, hence

$$\frac{1}{2}mv^2 = \frac{eV}{300}.$$ 

\[
\therefore v = \left(\frac{eV}{150m}\right)^{\frac{1}{2}},
\]

\[
\therefore \text{momentum } p = mv = \left(\frac{meV}{150}\right)^{\frac{1}{2}},
\]

\[
\therefore \text{de Broglie wavelength, } \lambda = \frac{h}{p} = h\left(\frac{150}{meV}\right)^{\frac{1}{2}}. \quad (11.1)
\]

### 11.5 Wave Mechanics

We have seen that if a large number of electrons are shot through a thin metal foil a diffraction pattern is formed on a screen placed beyond the foil, i.e. at some points on the screen a large number of electrons arrive, at other points very few, and perhaps at some places none at all. We may now ask what happens if just one electron is passed through the foil. In this case it is obvious that a pattern cannot be formed; the electron can go to only one point on the screen. What, then, determines the point to which it does go? The answer to this question was first suggested by Born.

Born put forward the idea that the complete pattern formed when a large number of electrons are used shows that at some points (high intensity) there is a high probability of an electron arriving and at others (low intensity) a small probability. Extending this idea to the case of one electron, one can predict that for each point on the screen there is a certain probability that the electron will arrive at that point.

We have seen that for a beam of electrons each having momentum $p$ the diffraction pattern formed is the same as that which would be formed by a wave of the de Broglie wavelength $\lambda = h/p$. For a single electron one may treat it as a wave of this wavelength, and then, in accordance with Born's idea, interpret the intensity (square of the amplitude) of this wave at any point as proportional to the probability that the electron will be found at that point. This is the basis of wave mechanics, which has to replace classical mechanics when dealing with small particles. The way in which one deals with the motion of these particles is to treat each one as a wave and then calculate the behaviour of this wave under the given conditions. The square of the amplitude of the resultant wave at some point at a given time gives the probability of the particle being at that point at that time.

### 11.6 Cause and Effect

One of the great precepts of classical mechanics is that if one is given the initial state of a system of interacting bodies, then the positions and
momenta of the bodies at time $t$ later are completely determined. In many cases the situation is too complex to enable future events to be calculated, but nevertheless they are exactly determined by the initial conditions. In other words, if the same set of initial conditions could be produced again and again the positions and momenta of the bodies after a time $t$ would always be exactly the same.

According to wave mechanics, however, this is not so. The wave function may be calculated precisely, but this leads only to probabilities when one enquires about the positions of the interacting bodies after time $t$. We can no longer speak in terms of a definite cause and effect in the interaction between objects.

The student may well ask how it is, then, that classical mechanics has been so successful. The simple answer to this is that when dealing with anything other than very small particles (e.g. electrons), the momentum $\lambda$ is so large that $\lambda = h/\rho$ is very small. Now the effect of a small $\lambda$ is that the diffraction pattern produced is so narrow that the probability of any divergence from behaviour predicted by classical methods is vanishingly small.

11.7 Heisenberg Uncertainty Relation

We may take the ideas of the previous section a stage further and say that not even the initial conditions may be known exactly. Suppose, for example, that we wish to measure the momentum and position of a particle at some instant. If we look at the particle with a microscope in order to determine its position it is necessary that the particle be bombarded by at least one photon which, after the collision, enters the microscope. Now, the collision will alter the momentum of the observed particle by a certain amount (Fig. 11.2). If we knew exactly which way the photon went after the collision it would be possible to calculate this change in momentum, but all that we do know about it is that it enters the aperture of the microscope. There is thus an uncertainty, $\Delta\rho$, in the momentum of the particle.

There is also an uncertainty, $\Delta x$, in our knowledge of the position of the particle; this arises because of the limit imposed by the resolving power of the microscope. Suppose that we try to reduce $\Delta x$ by increasing the aperture of the microscope (which improves its resolving power). The effect is to make $\Delta \rho$ larger because the uncertainty in the direction of the recoiling photon is increased. It seems, then, that we cannot know both $x$ and $\rho$ with precision; the more precisely we know one, the less well determined is the other. According to Heisenberg, this is expressed by the relation $\Delta x \cdot \Delta \rho \sim h$, and this sets a theoretical limit on all measurements.

Another way in which one might attempt to determine the position
of a particle (e.g. an electron) is to cause it to pass through a narrow slit. If it is known that the electron has gone through the slit, then its position at that instant is uncertain by an amount equal to the width of the slit \(\Delta x\). However, the passage of the electron through the slit introduces an uncertainty in the component of its momentum parallel to the plane of the slit. This arises because the wave associated with the electron is diffracted by the slit, so that the particle itself may be deflected from its original path. This is illustrated in Fig. 11.3. If the slit is very narrow \(\Delta x\) small) the diffraction fringes are wide \(\Delta p\) large), conversely, if \(\Delta p\) is to be small, \(\Delta x\) must be large. Suppose that the electron arrives at the slit with momentum for \(p\) perpendicular to the plane of the slit, and that the screen is far enough away for Fraunhofer diffraction to obtain. After passing through the slit there is a very high probability that the electron will move in a direction between the normal and the direction of either of the first zeros of the diffraction pattern. The uncertainty in the component of its momentum in a direction perpendicular to its original path is thus given by

\[
\Delta p \sim p \sin \theta_1,
\]

where

\[
\sin \theta_1 = \frac{\lambda}{\Delta x} \quad \text{(see section 4.21)},
\]

\[
\therefore \Delta p \sim \frac{p \lambda}{\Delta x}.
\]

But

\[
\lambda = \frac{h}{p} \quad \text{(de Broglie wavelength)},
\]

\[
\therefore \Delta p \Delta x \sim h,
\]

which is the Heisenberg uncertainty relation.

We have seen that two methods of observing the position of a particle have each caused the momentum of the particle to change. In fact, whatever measuring instrument one devises, it will always be subject to
the limitations imposed by the Heisenberg uncertainty relation. Another way of putting this is to say that any observations made are bound to influence the observed object. The disturbance produced by the observation is appreciable only when one is dealing with very small particles (i.e. fundamental particles).

11.8 The Bound Electron. Energy Levels

We shall now see how to apply the principles developed in previous sections to a particular case – that of an electron confined to a certain region. Consider the simple case of an electron free to move along the x-axis between two perfectly reflecting walls at $x = 0$ and $x = a$. According to classical (i.e. pre-quantum theory) ideas, the electron could have any energy and would bounce backwards and forwards between the two walls. Now let us examine what wave mechanics predicts.

If the electron has momentum $p$, then its de Broglie wavelength $\lambda = h/p$. This wave must form a standing-wave system, similar to that on a vibrating string. The square of the amplitude of the wave indicates the probability of the electron being at any particular point, so that, as this probability is zero at each wall, the standing wave must have a node at each end. Several modes of vibration are possible, corresponding to the harmonics of a vibrating string (see Fig. 11.4). There must be an integral number of half-wavelengths between the walls. The possible wavelengths are thus given by

$$a = \frac{\lambda}{2}, \frac{2\lambda}{2}, \frac{3\lambda}{2}, \ldots \frac{n\lambda}{2}, \ldots$$

i.e.

$$\lambda = 2a, \frac{2a}{2}, \frac{2a}{3}, \ldots \frac{2a}{n}, \ldots$$

This means that the momentum $p$ ($= h/\lambda$) can have only the values,

$$p = \frac{h}{2a}, \frac{2h}{2a}, \frac{3h}{2a}, \ldots \frac{n h}{2a}, \ldots$$

![Fig. 11.4](image-url)
This quantization must apply also to the energy because it is related to the momentum by the equations

\[ p = mv, \]
\[ E = \frac{1}{2} mv^2, \]

where \( m \) is the mass of the electron, \( v \) its velocity and \( E \) its kinetic energy.

Hence

\[ E = \frac{h^2}{2m} \] \hspace{1cm} (11.2)

so that the allowed values of energy are

\[ E = \frac{h^2}{8ma^2} \cdot 1^2, \frac{h^2}{8ma^2} \cdot 2^2, \frac{h^2}{8ma^2} \cdot 3^2, \ldots \frac{h^2}{8ma^2} \cdot n^2, \ldots \]

The orbital electrons in atoms are, of course, bound electrons. The situation is somewhat different from that discussed above, because the electrons are constrained to move in orbits and not simply in straight lines, but the same sort of considerations apply and account for the discrete energy levels which these electrons have. This point will be discussed more fully in Chapter 12.

11.9 The Schroedinger Equation

We now derive a general wave equation which may be used to predict the behaviour of a particle in some field of force – e.g. an electron in an electric field.

In Section 3.6 it was shown that a harmonic wave travelling along the axis with velocity \( v \) may be represented by

\[ \psi = a \sin \frac{2\pi}{\lambda} (x - vt) \] (see equation (3.11)).

Differentiating this partially with respect to \( x \) gives

\[ \frac{\partial \psi}{\partial x} = a \frac{2\pi}{\lambda} \cos \frac{2\pi}{\lambda} (x - vt), \]

and

\[ \frac{\partial^2 \psi}{\partial x^2} = -a \left( \frac{2\pi}{\lambda} \right)^2 \sin \frac{2\pi}{\lambda} (x - vt) \]

\[ = \left( \frac{2\pi}{\lambda} \right)^2 \psi. \]

\[ \therefore \frac{\partial^2 \psi}{\partial x^2} + \left( \frac{2\pi}{\lambda} \right)^2 \psi = 0. \]

If this equation is used to describe the behaviour of a particle then we must as usual put \( \lambda = h/p \).

\[ \therefore \lambda^2 = \frac{h^2}{p^2} = \frac{h^2}{2mE} \] (see equation (11.2)),

where \( E \) is the kinetic energy of the particle.

\[ \therefore \frac{\partial^2 \psi}{\partial x^2} + \frac{8\pi^2 mE}{h^2} \psi = 0. \]
In the case of an electron moving in an electric or magnetic field it will possess potential as well as kinetic energy. If we now use $E$ to denote its total energy, and $V$ to denote its potential energy, then its kinetic energy is $E - V$ and the equation becomes

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{8\pi^2 m(E - V)}{\hbar^2} \psi = 0.$$  

This is the one-dimensional form of the famous time-independent Schrödinger equation on which much of wave mechanics is based. In three dimensions the equation becomes

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} + \frac{8\pi^2 m(E - V)}{\hbar^2} \psi = 0.$$  

To use this equation to predict the motion of an electron in an electric field one inserts the appropriate value of $V$ (which may well be a function of $x, y, z$) and then solves the equation for $\psi$. $|\psi|^2$ is then interpreted as proportional to the probability of finding the particle at a given point, in the manner previously discussed.

**QUESTIONS**

1. Show that Planck’s radiation formula satisfies Wien’s law and that at long wavelengths ($\hbar/\lambda \ll kT$) it approximates to the Rayleigh–Jeans’ law (see Section 5.18).

2. Using $h = 6.63 \times 10^{-27}$ erg sec, $e = 4.80 \times 10^{-10}$ e.s.u, $m = 9.11 \times 10^{-28}$ g, calculate the wavelength of an electron which has been accelerated through a potential difference of 5 kV.
12

ATOMIC SPECTRA AND THE PERIODIC TABLE

12.1 Hydrogen Spectrum

Following the production by Rowland (1882) of good diffraction gratings there was a period of intense interest in spectroscopy. The most profitable field of study was that of the line-emission spectra of the elements. Several workers announced the discovery of series of lines in the spectra of certain elements. The first successful attempt to find a formula which related all the members of a series was made by Balmer in 1885 for the visible lines in the hydrogen spectrum. Balmer’s formula for this series may be expressed as

\[ \frac{1}{\lambda} = R \left( \frac{1}{4} - \frac{1}{m^2} \right) \]

each value of \( m (>2) \) giving the wavelength \( \lambda \), of one line in the series. \( R \) is a constant known as the Rydberg constant and has the value 109,700 cm\(^{-1}\). The series tends to a limit as \( m \rightarrow \infty \), given by \( 1/\lambda = R/4 \). Fig. 12.1 gives a rough indication of the nature of the series.

![Fig. 12.1](image)

It seemed probable that Balmer’s law might be a particular case of the more general formula

\[ \frac{1}{\lambda} = R \left( \frac{1}{n^2} - \frac{1}{m^2} \right) \]  \hspace{1cm} (12.1)

each value of \( n \) giving a complete series, the individual lines of which are given by integral values of \( m (>n) \). Several of these series have been discovered and named after their discoverers, e.g.

\[ n = 1, \quad \text{Lyman (ultra-violet)} \quad m = 2, 3, \ldots \]
\[ n = 2, \quad \text{Balmer (visible)} \quad m = 3, 4, \ldots \]
\[ n = 3, \quad \text{Paschen (infra-red)} \quad m = 4, 5, \ldots \]

etc.
12.2 The Bohr Atom

The first successful attempt to construct an atomic model which would explain how atomic spectra are produced, and give a quantitative agreement with Balmer’s formula, was made by Bohr in 1913. His theory of the structure of the atom was based on two theories by other great scientists of that era: the quantum theory postulated by Planck in 1901, and Rutherford’s nuclear model of the atom (1911). The first of these revolutionary theories has already been discussed in Chapter 11, and the reader should be acquainted with Rutherford’s theory. Briefly, Rutherford’s model is that of a central positively charged nucleus, containing most of the mass of the atom, surrounded by negatively charged electrons. The simplest atom, that of hydrogen, has one electron and a nucleus with a positive charge equal in magnitude to the negative charge on the electron. The hydrogen nucleus is known as a proton. Atoms of other elements have more electrons (the actual number being known as the atomic number of the element) and in each case nuclei containing a number of protons equal to the number of electrons outside the nucleus.

Bohr’s picture of the hydrogen atom is that of a proton of relatively large mass (1,845 times that of the electron), with one electron moving in a circular orbit round it. The electrostatic force of attraction between the proton and the electron provides the centripetal force necessary to keep the electron in its orbit (Fig. 12.2).

Suppose that the radius of the orbit is \( r \) cm and that the charges on the electron and the proton are respectively \(-e\) and \(+e\) e.s.u. The force, \( F \) dynes, acting on the electron is then given by \( F = \frac{e^2}{r^2} \), using Coulomb’s law of electrostatics. If the speed of the electron in its orbit is \( v \) cm sec\(^{-1}\) and its mass \( m \) g, then the central force necessary to maintain it in its orbit is \( mv^2/r \) dynes. Equating these two forces gives

\[
\frac{e^2}{r^2} = \frac{mv^2}{r}
\]

(12.2)

The kinetic energy of the electron is \( \frac{1}{2}mv^2 \), and its potential energy is \(-e^2/r\) if we take infinity as the zero of potential energy.

But, by equation (12.2),

\[
\frac{e^2}{r} = mv^2,
\]

\[
\therefore \text{kinetic energy} = \frac{1}{2} \frac{e^2}{r}.
\]
Thus the total energy of the electron in its orbit is
\[
\frac{1}{\frac{2}{3}} e^2 \frac{2}{r} - e^2 \frac{2}{r} = -\frac{1}{2} e^2 \frac{2}{r} .
\]

12.3 Explanation of Atomic Spectra by the Bohr Atom

It is in the development of his theory to explain the production of atomic spectra that Bohr laid the foundations for a major advance in atomic theory. This was such a notable achievement, and so revolutionized the whole approach to the subject, that it is worth studying in some detail. What has to be explained is why, under suitable conditions, an atom can emit radiation, and further, why the spectrum of this radiation is the characteristic line-emission spectrum.

Bohr’s theory of circular orbits as developed so far could account for the emission of radiation from an atom. An electron moving in a circle, and hence in accelerated motion, emits radiation of frequency equal to the frequency of rotation in its orbit. If an electron bound in an atom obeyed the same rule, then it would also radiate, and at first sight it seems that this might provide an explanation for the existence of line spectra. As soon as one attempts to develop a theory along these lines, however, certain grave objections arise.

The first difficulty encountered on this theory arises from energy considerations. If the electron is emitting radiation it is of necessity losing energy (the energy associated with the wave it emits), and if it is losing energy it must begin to spiral in towards the nucleus. It has to do this in order to comply with the principle of conservation of energy; the orbits of smaller radius are those of smaller energy (see equation (12.3)). It can be shown that if the electron did behave in this way it would come into the nucleus and the atom collapse in about $10^{-8}$ sec.

The second objection is that as the electron moved into smaller orbits its angular velocity, and hence its frequency of revolution, would increase. According to the theory of emission of radiation by a circulating charge, this would result in a change in frequency of the radiation. At any instant there would be atoms in all possible orbits, so that the spectrum of the radiation coming from the atoms would be a continuous one, whereas it is in fact a line spectrum.

The only conclusion one can draw is that the electrons in orbit in an atom do not radiate by virtue of their motion in a circle, i.e. they do not obey classical electromagnetic theory. So the question, ‘Why does an atom emit radiation?’ still remains unanswered, and to answer it Bohr had to extend his original ideas.

The first clue was provided by the ideas of the quantum theory which had been suggested by Planck in 1901. Planck had put forward evidence to indicate that radiant energy is emitted in packets or quanta of $hv$,
where \( \nu \) is the frequency of the radiation. It follows that when an atom emits a quantum of radiation the energy of the atom changes by a discrete amount, dependent on the frequency of the radiation emitted. Now an atom emits only radiation of certain frequencies (line spectrum), so that the changes in energy which take place on emission are limited and are the same for all atoms of the same element. These considerations led Bohr to postulate that the orbital electron can possess only certain discrete energies. This means that the electron can occupy only certain orbits: those whose radius is such that the electron has one of the allowed values of energy. While the electron remains in one of these orbits it cannot emit radiation, but when the electron moves from one orbit to another it undergoes a change in energy. If the electron moves from one of high energy to one of lower energy (smaller radius) the excess energy is emitted as a quantum of radiation.

Suppose that in its original orbit the electron has energy \( E_2 \), and that it then falls to an orbit where it has energy \( E_1 \). The change in energy is \( E_2 - E_1 \), and if this is emitted as one quantum its frequency is given by \( E_2 - E_1 = \nu h \).

In general, a transition from the \( m \)th orbit, in which the electron has energy \( E_m \), to the \( n \)th orbit, where it has energy \( E_n \), results in the emission of a quantum of frequency \( \nu \) given by

\[
\nu = E_m - E_n
\]  

(12.4)

This fits in very well with the production of a line spectrum, because a jump from one orbit to another results in radiation of one definite frequency, and when one takes into account changes between other pairs of possible orbits, then the line spectrum begins to take shape.

In the case of hydrogen this view is also strengthened by an examination of Balmer's formula. It will be seen that the expression for the frequency is also the difference between two quantities, just as it is in the equation above. If both sides of equation (12.1) are multiplied by \( h \), where \( h \) is Planck's constant, and \( c \) is the velocity of light, then

\[
\nu = \frac{hcR}{n^2} - \frac{hcR}{m^2}
\]

and the similarity to equation (12.4) is now obvious. Comparison of the two equations suggests that the energy of the \( n \)th orbit is given by

\[
E_n = -\frac{hcR}{n^2}
\]  

(12.5)

On the basis of this equation we may draw up a diagram showing the energy levels of the hydrogen atom. The level of lowest energy (\( n = 1 \)) is known as the ground state of the atom and higher levels (\( n = 2, 3, \ldots \)) as excited states. Fig. 12.3 shows such a diagram. The atoms in hydrogen gas are usually in the ground state, but if energy is supplied to some of them (e.g. by passing an electric current through the gas in a discharge
tube) they may be raised into one of the excited states. An atom in an
excited state is unstable, and in a very short time (the average is about
$10^{-8}$ sec) the electron jumps to a
smaller orbit (lower energy); the
difference in energy is emitted as a
quantum of radiation in accordance
with equation (12.4).

If the electron makes its transition
direct to the ground state ($n = 1$),
then the frequency emitted is in the
Lyman series, if to the first excited
state ($n = 2$) the line is in the Balmer
series and so on.

An atom in the excited state
$n = 3$ has two ways of returning to
its ground state: either directly,
giving a line in the Lyman series
(call its frequency $\nu_{31}$); or by making
a transition first to the $n = 2$
level (giving a line in the Balmer series of
frequency $\nu_{32}$), and then from $n = 2$
to $n = 1$, resulting in an emitted line
in the Lyman series ($\nu_{21}$). Reference
to Fig. 12.3 shows that

$$\nu_{31} = \nu_{32} + \nu_{21}.$$  

This principle was discovered empirically by Ritz some years before
Bohr postulated his theory; it is known as the Ritz combination principle.
This constitutes further evidence in favour of the Bohr atom.

12.4 The Frank–Hertz Experiment

This experiment, carried out in 1914, constitutes one of the strongest
pieces of direct evidence of the existence of energy levels. The principle
of the experiment is that if a
moving electron collides with an
atom it may cause the atom to
make a transition into one of its
excited states. According to
Bohr’s theory of energy levels, no
such excitation is possible unless
the electron has energy at least
equal to the difference between
the ground state and the first ex-
cited state ($E_2 - E_1$).

![Fig. 12.4](image)

The basis of the experimental arrangement is illustrated in Fig. 12.4.
Electrons from the heated cathode, C, were accelerated by a positive
potential difference, $V_3$, applied between $C$ and the grid $G$. A reverse potential difference, $V_2$ (about $\frac{1}{2}$ volt), applied to the plate $P$ was such that electrons with energy less than about $\frac{1}{3}$ electron-volt after passing through $G$ would not be able to reach $P$. The galvanometer, $M$, recorded the rate at which electrons were collected by $P$. The envelope was filled with a gas or vapour at low pressure.

As $V_1$ was increased the galvanometer showed a steadily increasing reading until a certain value of $V_1$ was reached, when the current suddenly dropped (Fig. 12.5). At this stage the gas or vapour began to emit radiation. Further increase in $V_1$ caused the current to rise again until a certain value of $V_1$ was reached at which there was another marked fall. This process continued as $V_1$ was increased.

The explanation in terms of energy levels is very simple. At low values of $V_1$ the electrons do not acquire sufficient energy to cause excitation of the atoms, so that a large number of them pass through $G$ with sufficient energy to reach $P$. As $V_1$ is increased the electrons acquire more energy until a value is reached at which they have sufficient energy just before reaching $G$ to cause excitation of the atoms of the gas. Those electrons which make such collisions lose all their energy, and are thus unable to reach $P$; hence a sharp drop in the current recorded by the galvanometer. The excited atoms then fall back to their ground state, with the corresponding emission of radiation.

As $V_1$ is increased further, electrons which have made collisions can still acquire sufficient energy before reaching $G$ to pass through and reach $P$. At an even higher value of $V_1$ an electron which excites one atom well before reaching $G$ can acquire sufficient energy to make a second collision of this type just before $G$, resulting in the second drop in galvanometer current.

A measurement of the difference between two peaks ($V$ in Fig. 12.5) enables one to calculate the energy which an electron needs if it is to cause excitation of an atom. This result agrees very well with the value of $E_2 - E_1$ obtained from spectroscopic data.
12.5 Quantization of Angular Momentum

We proceed now with some calculations involving the Bohr orbits. To do this it is necessary to introduce a further postulate made by Bohr; it concerns the angular momentum of the orbital electron. To account for the stationary states in the hydrogen atom Bohr had to introduce some quantization rule, and for various reasons he chose to apply it to angular momentum. His postulate was that the angular momentum of the electron in its orbit must be a multiple of $\hbar/2\pi$. This seems an entirely arbitrary assumption to make, and its justification at this stage must be that it leads to correct results. In Section 12.8 we shall see that the assumption makes sense in terms of wave mechanics.

Thus:

$$\text{angular momentum} = n\frac{\hbar}{2\pi} \quad \ldots \quad (12.6)$$

where $n$ is an integer.

For an electron moving with velocity $v$ in an orbit of radius $r$, we have:

$$\text{angular momentum} = mv = n\frac{\hbar}{2\pi}$$

But, from equation (12.2),

$$v = \frac{e}{(mr)^{1/2}}$$

$$\therefore \frac{me}{(mr)^{1/2}} = n\frac{\hbar}{2\pi}$$

$$\therefore r = \frac{n^2h^2}{4\pi^2me^2} \quad \ldots \quad (12.7)$$

Substituting this in equation (12.3), one obtains for the energy of the electron in the hydrogen atom:

$$E_n = -\frac{e^2}{r} = -\frac{2\pi^2me^4}{n^2h^2} \quad \ldots \quad (12.8)$$

It will be seen that this expression for $E_n$ depends on $n$ in the same way as that obtained in Section 12.3 (equation (12.5)) from a consideration of energy levels in conjunction with Balmer's formula.

Equating the right-hand sides of equations (12.5) and (12.8) gives

$$\frac{\hbar c R}{n^2} = -\frac{2\pi^2me^4}{n^2h^2}$$

$$\therefore R = \frac{2\pi^2me^4}{\hbar c}$$

(12.9)

This now enables us to make a numerical test of Bohr's theory. Substituting values of $e$, $m$, $c$ and $\hbar$, gives a value of $R$ in good agreement
with the value of the Rydberg constant deduced from spectroscopic data (see Section 12.1).

12.6 Radii of the Bohr Orbits

In the preceding section we obtained an expression (equation (12.7)) for the radii of the allowed orbits in the hydrogen atom:

\[ r = n^2 \frac{\hbar^2}{4\pi^2 \mu e^2}. \]

The radii of the orbits increases as \( n^2 \), so that the difference between successive orbits gets greater as \( n \) increases (Fig. 12.6).

![Fig. 12.6](image-url)

The radius of the ground-state orbit in hydrogen is found, by substituting the values of \( \hbar, m, \epsilon \) into equation (12.7), to be 0.53 Å. This gives a value of about 1 Å for the diameter of the hydrogen atom, which is of the same order as estimates based on kinetic theory calculations.

12.7 Energy Levels in Hydrogen. Ionization Potential

Putting \( n = 1 \) in equation (12.8) gives for the energy of the ground state:

\[ E_1 = -\frac{2\pi^2 \mu e^4}{\hbar^2} \]

Substituting numerical values gives

\[ E_1 = -2.17 \times 10^{-11} \text{ ergs} \]

\[ = -13.6 \text{ electron-volts}. \]

If this amount of (positive) energy is given to a hydrogen atom in its ground state it will just be sufficient to remove the electron completely from the attraction of the nucleus, i.e. to give it a total energy of zero. The atom is then ionized, so that \(-E_1\) is the ionization energy of the atom. The ionization energy for hydrogen, then, is \( 2.17 \times 10^{-11} \) ergs. This is often expressed as an ionization potential of 13.6 volts.
This value, obtained by calculation, is in good agreement with the experimentally determined value. This gives further support to Bohr's theories.

12.8 Application of Wave Mechanics

The Bohr theory of the atom outlined above gives a good account of the hydrogen atom, but when it is extended to more complex atoms discrepancies arise which cannot easily be accounted for. It should not come as a surprise to the critical student that the Bohr model has to be abandoned in favour of a more sophisticated approach, because he will already have noticed several unsatisfactory aspects of the theory. It is a strange combination of classical and quantum theory ideas, in which it has been assumed that expressions for the velocity and energy of the electron can be calculated by classical mechanics, whereas the energy and angular momentum are forced to obey certain quantum rules. Indeed, we have already seen in Section 11.8 that one cannot predict the behaviour of a bound electron in this way, and that the methods of wave mechanics must be used instead. Historically, of course, Bohr's work preceded the development of wave mechanics by some thirty years, and during this period considerable modification and development of his ideas was carried out.

If we apply the ideas developed in Section 11.8 to the hydrogen atom we shall attempt to fit in a whole number of electron waves round the Bohr orbit of the electron, thus:

momentum of electron, \( p = mv \),

\[ \therefore \text{de Broglie wavelength, } \lambda = \frac{\hbar}{p} = \frac{\hbar}{mv} \]

Length of orbit = \( 2\pi r \),

\[ \therefore \text{for exactly } n \text{ wavelengths fitted into the orbit} \]

\[ 2\pi r = n\lambda = \frac{nh}{mv} \]

\[ \therefore \text{momentum of orbit } n\frac{h}{2\pi} \]

which is Bohr's postulate of the quantization of angular momentum.

Although this may seem to prop up the Bohr theory and remove one arbitrary postulate from it, we still have not tackled the problem properly according to the rules of wave mechanics. To do this we abandon completely the idea of Bohr orbits and start afresh; and the starting-point this time must be the Schroedinger wave equation (Section 11.9), i.e.

\[ \nabla^2 \psi + \frac{8\pi^2 m(E - V)}{\hbar^2} \psi = 0, \]

where \( V \) is the potential energy of the electron and \( E \) its total energy. For the hydrogen atom the value of \( V \) is easily obtained if we assume the
validity of Coulomb's law. If the electron is at distance $r$ from the nucleus, then $V = -\frac{e^2}{r}$, so that the wave equation becomes

$$\nabla^2 \psi + \frac{8\pi^2 m(E + \frac{e^2}{r})}{\hbar^2} \psi = 0.$$  \hspace{1cm} (12.10)

In accordance with the usual rules of wave mechanics, solutions of the equation are sought in which $\psi$ and its first derivative are continuous, finite and single valued. We cannot give the solution of the equation at this stage, but must content ourselves with quoting some of the results obtained. It is found that the equation can be solved, subject to the above boundary conditions, for any positive value of $E$, but only for those negative values which satisfy the equation

$$E = -\frac{2\pi^2 me^4}{n^2 \hbar^2},$$

where $n$ is an integer. This is exactly the same quantum condition for the stationary states that we obtained from the Bohr orbital model, but it has arisen this time quite automatically once one has written down the wave equation; no arbitrary postulates have been necessary. We see, then, that equation (12.8) is still valid, and hence so is the energy level diagram shown in Fig. 12.3.

In the usual way the wave function $\psi$ may be used to determine the probability of finding the electron in a certain region. The precise orbits of the Bohr model are thus replaced by an electron probability distribution, which tells us, not where the electron is, but what the probability is that it is in a certain region. For hydrogen in the ground state it turns out that the most probable distance of the electron from the nucleus is just the same as the radius of the first Bohr orbit.

12.9 Elliptic Orbits

It was mentioned in the previous section that in the period between the publication of Bohr's papers and the replacement of his theory by a wave mechanical approach several modifications of his theory were put forward. One of the most important of these was introduced by Sommerfeld, and although we now know that his ideas have had to be replaced by the wave mechanics, it is still useful to pursue this historical approach for a while before turning to look at the predictions of the modern theory.

It is an established result of classical mechanics that the energy of a particle in an elliptic orbit (e.g. a planet) is the same as if it were moving in a circular orbit of radius equal to the semi-major axis of the ellipse. If we apply this result to the orbital electron in the hydrogen atom we arrive at the conclusion that it has an energy corresponding to one of its stationary states if it moves in any ellipse whose semi-major axis is equal to the radius of one of the circular Bohr orbits. If, however, we still in-
ATOMIC SPECTRA AND PERIODIC TABLE

sist that the rule of quantization of angular momentum is to be obeyed, then this further restricts the number of allowed elliptical orbits.

Suppose, for example, that \( n = 3 \), so that for the circular orbit the angular momentum is \( 3 \frac{\hbar}{2\pi} \). For the elliptic orbits the angular momentum is less than for the circular orbit, tending to zero as the length of the minor axis tends to zero. The quantization of angular momentum is fulfilled by those orbits which have angular momentum \( \frac{3\hbar}{2\pi} \) (circular), \( \frac{2\hbar}{2\pi}, \frac{1\hbar}{2\pi} \) (both ellipses) and \( \frac{0\hbar}{2\pi} \) (a straight line through the nucleus). Sommerfeld decided to exclude the straight line, as this would involve collision of the electron with the nucleus. The three allowed orbits are illustrated in Fig. 12.7.

To distinguish between the various orbits a second quantum number is needed. Sommerfeld used the symbol \( k \) to denote the number of units of angular momentum, so that in the above the orbits were described by

\[
\begin{align*}
  n &= 3, \ k = 3, \text{ circle}, \\
  n &= 3, \ k = 2, \text{ ellipse}, \\
  n &= 3, \ k = 1, \text{ ellipse}.
\end{align*}
\]

\( n \) is known as the total quantum number and \( k \) as the azimuthal quantum number.

We must now see whether wave mechanics predicts any results comparable to those discussed above, i.e. we must return to a discussion of the wave equation for the hydrogen atom (equation (12.10)). When the complete solution of this equation is carried out for an allowed value of \( E \) it is found that the solutions which satisfy the boundary conditions specified earlier are characterized by the occurrence of two constants which can take on only integral values. These constants are thus two further quantum numbers, and they are denoted by \( l \) and \( m \). It is found that \( l \) may have any integral value from \( n - 1 \) to zero inclusive, and that \( m \) may take on any integral value from \( +l \) to \( -l \) inclusive. It must be emphasized that these quantum numbers arise solely out of the mathematics of the solution, in fact from the boundary conditions which we have laid down for \( \psi \) and its first derivative.

The quantum number \( l \) is to be associated with angular momentum, and hence replaces \( k \) in the Sommerfeld theory. We may write \( l = k - 1 \) to give the corresponding values of the two symbols, so that in the example discussed previously, instead of \( n = 3 \) and \( k = 3, 2, 1 \), we now write \( n = 3 \) and \( l = 2, 1, 0 \). So far there seems to be no difference
between the predictions of the Sommerfeld treatment and those of wave mechanics (except for the use of a different symbol); however, when one uses the latter approach to calculate the value of the angular momentum a difference does arise. According to wave mechanics, the angular momentum is given by \( l(l + 1) \frac{\hbar}{2\pi} \). This differs from the previous result in two ways: first, and most important, it can have the value zero; and secondly, it is not an integral multiple of \( \frac{\hbar}{2\pi} \).

To denote the quantum state of an electron a notation is used which is based on terminology adopted in early spectroscopic work. Series were discovered which were named sharp, principal, diffuse and fundamental series. Using the first letters of these, the \( l = 0 \) state is denoted by \( s; l = 1 \) by \( p; l = 2 \) by \( d; \) and \( l = 3 \) by \( f \). Higher values of \( l \) are denoted by using successive letters of the alphabet; i.e. \( l = 4 \) by \( g \), etc. For example, \( n = 2, l = 1 \) is denoted by \( 2p \), and \( n = 4, l = 2 \) by \( 4d \). Table 12.1 should make this notation clear. The blank spaces indicate that these quantum states do not exist (\( l \) must be less than \( n \)).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( l )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1s</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2s</td>
<td>2p</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3s</td>
<td>3p</td>
<td>3d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4s</td>
<td>4p</td>
<td>4d</td>
<td></td>
<td>4f</td>
</tr>
</tbody>
</table>

Table 12.1

12.10 Effect of 'Elliptic Orbits' on Spectra

We have stated previously that the energy level is determined only by the value of \( n \): all states with the same \( n \) and different values of \( l \) have the same energy. However, when the effect of the change in mass of the electron with speed, as predicted by the theory of relativity, is taken into account it is found that there are slight differences in energy between states with the same \( n \) but different \( l \).

The energy level for each value of \( n \) is therefore split into \( n \) levels corresponding to the \( n \) values of \( l \). This means that there are far more possible transitions, and hence more lines in the spectrum.

The effect on the spectrum of the splitting of the energy levels may best be seen at this stage by reference to sodium. We may treat the sodium energy levels in a similar way to that already described for hydrogen, because it also produces a 'one-electron spectrum'. The electrons in the sodium atom are grouped in a number of closed shells, with one electron (the valence electron) left over. It is this electron which is responsible for the sodium spectrum.

Fig. 12.8 shows an energy level diagram for sodium, taking into account the changes in energy with \( l \). Not all transitions do in fact take place, and it is found that certain selection rules are always obeyed. These are that \( n \) may change by any amount, but \( l \) may change only by
Thus a $p \rightarrow s$ transition is allowed ($\Delta l = -1$) but not $d \rightarrow s$ ($\Delta l = -2$) or $s \rightarrow s$ ($\Delta l = 0$). Some possible transitions are shown in Fig. 12.8, together with the names of the series of which the emitted spectral lines are part. Incidentally, this also shows why the $s$, $p$, $d$, $f$ notation was adopted. The sharp series involves transitions from an $s$-state, the principal series from a $p$-state and so on.

12.11 The Zeeman Effect. Magnetic Quantum Number

It is found that when radiation is emitted by atoms subjected to a magnetic field each line of the normal spectrum is split into several closely spaced lines. This is known as the Zeeman effect. It seems to indicate that the presence of the magnetic field has the effect of splitting the energy levels of the atom.

Using the Bohr–Sommerfeld picture of the atom, this splitting can be explained in terms of the interaction of the magnetic moment of the orbital electrons with the applied magnetic field. The orbital electrons constitute an electric current flowing in a loop, and hence produce a magnetic moment. Suppose that the direction of this magnetic moment is not along the applied magnetic field when the latter is switched on. The atom then tries to turn so that its magnetic moment lies along the direction of the field, but is prevented from doing so by its angular momentum. The effect is similar to that of a gyroscope or spinning top which when it tries to tip over under the action of gravity does not do so, but instead its axis rotates or precesses. In the case of the atom the energy associated with the precession changes its energy level.

The observed fact that a given spectral line is split into a finite number of lines in a magnetic field shows that the energy levels of the atom are also split into a finite number of new levels. This can be explained in terms of a space quantization rule. According to this rule, it is assumed that the magnetic moment of the atom can be inclined only at
certain allowed angles to the direction of the field, so that only a limited number of precession energies are possible.

It may be shown by classical electromagnetic theory that the magnetic moment of the orbital electrons is proportional to their angular momentum and is in the same direction as the angular momentum vector (i.e. perpendicular to the plane of the orbit for a single electron). It is therefore convenient to express the space quantization rule in terms of the angular momentum vector. To give agreement with the observed Zeeman effect the rule is stated to be that the angular momentum vector may be only at those angles which result in its component along the field being an integral number of $\frac{h}{2\pi}$. This leads to the introduction of the magnetic quantum number, $m$, which indicates the component 

\[
\left(\frac{m}{\frac{h}{2\pi}}\right)
\]

of the angular momentum in the special direction.

![Fig. 12.9](image)

The wave mechanical treatment leads also to the introduction of a magnetic quantum number, which is in fact the quantum number $m$ mentioned in Section 12.9. Even though the ideas of space quantization have been superseded by the ideas of wave mechanics (in which the quantum number $m$ appears without the introduction of any ad hoc assumptions), a vector treatment of angular momentum is still often found to be a useful pictorial way of representing the situation. Fig. 12.9 illustrates this model for the case $l = 2$. The arrowed lines indicate the angular momentum vector, which in this case is \(\sqrt{6} \frac{h}{2\pi}\) (i.e.\([l(l + 1)]\frac{h}{2\pi}\)). These vectors are drawn in directions such that their components on the $z$-axis are $m\frac{h}{2\pi}$, where $m = 2, 1, 0, -1, -2$.

In general $m$ may take the $2l + 1$ values from $+l$ to $-l$ inclusive. We now have to quote the three quantum numbers $n, l, m$ in order to describe the state of an electron in an atom. But this is still not the whole story; we shall see in the next section that yet another quantum number has still to make its appearance.

### 12.12 The Spinning Electron

It was stated in Section 12.10 that the introduction of the quantum numbers $n$ and $l$ was able to explain the occurrence of the sharp, principal, diffuse and fundamental series in the sodium spectrum. In fact, each of the 'lines' in the series is not one line, but consists of two or three very closely spaced lines. In the principal series the 'lines' are double, the best-known example being that of the sodium D-lines of wavelengths 5,890 and 5,896 Å. The separation in this case is large
enough for the two separate lines to be seen with a grating spectrometer of moderate resolving power.

Analysis shows that this fine structure can be explained if it is assumed that all the energy levels except the $s$ level are actually two levels close together. The existence of these doublets cannot be explained without making some addition to the theory so far developed. The assumption made by Uhlenbeck and Goudsmit in 1925 was that, as an electron is of finite size, then it is capable of rotation, and can thus possess angular momentum and magnetic moment. This is referred to as electron spin. It was shown that the observations can be accounted for if the spin angular momentum is taken to be $\frac{\hbar}{2\pi}$.

The magnetic moment of the spinning electron was assumed to interact with that of the orbits and produce a change in the energy level. To account for the splitting into two levels it was assumed that the spin angular momentum could be either parallel or anti-parallel to the orbital angular momentum. The *spin quantum number* is denoted by $s$, and unlike the other quantum numbers this may take on only the one value, $\frac{1}{2}$. The two settings of the spin may be denoted by $+\frac{1}{2}$ and $-\frac{1}{2}$.

We must as usual enquire whether wave mechanics is able to predict this quantum number in a way which does not involve arbitrary assumptions of the type made in the electron spin theory outlined above. The Schroedinger equation so far discussed in connection with the hydrogen atom yields only the three quantum numbers $n$, $l$ and $m$ in its complete solution. However, it has been shown by Dirac that if one develops the wave equation to take into account the results of relativity theory (the relativistic wave equation), then the solutions of this equation do in fact involve the fourth quantum number $s$. The associated angular momentum is shown to be $[s(s + 1)]\frac{\hbar}{2\pi}$.

**12.13 Total Angular Momentum**

The total angular momentum of an electron is the vector sum of its orbital and spin moments. This is represented by the quantum number $j$, which may have the two values $l + \frac{1}{2}$ and $l - \frac{1}{2}$. For the $s$ terms ($l = 0$) can have only the value $\frac{1}{2}$, so that these are single.

The nomenclature used, is to show the value of $j$ as a subscript to the notation used previously.

*\text{e.g.} \quad 2p_3 \text{ means } n = 2, \ l = 1, \ j = \frac{3}{2}; \quad 4d_3 \text{ means } n = 4, \ l = 2, \ j = \frac{3}{2}.\)

The idea of space quantization may now be applied to the total angular momentum, and the magnetic quantum number (denoted by $m_j$ to distinguish it from the magnetic quantum number for $l$ which we shall in future write $m_l$) may have the $2j + 1$ values from $-j$ to $+j$ inclusive.
The energy-level diagram now becomes more complex, because it must show the effect of electron spin. Fig. 12.10 shows part of such a diagram for sodium. This should be compared with Fig. 12.8. The sodium D-lines are the transitions $2p_3 \rightarrow 1s_1$ and $2p_1 \rightarrow 1s_1$.

![Diagram](image)

Fig. 12.10

Certain selection rules for transitions apply. These are:

- $n$ may change by any amount;
- $l$ may change by $\pm 1$;
- $j$ may change by 0, $\pm 1$.

### 12.14 The Pauli Exclusion Principle and the Periodic Table

We are now in a position to understand the structure of the periodic table in terms of shells and sub-shells. This is based on the **Pauli exclusion principle**, which states that *no two electrons in an atom may be in exactly the same quantum state*. We may specify the state of an electron by the four quantum numbers $n$, $l$, $j$, $m_j$.

- $n$ may have any value;
- $l$ may have the values $n - 1$, $n - 2$, \ldots 0;
- $j$ may have the two values $l + \frac{1}{2}$ and $l - \frac{1}{2}$;
- $m_j$ may have the $2j + 1$ values from $-j$ to $+j$.

We first calculate how many electrons may have the same value of $l$. If, for a given value of $l$, $j$ has the value $l + \frac{1}{2}$, then $m_j$ may have $(2j + 1)$ values, i.e. $2(l + \frac{1}{2}) + 1 = 2l + 2$ values. If $j = l - \frac{1}{2}$, then $m_j$ may have $2(l - \frac{1}{2}) + 1 = 2l$ values. The total number of combinations of $j$, $m_j$ for a given $l$ is thus $2l + 2 + 2l = 2(2l + 1)$. If $l = 0$, $j$ may not have the value $l - \frac{1}{2}$, but the above formula still holds because we counted $2l$ values of $m_j$ for $j = l - \frac{1}{2}$, and this is zero if $l = 0$. For any value of $l$, then, there are allowed $2(2l + 1)$ electrons with that value of $l$.

With these facts before us we may now proceed to investigate the
Table 12.2 Periodic table of the elements
structure of the periodic table (Table 12.2). The elements are arranged in ascending order of their atomic numbers, \( Z \). \( Z \) denotes the number of electrons in the atom.

The simplest element, hydrogen, has \( Z = 1 \). In the ground state its one electron will be in the orbit of lowest energy, i.e. \( n = 1 \). Electrons with \( n = 1 \) are said to be in the \( K \) shell. \( l \) may have only the one value \( l = 0 \), because \( l \) must be less than \( n \). \( j \) has the value \( \frac{1}{2} \) and \( m_j \) may have either of the values \( +\frac{1}{2}, -\frac{1}{2} \). There is thus room for one more electron having \( n = 1 \). Helium, \( Z = 2 \), thus has both its electrons in the \( K \) shell, which is then complete; any further electrons must have \( n = 2 \) (\( L \) shell). An element whose electrons fill a complete number of shells is particularly stable chemically; it is the electrons in incomplete shells which are responsible for chemical activity.

Lithium, \( Z = 3 \), has two electrons in the \( K \) shell and one in the \( L \) shell. This one electron outside a completed shell is a valence electron and is typical of the alkali metals. For the \( L \) shell \((n = 2)\), \( l \) may have the values \( 1, 0 \); electrons with \( l = 0 \) are said to be in the \( s \) sub-shell, and with \( l = 1 \) in the \( p \) sub-shell. Application of the expression \( 2(2l + 1) \) shows that there may be 2 electrons in the \( s \) sub-shell and 6 in the \( p \) sub-shell, making a total of 8 electrons in the \( L \) shell. The \( L \) shell is thus complete for the element \( Z = 10 \) (neon).

Sodium, \( Z = 11 \), has complete \( K \) and \( L \) shells and one electron in the \( M \) shell. The \( M \) shell may have \( s, p, d \) sub-shells containing respectively 2, 6, 10 electrons. Argon, \( Z = 18 \), has the 3\( s \) and 3\( p \) sub-shells complete; it is an inert gas, although the \( M \) shell is not complete. However, the next electron \((Z = 19, \text{potassium})\) goes, not into the \( 3d \) state but into \( 4s \), which is in the \( N \) shell. This occurs because it requires less energy to put the electron into \( 4s \) than to put it into \( 3d \).

The way in which the build-up of electrons proceeds after this may be seen from a careful study of Table 12.2. It may also be seen in tabular form in many of the books mentioned in the bibliography (e.g. Born).

<table>
<thead>
<tr>
<th>( n )</th>
<th>Shell</th>
<th>( l )</th>
<th>Sub-shell</th>
<th>Number of electrons in sub-shell</th>
<th>Number of electrons in shell</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( K )</td>
<td>0</td>
<td>( s )</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>( L )</td>
<td>0</td>
<td>( s )</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1</td>
<td>( p )</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>( M )</td>
<td>-1</td>
<td>( p )</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2</td>
<td>( d )</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-3</td>
<td>( d )</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>( N )</td>
<td>-1</td>
<td>( p )</td>
<td>10</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2</td>
<td>( d )</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-3</td>
<td>( f )</td>
<td>14</td>
<td></td>
</tr>
</tbody>
</table>

Table 12.3
ATOMIC SPECTRA AND PERIODIC TABLE  235

Table 12.3 shows how many electrons are allowed in each shell and sub-shell. It will be noticed that there are $2n^2$ electrons in each shell.

QUESTIONS

1. Using the values $m = 9.11 \times 10^{-28}$ g, $e = 4.80 \times 10^{-10}$ e.s.u., $c = 3.00 \times 10^{10}$ cm sec$^{-1}$, $h = 6.63 \times 10^{-27}$ erg sec, calculate the value of $R$ from equation (12.9), and compare this with the value quoted in Section 12.1.

2. Check the values for $r$ and $E$, given in Section 12.7.

3. Show that the total number of electrons allowed in a shell is $2n^2$. 
13

ELECTRICAL AND MAGNETIC PROPERTIES OF MATTER

ELECTRICAL PROPERTIES

13.1 Classification

Solids may be classified into three groups according to their electrical properties:

(i) Conductors. These have appreciable electrical conductivity, which decreases as the temperature rises. This group comprises all metals and their alloys.

(ii) Semi-conductors. These have a small, but measurable, conductivity at room temperature. The conductivity of semi-conductors increases with rise in temperature.

(iii) Insulators (dielectrics). The conductivity is virtually zero under normal conditions.

We shall study some of the properties of each of these groups in the following sections.

13.2 Drude’s Theory of Metallic Conduction

Conduction in a metal is by ‘free’ electrons. These electrons are the outermost (valence) electrons of the atoms of the metal. The property which makes a substance a metal is that these electrons are not tightly bound to individual atoms, but are able to wander from atom to atom, even in the absence of an electric field.

Under the influence of an electric field these electrons drift through the metal and constitute an electric current. The number of electrons in the metal does not change because each electron that leaves one end of the metal is replaced by one entering at the other end.

If this picture is to be used to give a quantitative account of conduction we shall need to know the laws that govern the motion of the electrons in the metal. Drude assumed that the electrons are free to move within the confines of the metal, like the molecules of a perfect gas in a box; they are said to form a free electron gas. The electrical conductivity of such a gas would be infinite because there would be no hindrance to the flow of electrons under an applied electric field. To explain the finite conductivity of a metal it is necessary to postulate that the electrons have their motions impeded and their velocities reduced by some sort of scattering process in the metal. An obvious suggestion is that this occurs due to collisions of the electrons with the ions fixed in the lattice, and this was assumed in the original theory. This has subsequently been shown to be a false picture, and the scattering is now
attributed to the thermal vibrations of the ions and to imperfections or impurities in the lattice structure. Thus at absolute zero (where there are no thermal vibrations of the ions) the electrical conductivity of a perfect metallic crystal should be infinite; the residual resistance found in practice is due to imperfections and depends on the way in which the specimen has been formed.

We can develop the theory as far as is necessary for our purpose without making assumptions about the nature of the scattering process. All we need to postulate is that the electrons do make ‘collisions’ and that their drift velocities are thereby reduced to zero.

Consider first the situation within the metal in the absence of an electric field. The free electrons are in random motion due to their thermal energies, and they make frequent ‘collisions’ with fixed irregularities. Suppose that the average time between ‘collisions’ is \( t \) and that the mean velocity of the electrons is \( \bar{v} \).

In the calculation which follows no mention is made of units, as the formulae are applicable in both the c.g.s. and m.k.s. systems. Consider a specimen of metal in the form of a rod of uniform cross-section of area \( A \) and length \( l \) and assume a potential difference \( V \) is maintained across its ends. This gives rise to an electric field \( E (= V/l) \) in the metal in a direction parallel to the axis of the rod. Each free electron is acted on by a force, \( F \), given by

\[
F = eE,
\]

and hence undergoes an acceleration,

\[
a = \frac{eE}{m},
\]

where \( e \) denotes the charge on an electron and \( m \) its mass.

After a time \( t \) (on average) the electron will be scattered and its drift velocity reduced to zero. It will have acquired velocity equal to \( at \), i.e.

\[
\frac{eEt}{m},
\]

in this time, so that the average drift velocity, \( v \), is given by

\[
v = \frac{eEt}{m}.
\]

Note that we have assumed that \( v \) is much less than \( \bar{v} \), so that the time between ‘collisions’ is not appreciably altered by the electric field.

Now, if there are \( n \) free electrons per unit volume in the metal, the number passing through any cross-section of the rod in 1 second is the number contained in length \( v \) of the rod, i.e. \( nAv \). But the electric current, \( I \), is the rate of flow of charge, and hence

\[
I = nAv
\]

\[
= \frac{nAv^2Et}{2m}.
\]

on substituting for \( v \).
But
\[ E = \frac{V}{l} \]
\[ : : I = \frac{nAe^2t}{2ml} V. \]
Comparing this with Ohm’s law, gives for the resistance
\[ R = \frac{2ml}{nAe^2t} \]
\[ = \frac{l}{\left(\frac{ne^2t}{2m}\right)} A, \]
and hence the conductivity
\[ \sigma = \frac{ne^2t}{2m}. \]
The electrons have a \textit{mean free path}, \( \lambda \), between ‘collisions’ given by
\[ = \bar{\ell}t, \] so that we may write
\[ \sigma = \frac{ne^2\lambda}{2mc} \quad . \quad . \quad . \quad \quad (13.2) \]
The electron gas will also have a thermal conductivity, \( K \), which may be obtained in terms of the properties of the ‘gas’ by the methods of kinetic theory (Section 6.10). Equation (6.7) gives
\[ K \propto n\lambda \bar{c} \frac{dE}{dT}. \]
The electron gas has three degrees of freedom, so that at temperature \( T \) the average kinetic energy of an electron is \( \frac{3}{2}kT \) (Section 6.5).
\[ : : \frac{dE}{dT} = \frac{3}{2}k, \]
so that
\[ K \propto n\lambda \bar{c} \quad . \quad . \quad . \quad \quad (13.3) \]
Dividing equation (13.3) by equation (13.2) gives
\[ \frac{K}{\sigma} \propto \frac{mc^2k}{\bar{c}^2}. \]
But the average kinetic energy of an electron, \( \frac{1}{2}mc^2 = \frac{3}{2}kT \), where \( \bar{c}^2 \) is the mean square velocity of the electrons,
\[ : : \frac{K}{\sigma} \propto \frac{k^2}{\bar{c}^2} T \text{ because } \bar{c}^2 \propto \bar{c}^2, \]
\[ : : \frac{K}{\sigma T} = \text{constant.} \]
ELECTRICAL AND MAGNETIC PROPERTIES

This is the same as the empirical Wiedemann–Franz law mentioned in Section 5.2, where \( K \) is the thermal conductivity of the metal. In the above theory \( K \) is the thermal conductivity of the electron gas. The agreement between theory and experiment shows that the high thermal conductivity of metals may be attributed to the free electrons, the same particles that are responsible for the electrical conductivity.

13.3 The Hall Effect

If a conductor in the form of a rod or slab is carrying a current in one direction and has a magnetic field applied perpendicular to the direction of the current, then a potential difference is developed across the specimen in the direction perpendicular to both current and magnetic field (Fig. 13.1). This is known as the Hall effect; it may be explained in terms of the simple free electron model. The calculation which follows applies either in c.g.s. (e.m.u.) or m.k.s. units.

Consider a current, \( I \), produced by the flow of particles, each carrying an electric charge, \( +e \), with drift velocity \( v \). Suppose that a magnetic field of flux density \( B \) is applied perpendicular to the direction of motion of these particles. A force, \( F \), of magnitude \( Bev \) acts on each particle and causes it to move towards one side of the specimen (Fig. 13.1). This movement continues until an electric field, \( E \), has been set up in the direction shown, such that the force it exerts on the charges just balances that due to the magnetic field,

\[
e E = Bev
\]

But

\[
I = nevA \quad \text{(see equation (13.1))},
\]

where \( n \) is the number of charged particles per unit volume.

\[
E = \left( \frac{IB}{A} \right) \times \frac{1}{ne}.
\]

The quantity \( 1/ne \) is called the Hall coefficient. One important fact about the Hall coefficient is that its sign indicates the sign of \( e \), and thus tells us the nature of the carriers responsible for conduction. It also enables the density \( n \) of the carriers to be determined.
Most metals have a negative value of the Hall coefficient, indicating that the current is carried by electrons, but there are exceptions.

13.4 Application of the Quantum Theory. Fermi Energy

We have discussed examples where the free electron model has led to results in fair agreement with experimental measurements; in other cases, however, this simple picture is not so successful in predicting the properties of metals. For example, if we use the ideas developed in previous sections to calculate the specific heat of a metal we arrive at a completely erroneous conclusion.

It was shown, in Section 6.13, that the atomic heat of a solid element should be $3R$, corresponding to the six degrees of freedom of the atoms. In a pure metal the free electrons contribute another three degrees of freedom, so that (assuming one free electron per atom) the atomic heat should be increased by a further $\frac{3}{2}R$. In other words, the atomic heat of a solid metallic element should be half as high again as that of any other solid element. Measurement shows that this is not so; there is little difference between the atomic heats of the two groups at ordinary temperatures. In order to resolve this and other difficulties, we must abandon the classical free-electron model and rethink the problem in terms of quantum theory. In the kinetic theory approach which we have previously used we made the implicit assumption that the free electrons may possess any value of energy; we know, however, that this is not a valid assumption for particles confined to a 'box', and that according to quantum theory the electrons have only certain allowed energy levels. We were tempted to ignore this effect by analogy with gas kinetic theory, which makes excellent predictions about the behaviour of gases without reference to the quantum theory at all. This is because the energy levels for an actual gas are so close together as to produce, at normal temperatures, an effective continuum.

When we deal with an 'electron gas' two factors arise which make classical assumptions completely invalid. First, the mass of an electron is very much less than that of a molecule, and this results in a wider spacing of the energy levels. Secondly, and far more important, the electrons must obey the Pauli exclusion principle (Section 12.14), which the molecules of a gas do not. According to this principle no two electrons may be in exactly the same quantum state; this means that only two electrons may have the same kinetic energy. Two electrons are allowed in each energy level because they may be distinguished by opposite spins.

The problem is, then, to find the allowed energy levels of the free electrons in the metal and put at most two electrons into each level. The number of electrons involved ($\sim 10^{22}$ per cc) means that a very large number of energy levels are filled. The method of approach is similar to that outlined in Section 11.8, extended to three dimensions. We are still using the free electron model and regarding each of these electrons to be
confined to the volume of the metal. The electrons are so confined, because if an electron leaves the surface of the metal the latter becomes positively charged by induction, and there is a force pulling the electron back into the metal.

If electrons at rest outside the metal (at infinity) are regarded as having zero energy, then the electrons in the metal have negative electrical potential energy. This is depicted by representing the metal as a ‘potential well’ (Fig. 13.2). The potential energy of an electron in the metal is indicated by the line PP'.

![Diagram showing potential energy of electrons in a metal](image)

**Fig. 13.2**

It is not proposed to attempt here a solution of the problem of calculating the allowed energy levels of the free electrons in the metal (although it is not too difficult, and the student should be able to follow it in the books mentioned in the bibliography), but simply to state some of the results obtained.

When the allowed energy levels have been determined the next step is to decide which of these will actually be occupied by electrons. At absolute zero this is easy; the electrons have the lowest possible kinetic energies, so we simply fill the levels from the lowest upwards, putting two electrons in each level. It turns out that the highest level reached when all the electrons are used up corresponds to a kinetic energy which would be attained, according to the classical picture, at temperatures in excess of 10,000°K. In face of this, it is not surprising that the classical and quantum theory predictions are somewhat different!

The kinetic energy of the electrons in the highest level filled at absolute zero is referred to as the *Fermi energy*, \( W \), of the electrons, and the level itself as the *Fermi level*. Typically, the Fermi energy is of the order of a few electron-volts.

When the temperature of the metal is raised some of the electrons gain kinetic energy and move into higher energy levels. The electrons affected can come only from levels close to the Fermi level, because electrons farther down need to gain large amounts of energy if they are to move up into vacant levels, all levels immediately above being already occupied. Fig. 13.3 illustrates this point. An electron such as \( A \) would need to acquire energy at least equal to that shown by the arrow. Except at very high temperatures there is not sufficient energy available and the transition does not occur. Electrons like \( B \) can gain sufficient
energy to move to a vacant level. If the temperature is high some electrons gain enough energy to 'overflow' from the potential well; this constitutes thermionic emission.

Fig. 13.3

Fig. 13.4 shows how the energy is distributed amongst the electrons at absolute zero and at a high temperature. The shaded areas represent the small number of electrons whose energies are increased at the high temperature.

Fig. 13.4

To return now to the problem of electronic specific heat mentioned at the beginning of this section. We saw that the classical theory was unable to explain the negligible value of this compared with the specific heat of the ionic lattice. The quantum theory offers a simple explanation. We have seen that only a few electrons near the top of the Fermi distribution can gain energy when the metal is heated; these are the only electrons that contribute to the electronic specific heat, which is accordingly small. Detailed calculation shows good agreement between theory and experiment.

Similarly, only the electrons near the Fermi level can take part in electrical conduction; those that take part in conduction gain kinetic energy and move into higher energy levels. We shall see later that this is very important in the explanation of the difference between conductors, semi-conductors and insulators.
13.5 Electronic Work Function and Contact Potential

To remove an electron from a metal to infinity requires a certain amount of energy to be supplied to the electron. It is seen from Fig. 13.3 that at absolute zero this energy must be at least equal to the difference between the Fermi level and the top of the potential well. This energy is represented by $\phi$ in Fig. 13.3 and is known as the electronic work function; it varies slightly with temperature. Its value is of the order of a few electron-volts.

Two metals have different Fermi energies and different electronic work functions. When the two metals are placed in contact some electrons from the top levels of the one with the smaller work function flow into the other metal, where they have lower energy (from A to B in Fig. 13.5 (a)). This flow causes a separation of charge across the boundary, so that a potential difference is set up which opposes the transfer of electrons from A to B. In equilibrium this potential difference prevents a further net flow of electrons; this means that the ‘potential well’ of B has been raised until the Fermi levels in the two metals are the same (Fig. 13.5 (b)). The potential difference between the metals is known as a contact potential, and may be of the order of 1 volt.

The number of electrons transferred before equilibrium is reached is a very small fraction of the whole, so that the Fermi distribution in each metal is not appreciably affected. It may be seen from Fig. 13.5 (b) that the contact potential of the two metals is equal to the difference between their electronic work functions.

13.6 Energy Bands

We must now abandon the free electron picture and cease to talk about an electron gas; instead we must try to apply the quantum theory of the atom, developed in Chapter 12, to the solid state. We have seen that for a single atom there are certain allowed energy levels which its electrons may occupy; we now enquire how these energy levels are modified when the atoms are brought close together in a solid. Consider sodium as an example. The sodium atom in its ground state has its 1s, 2s and 2p levels complete and one electron in the 3s level. This latter
level could hold two electrons, and is thus only half-full; it is this which makes sodium a metal, as we shall see.

Imagine now that two sodium atoms are brought close together. This causes a slight change in the 3s level in each atom and results in the single level splitting into two closely spaced levels. As more atoms are brought close further splitting occurs, so that for $2N$ sodium atoms the 3s level has become an energy band consisting of 2$N$ closely spaced levels. At absolute zero only $N$ energy levels in this band are occupied, because each level can accommodate two electrons (of opposite spins), i.e. $4N$ electrons in all, whereas there are only $2N$ valence electrons available. Another way of expressing this is to say that the Fermi level comes half-way up the 3s band. Sodium is therefore a conductor, because there are vacant levels immediately above the Fermi level into which electrons can be raised under the influence of an electric field (Fig. 13.6).

It would seem that, according to this theory, the alkaline-earth metals should not be metals. These elements have two valence electrons in an s sub-shell, which is therefore full; e.g. the valence electrons of calcium fill the 4s level. For a solid consisting of $N$ calcium atoms there are $2N$ valence electrons which just fill the 4s band, i.e. the Fermi level is just at the top of this band. How is it, then, that calcium is a conductor of electricity? The answer to this question comes when we consider the next higher energy band, i.e. the 4p band.

The 4s and 4p levels in an atom are quite distinct, but when they be-
ELECTRICAL AND MAGNETIC PROPERTIES

come broadened into bands in the solid state overlapping between the two bands occurs. Fig. 13.7 shows how the bands broaden as the atoms are brought closer together. This clearly illustrates that, although the Fermi level comes at the top of the $4s$ band, it is in the $4p$ band and there are vacant levels for conduction to occur.

This band theory provides a satisfactory explanation of the differences between conductors, semi-conductors and insulators. We have already seen that for a conductor the Fermi level is in an allowed energy band; for a semi-conductor or an insulator the Fermi level comes right at the top of such a band, there being a forbidden zone immediately above. This means that at absolute zero the valence electrons cannot gain the small amounts of kinetic energy which they would have to gain if they took part in electrical conduction. All such substances are therefore insulators at absolute zero (semi-conductors are insulators at this temperature).

![Diagram showing energy bands for conductor, semi-conductor, and insulator](image)

When the temperature is raised there is a possibility that some of the electrons may be able to gain sufficient thermal energy to jump to the next allowed energy band. Whether or not this happens depends on the width of the forbidden zone. If this is narrow some electrons do make transitions into the higher band, and these electrons can then take part in conduction; the substance is an intrinsic semi-conductor. Carbon, silicon and germanium are examples of this class. If the energy gap is too wide the element is an insulator. Fig. 13.8 illustrates these properties.

We now proceed to study the properties of semi-conductors in some detail.

13.7 Intrinsic Semi-conductors

We have seen that the nature of an intrinsic semi-conductor is that some of the electrons from the top of the Fermi distribution are raised, at ordinary temperatures, into what we may call the conduction band. The higher the temperature, the more electrons there are in the conduction band, so that the conductivity increases with temperature. Detailed calculation shows this to be an exponential increase, in accord
with experimental measurements. At room temperature the number of conduction electrons is still small and semi-conductors have a conductivity about $10^{-7}$ to $10^{-8}$ times that of metals.

The electrons which are raised into the conduction band leave gaps in the band from which they have been removed. Some of the other electrons, of lower energy, in that band may now gain kinetic energy and move into these gaps, i.e. electrical conduction is possible. It is convenient to think of these gaps in an otherwise completely filled band as positively charged holes. A hole, then, is the lack of an electron, and it may move through the semi-conductor in the opposite direction to the electron flow, so contributing to the flow of charge which constitutes the electric current.

The way in which this hole movement takes place may be visualized in this way: suppose an atom of a semi-conductor has one of its valence electrons raised into the conduction band so that this electron, under the influence of an applied electric field, moves away from the atom.

![Diagram of electron and hole movement](image)

Fig. 13.9

This atom is now one electron short, so that it has a single positive charge (atom A in Fig. 13.9). An electron from a neighbouring atom (e.g. B) can now move to A and make it neutral again; but B is now positively charged, so that an electron from C may move to it and so on. It will be noticed in Fig. 13.9 that the effect is that of the movement of a positive charge (hole) to the right. The conduction electron released from A (— in Fig. 13.9), is moving to the left; both of these effects contribute to the total current flow. There are an equal number of holes and electrons in this case, but they may travel with different speeds, and hence make different contributions to the total current. For both germanium and silicon the electron mobility (drift velocity in unit field) is greater than the hole mobility, so that the former makes the larger contribution to the conduction process, resulting in a negative Hall coefficient.

13.8 Impurity Semi-conductors

By inserting suitable impurity atoms into the crystal lattice of germanium or silicon two effects may be achieved: the number of carriers
ELECTRICAL AND MAGNETIC PROPERTIES

(holes or electrons) may be increased, and one type may be produced in excess of the other. To see how this arises we must consider the arrangement of the atoms in a perfect crystal of germanium (all that follows applies in exactly the same way to silicon). Germanium has a valency of four and crystallizes such that each atom has four nearest neighbours with which it forms covalent bonds; that is, each atom shares four of its electrons with its nearest neighbours (Fig. 13.10 (a)).

If an atom of a group V element (e.g. phosphorus) of valency five is introduced into the crystal lattice, in place of one of the germanium atoms, one of its valence electrons is spare (Fig. 13.10 (b)). The energy level of this electron is just below the conduction band in germanium (see Fig. 13.11), so that at normal temperatures this electron may be raised into this band and become a conduction electron. If a number of impurity atoms are present there are enough electrons in the conduction band to cause appreciable conductivity – much greater than that of an intrinsic semi-conductor. A semi-conductor which has conduction electrons due to the presence of the impurity atoms is referred to as n-type (negative carriers).

The intrinsic semi-conductor action still takes place so that there are also a relatively small number of holes; these may also be present due to the presence of other impurities. The presence of these minority carriers is important in the action of the junction diode and the transistor.

If a group III element (e.g. aluminium or indium) of valency three is introduced into the crystal lattice of germanium there is a deficiency of a valence electron near the impurity atom (Fig. 13.10 (c)). This vacancy has an energy level just above the filled band in germanium (Fig.

![Diagram](image-url)
13.11, so that an electron from this band may be raised to this level at normal temperatures and leave a positive hole. This is a $p$-type (positive carriers) semi-conductor, in which the conduction is by a hole current. There are also a minority of negative carriers (electrons) present.

Fig. 13.11

13.9 The Thermistor

The large negative temperature coefficient of resistance of a semi-conductor is made use of in a device known as a thermistor. This consists of a small bead or disc of the semi-conductor.

A thermistor may be used as a resistance thermometer; its large temperature coefficient makes the measurement of the change in resistance with temperature easier than for a metal.

Another use of the thermistor is to limit surge currents in a circuit, by virtue of its negative temperature coefficient. For example, the heaters of valves have a low resistance when cold so that large initial currents flow at switch on. If a thermistor is put in series its resistance will be high when the current is first switched on, thereby reducing the surge. As the thermistor heats up its resistance falls rapidly and allows the current to reach its correct steady value.

13.10 The Junction Diode

This is made by forming a junction between a $p$-type and an $n$-type semi-conductor, within a single crystal. This may be accomplished by diffusing a pellet of the $p$-type impurity (e.g. indium) into one side of a crystal of the $n$-type semi-conductor. The diode has the property that it conducts when the $p$-region is positive with respect to the $n$-region, but allows very little current to flow in the reverse direction. The diode is represented schematically in Fig. 13.12 (a). When no external potential difference is applied (no bias) some of the holes from the $p$-type diffuse into the $n$-type region; electrons diffuse in the opposite direction. The $p$-type gains excess negative charge and the $n$-type excess positive charge, which results in a potential difference across the boundary, as indicated in Fig. 13.12 (a). There are also minority carriers present in both regions, and these flow across the boundary, aided by the potential difference. In equilibrium there are a few holes in the $p$-region and a few electrons in the $n$-region which have sufficient energy to overcome
the potential barrier and cross into the other region. This flow is cancelled by the flow of minority carriers in the opposite direction, so that the net current is zero.

Suppose that an external potential difference is connected as in Fig. 13.12 (b), with its positive side to the $n$-region and its negative to the $p$-region (reverse bias). This increases the potential difference across the boundary, so that the flow of majority carriers is almost stopped. The flow of minority carriers is still as before, so that a very small, almost constant, current flows from $n$ to $p$. For a small diode this current is of the order of microamps.

If the terminals of the external supply are reversed (forward bias) the potential barrier is reduced and the flow of majority carriers increased. The holes entering the $n$-region soon recombine with free electrons there; electrons diffusing into the $p$-region recombine with holes. A current thus flows from $p$ to $n$; this current depends on the magnitude of the applied potential difference. For small diodes the current is of the order of milliamps. Large diodes are now made which will handle currents of the order of tens of amps.

Fig. 13.13 is a graph showing the characteristic of a typical junction diode. $V$ is shown as positive when the junction is forward biased.
13.11 Metal – Semi-conductor Diodes

(a) The point-contact diode

The earliest diode was of this type, although its action was not then understood. This was the 'cat's whisker' used as a detector in early radio receivers. It consisted of a metal wire in contact with a crystal of galena (lead sulphide), which is a semi-conductor.

A modern point contact diode consists of a fine platinum wire in contact with an n-type semi-conductor; it has a low resistance for current flowing from the wire into the crystal and a very high resistance in the reverse direction. It is chiefly used in high-frequency circuits, where its very low capacitance is essential.

(b) The metal rectifier

It has been found that certain other metal – semi-conductor junctions have good rectifying properties; selenium and copper oxide are two of the most commonly used semi-conductors for this purpose. In this case the semi-conductor is formed as a layer on a metal plate, and much larger currents can be handled than is possible with the point-contact type. Metal rectifiers are frequently used in power packs which produce direct current from the A.C. mains.

13.12 The Junction Transistor

This consists of a semi-conductor crystal with two junctions; the two possible types are illustrated in Fig. 13.14 (a), (b). They are known respectively as \( p-n-p \) and \( n-p-n \) transistors. The three regions in the transistor are known as the emitter, base and collector. Fig. 13.14 (a), (b) also shows the symbols used to represent the two types of transistor in circuit diagrams.

![Diagram of p-n-p and n-p-n transistors](image)

Fig. 13.14

We shall consider the operation of the \( p-n-p \) type. This is made by diffusing an indium pellet into each face of a thin slice of \( n \)-type germanium; Fig. 13.14 (c) shows the construction. It is arranged that the base has a low concentration of electrons compared with the concentration of holes in the emitter and collector regions.
In use the emitter–base junction is given a small forward bias (less than 0.3 volt), and the collector–base junction a larger reverse bias (e.g. 5 volts), as indicated in Fig. 13.15. Holes from the emitter diffuse into the base, and most of these continue into the collector, causing a current in the collector circuit. A few of the holes combine with electrons in the base, causing a very small base current to flow, as shown.

A small change in the base current ($\delta i_b$) causes a much larger change in the collector current ($\delta i_c$), so that the transistor may be regarded as a current amplifier. The current amplification factor ($\alpha'$ or $\beta$) is defined as $\delta i_c / \delta i_b$.

Fig. 13.16 shows a circuit, suitable for investigating the characteristics of a $p-n-p$ transistor. Figs. 13.17 and 13.18 show simple circuits for a transistor amplifier.

13.13 Applications of Semi-conductor Devices

Semi-conductor devices have several advantages over thermionic valves:

(i) they work at much lower voltages – tens of volts instead of hundreds of volts;
(ii) there are no heaters, so that the heat dissipation is only a small fraction of that in valve circuits;
(iii) they are small and mechanically robust.

These advantages have led to a large-scale use of semi-conductor diodes and transistors in military and space-research equipment, in large computers and in many other spheres of industrial electronics. The reader will already be familiar with their use in portable and car radios and in deaf aids.

13.14 Dielectrics

In this discussion we use the term electric dipole to denote a positive and negative charge of equal magnitude a small distance apart. This dipole, when held at right angles to an electric field, experiences a couple tending to turn it into the direction of the field. The moment of this couple in a field of unit strength is called the electric dipole moment.

The arrangement of the electrons in an atom is symmetrical about the nucleus, so that the atom has no dipole moment. In a molecule the arrangement may be symmetrical, or it may be such that the molecule has a permanent dipole moment; molecules of the latter type are said to be polar. The molecules of potassium and hydrogen halides are polar, and so are those of water; molecules of hydrogen, nitrogen and carbon dioxide are non-polar. We may thus divide dielectrics into two classes according to this property of their molecules.

![Diagram of dipole moments](image)

Fig. 13.19

We deal first with non-polar dielectrics. When such a dielectric is subjected to an electric field the nuclei and the electrons in the molecules experience oppositely directed forces of equal magnitude. There is no net translational force on a molecule, but it does experience distorting forces which alter the symmetrical arrangement of the electrons, so that the molecule acquires a dipole moment; the dielectric is said to be polarized (Fig. 13.19).

If a slab of dielectric is placed between the charged plates of a parallel-plate capacitor the polarization results in one face of the dielectric becoming positively charged and the other negatively charged
These charges result in a reduction of the potential difference between the plates of the capacitor (if it is isolated), and hence an increase in its capacitance.

If the molecules of the dielectric are polar they turn so that the dipole points along the direction of an applied electric field. The charges on the faces of the dielectric are then much greater than for a non-polar dielectric; this explains the high dielectric constant of polar dielectrics (e.g. water).

13.15 Thermo-electric Phenomena

(a) The Seebeck effect

If two metals (A and B) are joined together as in Fig. 13.21 and the junctions maintained at different temperatures an e.m.f. is set up which drives a current round the circuit. The magnitude of this e.m.f. depends on the temperature difference between the junctions, and typically is of the order of microvolts per degree temperature difference. The student will already be familiar with the use of this effect in thermometry.

The insertion of a galvanometer into the circuit has no effect on the e.m.f., provided that both its terminals are at the same temperature. This follows from an empirical law that

If a circuit consisting of two metals A and B has another metal C inserted into it (Fig. 13.22) the e.m.f. produced when the junctions of A with B are at different temperatures is unaffected by the presence of C, provided that both ends of C are at the same temperature.

This is referred to as the law of intermediate metals. The law may also be deduced theoretically.

Another law, the law of intermediate temperatures, may be stated as follows.

If a thermocouple has an e.m.f. $E_{12}$ when its junctions are at temperature $T_1$ and $T_2$, and an e.m.f. $E_{23}$ when its junctions are at temperatures $T_2$ and $T_3$, then the e.m.f. $E_{13}$ when its junctions are at temperatures $T_1$ and $T_3$ is given by $E_{13} = E_{12} + E_{23}$.

For most pairs of metals the variation of e.m.f. with the temperature difference between the junctions is approximately parabolic (Fig. 13.23). The e.m.f. reaches a maximum at a certain temperature difference
neutral temperature) and becomes zero again at approximately twice this value (temperature of inversion). For higher temperature differences the direction of the e.m.f. is reversed. If the neutral temperature is high the curve is approximately linear up to 100°C, so that the thermocouple would be convenient for measuring temperatures in this range.

![Diagram](image)

**Fig. 13.23**

The rate of change of e.m.f. with temperature \( \frac{dE}{dT} \) is called the **thermo-electric power**, \( P \). If the relation between \( E \) and \( T \) is parabolic, then the relation between \( P \) and \( T \) is linear.

(b) The Peltier effect

This is the reverse of the Seebeck effect. The current flowing in the circuit in Fig. 13.24 causes a temperature difference between the two junctions, i.e. heat is absorbed at one junction and liberated at the other. There is a Seebeck e.m.f. set up by this temperature difference and, in order not to violate the principle of conservation of energy, this e.m.f. must be in such a direction that it opposes the current already flowing in the circuit.

![Diagram](image)

**Fig. 13.24**

The heat (in joules) liberated or absorbed when 1 coulomb flows across a junction is called the **Peltier coefficient** \( \pi \). As \( \pi \) has the units \( \text{joules/coulombs} \), it is a potential difference measured in volts. The origin of this potential difference is the contact potential between two metals (Section 13.5).

The Seebeck effect may also be explained in terms of contact potentials. If the two junctions of a thermocouple are at the same temperature the two contact potentials are equal and opposite, so that the net
ELECTRICAL AND MAGNETIC PROPERTIES 255
e.m.f. is zero. However, the contact potential varies slightly with temperature, so that there is a net e.m.f. when the two junctions are at different temperatures.

(c) Thomson effect

When there is a temperature gradient along a conductor there is also a potential gradient, which is usually directed from the end at lower temperature to that at higher temperature, although in some metals it is the opposite way.

The effect is explained in terms of the different mobilities of the conduction electrons at different temperatures, causing a greater rate of diffusion along the bar in one direction than the other, until the distribution of electrons is such that a potential difference is set up which just counteracts the effect of diffusion. Thomson e.m.f.s are much smaller than Peltier or Seebeck e.m.f.s.

The Thomson coefficient \( \delta \) is defined as the ratio \( \delta E/\delta T \), where \( \delta E \) is the e.m.f. between two points in the metal which are at a temperature difference \( \delta T \). The Thomson e.m.f. between two points in a metal whose temperatures are \( T_1 \) and \( T_2 \) is \( \int_{T_1}^{T_2} \sigma \, dT \).

For a thermocouple the Seebeck e.m.f. is the sum (taking into account signs) of the Peltier coefficients at the two junctions and the Thomson e.m.f.s along the two wires.

MAGNETIC PROPERTIES

13.16 Definitions

These are phrased in terms of c.g.s. units; for the m.k.s. versions see Chapter 10.

The intensity of magnetization, \( J \), of a substance is defined as its magnetic moment per unit volume,

\[
J = \frac{M}{V}
\]

i.e.

\[
M = \frac{2ml}{A} = \frac{m}{A}
\]

We have assumed that the specimen is uniformly magnetized.

The magnetic susceptibility, \( \kappa \), of a material is defined as the ratio between its intensity of magnetization and the magnetizing field:

\[
\kappa = \frac{J}{H}
\]

If a magnetic field exists in a vacuum there are a certain number of lines of force associated with this field. A unit magnetic field (1 oersted)
is represented by one line of force per unit area perpendicular to the field; it follows that $4\pi$ lines of force originate from a unit magnetic pole (cf. Section 9.1). In the presence of a material medium the number of lines is increased due to the induced magnetism in the medium, and the total number of lines is called a magnetic flux, $\Phi$, measured in maxwell.

The flux density or magnetic induction, $B$, in a substance is the number of lines of flux crossing unit area drawn perpendicular to these lines. A flux density of 1 maxwell cm$^{-2}$ is called one gauss.

The permeability, $\mu$, of a material is defined as the ratio of the flux density to the strength of the magnetizing field: $\mu = B/H$.

We may establish a relation between $B$, $H$ and $J$ by considering a substance in a magnetic field, $H$, and calculating the number of lines of flux per unit area in the medium. Imagine a thin disc-shaped cavity cut in the material (Fig. 13.25). If the area of each face of the disc is $A$ and the intensity of magnetization of the material is $J$ there are magnetic poles of strength $m = JA$ on each face. The number of lines of flux due to these poles is thus $4\pi JA$, and the total number of lines is

$$\Phi = HA + 4\pi JA.$$  

Hence

$$B = \frac{\Phi}{A} = H + 4\pi J.$$  

Dividing by $H$ gives

$$\frac{B}{H} = \mu = 1 + 4\pi \kappa.$$  

### 13.17 Classification

All substances belong to one of the following groups:

(i) **Diamagnetics.** These have a small negative susceptibility of the order of $10^{-8}$; this means that they become magnetized in such a direction that the flux they produce opposes the applied field. A small piece of a diamagnetic material in a non-uniform magnetic field experiences a net force away from the region where the field is strongest. $\kappa$ is practically independent of temperature, and of the strength of the magnetic field.

(ii) **Paramagnetics.** These have a small positive susceptibility of the order of $10^{-4}$; they become magnetized in such a direction that the flux they produce is in the same direction as the applied field. A small piece of a paramagnetic material in a non-uniform magnetic field experiences a net force towards the region where the field is strongest. $\kappa$ is approximately proportional to $1/T$, and independent of the strength of the magnetic field except for high fields and low temperatures.
(iii) Ferromagnetics. These have a large positive susceptibility of the order of $10^8$. $\kappa$ depends on temperature and on the strength of the magnetic field. Unlike paramagnetics and diamagnetics, ferromagnetics retain some of their magnetism when the magnetizing field is switched off. This group comprises substances often referred to as magnetic materials (iron, nickel).

13.18. Properties of Ferromagnetics

The susceptibility of a ferromagnetic varies with $H$ at moderate temperatures, but above a certain temperature (the Curie point) the material behaves like a paramagnetic and $\kappa$ is independent of $H$. The variation of $\kappa$ with $H$ for iron at room temperature is shown in Fig. 13.26; the Curie temperature of iron is about 1,000°K.

Another way of representing the behaviour of a ferromagnetic graphically is to plot $J$ against $H$. Fig. 13.27 shows such a graph for iron. Suppose that initially the specimen is unmagnetized so that it is represented by the point O. As $H$ is increased the curve OA results; at A the iron is saturated, and further increase in $H$ does not affect $J$.

If the field is now reduced the curve AB is obtained. This shows that when $H = 0$ (point B) the specimen still retains some of its magnetism. OB is called the remanence. If $H$ is reversed the curve BCA' is followed. OC is a measure of the reverse field needed to demagnetize the specimen; it is called the coercive force. If the reverse field is de-
increased from the point represented by A' to zero, and then increased in the forward direction again, the curve A'B'C'A is obtained.

It is seen from Fig. 13.27 that the magnetization of the iron lags behind changes in $H$; this phenomenon is known as hysteresis, and the curve ABCA'B'C' as a hysteresis loop.

A similar loop is obtained if B is plotted against $H$. As $B = H + 4\pi J$ and $H$ is much less than B, the two loops are of similar shape. It may be shown that the area enclosed by the $J$-$H$ curve is the energy dissipated as heat when unit volume of the material is taken once round the cycle.

The shape of the hysteresis loop is an important factor in the selection of a magnetic material for a particular purpose. For permanent magnets alloys have been developed which have high values of remanence and coercive force. For transformers it is essential to use a material with a very narrow hysteresis loop so that the energy dissipated as heat in each cycle of the A.C. supply shall be as small as possible.

13.19 Experimental Determination of the Hysteresis Loop

For a specimen in the form of a toroid a method based on electromagnetic induction may be used (Fig. 13.28). $H$ is produced by a primary coil of many turns of wire carrying a current $I$. On top of part of the primary is wound a short secondary coil which is connected to a ballistic galvanometer. If the current in the primary is changed suddenly the resulting change in flux causes an induced charge to circulate through the ballistic galvanometer. This charge is proportional to the change in $B$ so that the required results are obtained. If qualitative results only are required it is sufficient to use the deflection of the galvanometer to denote changes in $B$ and the change in current to represent the change in $H$.

13.20 Explanation of the Magnetic Properties of Matter

All the magnetic properties of matter arise from the orbital electrons in the atoms of which matter is made up. These electrons constitute a circulating electric charge, i.e. an electric current, and hence have a magnetic effect. It may be shown that the magnetic moment of an orbital electron is proportional to its angular momentum.
ELECTRICAL AND MAGNETIC PROPERTIES

We see, then, that magnetism is not a separate phenomenon, but is always associated with electricity: magnetism is electricity in motion. Electromagnetism and permanent magnetism are not fundamentally different; they both arise from moving electric charges.

The total angular momentum of the electrons in an atom is the vector sum of the angular momenta of the individual electrons, and usually this is not zero, so that the atom has a magnetic moment. In molecules, however, the resultant angular momentum frequently is zero and there is no magnetic moment; such molecules are diamagnetic.

When a diamagnetic molecule is placed in a magnetic field forces act on the individual magnetic moments of the electrons and cause them to alter their motions in such a way that the molecule acquires a magnetic moment in the opposite direction to the field. When the field is applied the flux through the plane of each electron orbit alters so that an e.m.f. is induced. This e.m.f. alters the speed of the electrons in such a way that they produce an additional magnetic flux in a direction opposing the direction of the applied magnetic field (Lenz’s law); this accounts for the diamagnetic properties.

All molecules must show diamagnetic effects, but if the molecule has a permanent magnetic moment in zero field this masks the diamagnetism; such a molecule is paramagnetic. When the field is switched on the molecules tend to align themselves in the direction of the applied field, but are prevented from doing so completely by their thermal vibrations. The stronger the field, the greater is the tendency to alignment, so that \( J \) increases with \( H \). Detailed theory shows that the increase is linear over a wide range of conditions, so that \( \kappa \) is constant. At low temperatures thermal vibrations are considerably reduced and \( \kappa \) is much larger than at ordinary temperatures. It is also possible to achieve saturation (i.e. when all the magnetic moments are pointing the same way) at low temperatures.

Ferromagnetics are a particular type of paramagnetic in which the molecules form themselves into groups in such a way that within one group all the molecules are pointing the same way. The groups are referred to as domains. In an unmagnetized specimen the domains are pointing in random directions. When a steadily increasing magnetic field is applied the domain boundaries move, and the domains whose magnetic moments are pointing roughly in the direction of the field grow in size at the expense of the others. At higher fields rotation of the domains takes place until saturation is reached.
BIBLIOGRAPHY

The books listed below contain further information about the topics covered in the chapter(s) indicated in the left-hand column.

Chapters

1, 2, 6 CHAMPION and DAVY, Properties of Matter (Blackie)
2, 3, 6 NEWMAN and SEARLE, The General Properties of Matter (Benn)
1, 2, 6 TEMPERLEY, Properties of Matter (University Tutorial Press)
3 COULSON, Waves (Oliver and Boyd)
3, 4, 11 DITCHBURN, Light (Blackie)
3, 4, 11 JENKINS and WHITE, Fundamentals of Optics (McGraw-Hill)
4 ANDREWS, Optics of the Electromagnetic Spectrum (Prentice Hall)
4 WILLIAMS, Applications of Interferometry (Methuen)
5, 6, 7, 8 ROBERTS, Heat and Thermodynamics (Blackie)
5, 7, 8 ZEMANSKY, Heat and Thermodynamics (McGraw-Hill)
6 JEANS, An Introduction to the Kinetic Theory of Gases (Cambridge)
7 PIPPARD, Elements of Classical Thermodynamics (Cambridge)
8, 13 JACKSON, Low Temperature Physics (Methuen)
9, 10, 13 BLEANEY and BLEANEY, Electricity and Magnetism (Oxford)
9 COULSON, Electricity (Oliver and Boyd)
9 FEWKES and YARWOOD, Electricity, Magnetism and Atomic Physics (Vol I) (University Tutorial Press)
9, 10, 13 HARNWELL, Principles of Electricity and Electromagnetism (McGraw-Hill)
9 RAMSEY, Electricity and Magnetism (Cambridge)
11 BOHM, Quantum Theory (Constable)
11, 12 BORN, Atomic Physics (Blackie)
11 HEITLER, Wave Mechanics (Oxford)
11 MOTT, Elements of Wave Mechanics (Cambridge)
11, 12 RICHTMEYER, KENNARD and LAURITSEN, Introduction to Modern Physics (McGraw-Hill)
11, 12 TOLENSKY, Introduction to Atomic Physics (Longmans)
11, 12 YARWOOD, Electricity, Magnetism and Atomic Physics (Vol II) (University Tutorial Press)
12 CROWTHER, Ions, Electrons and Ionizing Radiations (Arnold)
12 HARNWELL and LIVINGOOD, Experimental Atomic Physics (McGraw-Hill)
13 DEKKER, Solid State Physics (Macmillan)
13 STONER, Magnetism (Methuen)
13 WRIGHT, Semi-conductors (Methuen)

260
INDEX

Absorptivity, 110
Acceleration, angular, 1
  of gravity, determination, 9-13
  variation, 20-2
Acceptor circuit, 185
A.C. circuits, 182-8
  use of complex numbers, 186
Adiabatic, change, work done, 135
  demagnetization, 159
Ampere, the, 194
Angle of contact, 25
Angular acceleration, 1
  momentum, 6
    conservation of, 7
    quantization of, 223
motion, equations, 6
  velocity, 1
Atomic heat, 131, 240
Attracted disc electrometer, 168, 207
Avogadro's number, 118
  determination, 129
Ballistic galvanometer, 174
Balmer's formula, 217
Band, conduction, 245
  energy, 243
  theory, 244
Base, 250
Beats, 54
Biot-Savart formula, 207
Black body, 110
  radiation, 110, 208
  spectrum, 113
Bohr atom, 218
  orbits, radii, 224
Boiling, effect of surface tension, 29
Boltzmann's constant, 118
Boyle temperature, 156
Brownian motion, 128
Capacitance in D.C. circuit, 178, 181
  in A.C. circuit, 182-7
  of coaxial cable, 167, 207
  of parallel plates, 205
Carnot, cycle, 137
  refrigerator, 138
Carnot's theorem, 140
Cascade process, 154
Charged conductor, force on surface, 167, 206
Claude process, 157
Clausius', statement of second law of
  thermodynamics, 140
  theorem, 145
Clausius-Clapeyron equation, 150
Coaxial cable, capacitance of, 107, 207
  Coercive force, 257
  Collector, 250
  Collins helium liquefier, 158
  Complex numbers in A.C. problems, 186
  Compound pendulum, 9
  Compton effect, 210
  Condensation, and surfaced tension, 29
    on electrically charged particles, 30
  Conduction, band, 245
    of heat 102-8
  Contact potential, 243
  Convection, forced, 108, 109
    natural, 108, 109
  Cornu spiral, 96-9
  Coulomb, the, 195
  Coulomb's law, 202
  Couple, work done by, 7
  Critical, damping, 47
    temperature, 153
    velocity, 32
  Current elements, 192
  Cycle, Carnot, 137
  Cycles of changes, 136
Damping, correction, 176
  critical, 47
  electromagnetic, 175
D.C. circuits, 176-81
  de Broglie wavelength, 210
  Debye temperature, 132
  Degrees of freedom, 121
  Demagnetization, adiabatic, 159
  Diamagnetics, 256, 259
  Dielectrics, 204, 252
  Diffraction, Fraunhofer, 80-5
  Fresnel, 91-9
  grating, 87-90
  resolving power of, 90
Diode, junction, 248
  point contact, 250
Die, electric, 252
  moment, magnetic, 200
Dispersion, 56
Domains, 259
Drift velocity, 237
Drude's theory of metallic conduction,
  236
Dulong and Petit's law, 131
Dyne, the, 190
Echelon gratings, 91
Efficiency of heat engine, 138
Electric, dipole, 252
  field, intensity, 166, 203
  flux, 203
Electromagnetic induction, 196
Electrometer, attracted disc, 168, 207
  quadrant, 169

261
<table>
<thead>
<tr>
<th>Index Word(s)</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electron, bound</td>
<td>214, 219</td>
</tr>
<tr>
<td>'free',</td>
<td>236</td>
</tr>
<tr>
<td>gas,</td>
<td>236</td>
</tr>
<tr>
<td>probability distribution,</td>
<td>225</td>
</tr>
<tr>
<td>shells,</td>
<td>232</td>
</tr>
<tr>
<td>spin,</td>
<td>230</td>
</tr>
<tr>
<td>Electronic work function,</td>
<td>243</td>
</tr>
<tr>
<td>Elliptic orbits,</td>
<td>226</td>
</tr>
<tr>
<td>Emissive power,</td>
<td>110</td>
</tr>
<tr>
<td>Emissivity,</td>
<td>110</td>
</tr>
<tr>
<td>Emitter,</td>
<td>250</td>
</tr>
<tr>
<td>Energy, bands,</td>
<td>244</td>
</tr>
<tr>
<td>conservation of,</td>
<td>8, 69</td>
</tr>
<tr>
<td>equipartition of,</td>
<td>121</td>
</tr>
<tr>
<td>levels,</td>
<td>214, 220, 240</td>
</tr>
<tr>
<td>in hydrogen,</td>
<td>224</td>
</tr>
<tr>
<td>unavailable,</td>
<td>148</td>
</tr>
<tr>
<td>Entropy,</td>
<td>144</td>
</tr>
<tr>
<td>change in reversible cycle,</td>
<td>146</td>
</tr>
<tr>
<td>principle of increase,</td>
<td>147</td>
</tr>
<tr>
<td>Equipartition of energy,</td>
<td>121</td>
</tr>
<tr>
<td>Equivalent simple pendulum,</td>
<td>11</td>
</tr>
<tr>
<td>Erg, the,</td>
<td>190</td>
</tr>
<tr>
<td>Escape velocity,</td>
<td>19</td>
</tr>
<tr>
<td>Evaporation, effect of surface tension on,</td>
<td>29</td>
</tr>
<tr>
<td>Excited states of atom,</td>
<td>220</td>
</tr>
<tr>
<td>Exclusion principle, Pauli,</td>
<td>232, 240</td>
</tr>
<tr>
<td>Faraday's law,</td>
<td>196</td>
</tr>
<tr>
<td>Fermat's principle,</td>
<td>63-5</td>
</tr>
<tr>
<td>Fermi, energy,</td>
<td>241</td>
</tr>
<tr>
<td>level,</td>
<td>241</td>
</tr>
<tr>
<td>Ferromagnetics,</td>
<td>257, 259</td>
</tr>
<tr>
<td>properties,</td>
<td>257</td>
</tr>
<tr>
<td>Field, conservative,</td>
<td>14</td>
</tr>
<tr>
<td>electric,</td>
<td>166, 203</td>
</tr>
<tr>
<td>gravitational,</td>
<td>13-19</td>
</tr>
<tr>
<td>magnetic,</td>
<td>171</td>
</tr>
<tr>
<td>Flow, of liquid through tube,</td>
<td>33</td>
</tr>
<tr>
<td>streamline,</td>
<td>32</td>
</tr>
<tr>
<td>turbulent,</td>
<td>32</td>
</tr>
<tr>
<td>Flux density, electric,</td>
<td>203</td>
</tr>
<tr>
<td>magnetic,</td>
<td>193, 256</td>
</tr>
<tr>
<td>Flux, electric,</td>
<td>203</td>
</tr>
<tr>
<td>magnetic,</td>
<td>195, 256</td>
</tr>
<tr>
<td>Fountain effect,</td>
<td>162</td>
</tr>
<tr>
<td>Force, magnetizing,</td>
<td>197</td>
</tr>
<tr>
<td>on surface of charged conductor,</td>
<td>167</td>
</tr>
<tr>
<td>Frank–Hertz experiment,</td>
<td>221</td>
</tr>
<tr>
<td>'Free' electrons,</td>
<td>236</td>
</tr>
<tr>
<td>Fringes, Young's,</td>
<td>65-70, 77</td>
</tr>
<tr>
<td>white light,</td>
<td>74</td>
</tr>
<tr>
<td>wedge,</td>
<td>75</td>
</tr>
<tr>
<td>localized and non-localized,</td>
<td>77</td>
</tr>
<tr>
<td>Galvanometer, ballistic,</td>
<td>174</td>
</tr>
<tr>
<td>Gases, kinetic theory,</td>
<td>116-28</td>
</tr>
<tr>
<td>liquefaction,</td>
<td>153-8</td>
</tr>
<tr>
<td>specific heats,</td>
<td>121-4</td>
</tr>
<tr>
<td>viscosity,</td>
<td>40</td>
</tr>
<tr>
<td>Gauss' theorem,</td>
<td>165, 207</td>
</tr>
<tr>
<td>applications,</td>
<td>166</td>
</tr>
<tr>
<td>Gauss, the,</td>
<td>256</td>
</tr>
<tr>
<td>Gravitation, Newton's law,</td>
<td>13</td>
</tr>
<tr>
<td>Gravitational, field,</td>
<td>13-19</td>
</tr>
<tr>
<td>potential,</td>
<td>13-18</td>
</tr>
<tr>
<td>Gravity, acceleration due to,</td>
<td>9-13, 20-2</td>
</tr>
<tr>
<td>Ground state of atom,</td>
<td>220</td>
</tr>
<tr>
<td>Group velocity,</td>
<td>57</td>
</tr>
<tr>
<td>Half-period zones,</td>
<td>91</td>
</tr>
<tr>
<td>Hall, effect,</td>
<td>239</td>
</tr>
<tr>
<td>coefficient,</td>
<td>239, 246</td>
</tr>
<tr>
<td>Hampson process,</td>
<td>156</td>
</tr>
<tr>
<td>Heat, engine,</td>
<td>133, 137, 140</td>
</tr>
<tr>
<td>exchanger,</td>
<td>156</td>
</tr>
<tr>
<td>Heisenberg uncertainty relation,</td>
<td>212</td>
</tr>
<tr>
<td>Helium, liquefaction,</td>
<td>158</td>
</tr>
<tr>
<td>liquid,</td>
<td>161</td>
</tr>
<tr>
<td>solid,</td>
<td>163</td>
</tr>
<tr>
<td>Henry, the,</td>
<td>173</td>
</tr>
<tr>
<td>Heylandt process,</td>
<td>157</td>
</tr>
<tr>
<td>Holes,</td>
<td>246</td>
</tr>
<tr>
<td>Huygens' construction,</td>
<td>61</td>
</tr>
<tr>
<td>Hydrogen spectrum,</td>
<td>217</td>
</tr>
<tr>
<td>Hysteresis,</td>
<td>258</td>
</tr>
<tr>
<td>Impedance,</td>
<td>183</td>
</tr>
<tr>
<td>Impulse,</td>
<td>6</td>
</tr>
<tr>
<td>Impurity semi-conductors,</td>
<td>246</td>
</tr>
<tr>
<td>Indicator diagram,</td>
<td>133</td>
</tr>
<tr>
<td>Intensity of magnetization,</td>
<td>198, 255</td>
</tr>
<tr>
<td>Interference,</td>
<td>65-80</td>
</tr>
<tr>
<td>in thin films,</td>
<td>70-4</td>
</tr>
<tr>
<td>Interferometry,</td>
<td>78</td>
</tr>
<tr>
<td>Inductance, in D.C. circuit,</td>
<td>180-2</td>
</tr>
<tr>
<td>in A.C. circuit,</td>
<td>183-7</td>
</tr>
<tr>
<td>Induction, electromagnetic,</td>
<td>196</td>
</tr>
<tr>
<td>mutual,</td>
<td>173</td>
</tr>
<tr>
<td>self,</td>
<td>173</td>
</tr>
<tr>
<td>Intrinsic semi-conductors,</td>
<td>245</td>
</tr>
<tr>
<td>Inversion temperature,</td>
<td>155</td>
</tr>
<tr>
<td>Ionization, energy,</td>
<td>224</td>
</tr>
<tr>
<td>potential,</td>
<td>224</td>
</tr>
<tr>
<td>Irreversible, changes,</td>
<td>135, 147</td>
</tr>
<tr>
<td>cycle, entropy change,</td>
<td>146</td>
</tr>
<tr>
<td>Isothermal change, work done,</td>
<td>134</td>
</tr>
<tr>
<td>Joule, the</td>
<td>190</td>
</tr>
<tr>
<td>Joule–Kelvin effect,</td>
<td>155</td>
</tr>
<tr>
<td>Junction diode,</td>
<td>248</td>
</tr>
<tr>
<td>transistor,</td>
<td>250</td>
</tr>
<tr>
<td>Kater's pendulum,</td>
<td>12</td>
</tr>
<tr>
<td>Kelvin, statement of second law of thermodynamics,</td>
<td>139</td>
</tr>
<tr>
<td>temperature scale,</td>
<td>141</td>
</tr>
<tr>
<td>Kinetic energy of rotation,</td>
<td>2</td>
</tr>
<tr>
<td>theory of gases,</td>
<td>116-28</td>
</tr>
<tr>
<td>of liquids,</td>
<td>128-31</td>
</tr>
<tr>
<td>of solids,</td>
<td>131-2</td>
</tr>
<tr>
<td>Kirchhoff's, law of radiation,</td>
<td>111</td>
</tr>
<tr>
<td>laws in electricity,</td>
<td>176</td>
</tr>
</tbody>
</table>
INDEX

Linde process, 156
Liquid helium, 161

Mutual induction, 173
Magnetic circuits, 198
field, 171-3
flux, 195, 256
flux density, 193, 256
induction, 193
materials, 198
moment, 200
susceptibility, 256
Magnetization, intensity of, 256
Magnetizing force, 197
Magnetomotive force, 198

Matter waves, 210
Maxwell, the 256
Maxwell's distribution law, 120
Mean free path, 124, 238
Metal rectifier, 250
Michelson interferometer, 78
Mobility of electrons and holes, 246
Modulation, 58
Moment of inertia, 2
of certain bodies, 3

Neutral temperature, 254
Newton, the, 190
Newton's law of cooling, 109
law of gravitation, 13
rings, 75, 77
n-type semi-conductors, 247

Orbits, Bohr, 218, 224
eliptic, 226
Oscillations, 43-9
of rigid body, 8-9

Parallel axes theorem, 3
Paramagnetics, 159, 256, 259
Pauli exclusion principle, 232, 240
Peltier effect, 254
Pendulum, compound, 9
equivalent simple, 11
Kater's, 12
Periodic table, 233
Permeability, 256
of free space, 192
relative, 198
Permittivity, of free space, 202
relative, 204
Perpendicular axes theorem, 3
Perrin's determination of Avogadro's number, 129
Photo-electric effect, 208
Photons, 208, 210
Pictet process, 153
Planck's constant, 208
Potential, contact, 243
difference, 203
gravitational, 13-18
ionization, 224
Point-contact diode, 250

Poiseille's formula, 33
Polar molecules, 252
Power, in A.C. circuits, 187
factor, 188
thermo-electric, 254
Pressure, difference across curved liquid surface, 26
variation with height in atmosphere, 129
p-type semi-conductors, 248

Quadrant electrometer, 169
Quantization of angular momentum, 223
Quantum numbers, 227

Radiation, black-body, 110, 208
thermal, 110
Radius of gyration, 2
Rayleigh, criterion, 84
interferometer, 78
Rayleigh-Jeans formula, 114, 208
Reactance, 182
Rectifier, metal, 250
Rectilinear propagation, 99
Regenerative process, 156
Reluctance, 199
Remanence, 257
Resolving power, 83
doifraction grating, 90
of microscope, 85
of telescope, 85
Resonance, 49, 185
Reversible, changes, 135, 146
cycles, 149-52
heat engine, 140
pendulum, 12
Reynolds' number, 32, 36
Ritz combination principle, 221
Root mean square velocity, 117
Rutherford-Bohr atom, 218
Rydberg constant, 217, 224

Schroedinger equation, 215, 225
Seebeck effect, 253
Selection rules, 228
Self-induction, 173
Semi-conductors, 236
impurity, 246
intrinsic, 245
n-type, 247
p-type, 248
Sheils, electron, 234
Simple harmonic motion, angular, 8
damped, 44
forced, 47
free, 43
Simon helium liquefier, 158
Solid helium, 163
Sommerfeld, elliptic orbits, 226
Space quantization, 229
Specific heat of gases, 121
variation with temperature, 123
Spectrum, hydrogen, 217
INDEX

Spinning electron, 230
Stationary waves, 53
Stefan's law, 112
constant, 112
Stokes' formula, 39
Streamline flow, 32
Superconductivity, 163
Superposition, 53
Surface energy, 24, 151
tension, 24-30
Susceptibility, magnetic, 255

Temperature, Boyle, 156
critical, 153
inversion, 155
neutral, 254
Temperature scale, Kelvin, 141
perfect gas, 142
Terminal velocity, 39
Thermal conductivity, 101-8
kinetic theory, 127
Thermal radiation, 110
Thermionic emission, 242
Thermistor, 248
Thermodynamics, first law, 133
second law, 139
Thermo-electric effects, 253-5
power, 254
Thermometers, low temperature, 160-1
Thin film interference, 70-4
Thin lens formula, 62
Thomson effect, 255
Throttling process, 155
Time constant, 178
Torque on cylinder in rotating liquid, 37
Transformer, 188
Transients, 185
Transistor, 250
Turbulent flow, 32

Unavailable energy, 148
Uncertainty relation, 212

Vapour pressure above a curved surface, 27
Velocity, angular, 1
critical, 32
drift, 237
escape, 19
group, 57
root mean square, 117
terminal, 39
Viscosity, 30-42
determination, 35-42
kinetic theory, 125
of gases, 40-2
Volt, the, 203

Watt, the, 190
Waves, 49-59
harmonic, 49
matter, 210
stationary, 53
Wave, mechanics, 211, 225
motion, equation of, 52
Weber, the, 196
Wedge fringes, 75
White light fringes, 74
Wiedemann-Franz law, 101, 236
Wien's laws, 114
Wilson cloud chamber, 30
Work function, 209, 243
Working substance, 137

Young's fringes, 65-70, 77

Zeeman effect, 229
Zone plate, 95