

This textbook explains the concepts and most important advances of modern physics without resort to higher mathematics. Avoids the traditional division between classical and modern physics and endeavours to present all material so as to develop quantum mechanical concepts.

The textbook is intended for secondary schools and as a teaching aid for physics teachers in general and technical secondary schools. Will be found useful by correspondence students studying 'A' level and first year physics.

Contents. Vol. II. Vibrations and Waves. Quantum Physics of Atoms, Molecules and Solids. Physics of the Nucleus and Elementary Particles.

Prof. BORIS YAVORSKY, Doctor of Physical and Mathematical Sciences, is in the department of theoretical physics at the Lenin State Pedagogical Institute in Moscow. He has been lecturing in higher educational institutions of the Soviet Union for the last 35 years. Prof. Yavorsky has written about 300 published works: books and articles in various scientific journals and collected papers, among them the *Complete Course in Physics* (in three volumes) and some other textbooks and aids for institute teachers.

Assoc. Prof. ARKADY PINSKY, Candidate of Pedagogical Sciences, is a senior scientist at the Scientific Research Institute for Teaching Practice of the USSR Academy of Pedagogical Sciences. He has specialized in the methods of physics teaching at school and university levels. In this field he has published over 50 works, including several books on methods of teaching physics.

Б. М. Яворский, А. А. Пинский
ОСНОВЫ ФИЗИКИ
Том I

Движение и силы;
законы сохранения;
молекулярно-кинетическая теория газа;
молекулярные силы
и агрегатные состояния
вещества;
электродинамика

Издательство «Наука»

B. M. YAVORSKY and A. A. PINSKY

FUNDAMENTALS of PHYSICS

VOLUME

I

MOTION AND FORCES.

CONSERVATION LAWS.

MOLECULAR-KINETIC THEORY OF GASES.

MOLECULAR FORCES AND STATES
OF AGGREGATION OF MATTER.

ELECTRODYNAMICS.

Translated from the Russian
by

NICHOLAS WEINSTEIN

MIR PUBLISHERS MOSCOW

First published 1975
Revised from the 1974 Russian edition

На английском языке

© Издательство «Наука», 1974
© English translation, Mir Publishers, 1975

CONTENTS

FOREWORD	15
---------------------------	-----------

Part one MOTION AND FORCES

Chapter 1 VELOCITY	19
1.1. Mechanical motion	19
1.2. Frames of reference. Paths	20
1.3. Rectilinear motion. Motion equations	22
1.4. Uniform motion	23
1.5. Variable motion	24
1.6. Average velocity	25
1.7. Instantaneous velocity of variable motion	27
Chapter 2 INERTIA	28
2.1. The principle of inertia	28
2.2. Inertial reference frames	30
2.3. The principle of relativity	32
2.4. Galilean transformations	33
2.5. Classical law of the addition of velocities	35
Chapter 3 SCALARS AND VECTORS. VELOCITY VECTOR	35
3.1. Scalar quantities	35
3.2. Vector quantities	36
3.3. Certain operations on vectors	37
3.4. Resolving a vector into two components	39
3.5. Velocity is a vector	40
3.6. Addition of velocities	42
Chapter 4 ACCELERATION	43
4.1. Average and instantaneous acceleration	43
4.2. Rectilinear variable motion	44
4.3. Uniformly accelerated rectilinear motion	44
4.4. Velocity graph for uniformly accelerated motion	45
4.5. Graphical calculation of displacement	45
4.6. Displacement and average velocity in uniformly accelerated motion	46
4.7. Uniform circular motion of a particle	48
4.8. Acceleration in uniform circular motion of a particle	48

Chapter 5 FORCE	50
5.1. Force, a measure of the interaction of bodies	50
5.2. Elastic and plastic deformation	51
5.3. Force is a vector	52
5.4. Vector addition and resolution of forces applied to a particle	54
Chapter 6 WEIGHT AND MASS	55
6.1. Force of gravity. Weight	55
6.2. Free fall	57
6.3. Mass of a body	58
6.4. The density of substances	59
Chapter 7 FUNDAMENTAL LAW OF DYNAMICS	60
7.1. Force and acceleration	60
7.2. Applying the fundamental law of dynamics	63
7.3. Weightlessness	66
7.4. System of units	67
7.5. The international system of units	68
7.6. The cgs and mk(force)s systems of units	69
Chapter 8 EQUATIONS OF MOTION AND INITIAL CONDITIONS	69
8.1. The basic problem of dynamics	69
8.2. Motion of a particle subject to the force of gravity	70
8.3. Numerical solution of the basic problem of dynamics	72
8.4. Motion of a body subject to an elastic force	73
8.5. Quantities determining the motion equation of a particle	77
Chapter 9 GRAVITATION	78
9.1. Discovery of the law of gravitation	78
9.2. Newton's law of universal gravitation	80
9.3. The Cavendish experiment	81
9.4. Determining the distances from the sun to the planets	82
9.5. The gravitational field	83
9.6. Gravitational field intensity	84
9.7. The earth's gravitational field	84
9.8. Effect of the earth's rotation of free-fall acceleration	86
Chapter 10 ELECTRIC FORCES	87
10.1. Electric charge	87
10.2. Coulomb's law	88
10.3. Units of charge and systems of units	90
10.4. The electric dipole	92
10.5. The electric field. Field strength	94
10.6. Electric field of a point charge and of a dipole	94
Chapter 11 FRICTION	97
11.1. External and internal friction	97
11.2. Static friction	97
11.3. The angle of friction	100
11.4. Sliding friction	101
11.5. Rolling friction	102
11.6. Motion of bodies subject to the force of friction	102
11.7. Internal friction	103
11.8. Motion of bodies in fluids	105
11.9. Bodies falling in a fluid	108

Chapter 12 THE THEORY OF RELATIVITY	110
12.1. Velocity of light and the law of addition of velocities	110
12.2. Basic postulates of the special theory of relativity	113
12.3. Simultaneity of events	113
12.4. Simultaneity and length	116
12.5. Relativistic law for the addition of velocities	116
12.6. Limiting nature of the velocity of light	118
12.7. Lorentz transformations	119
12.8. Length or distance	120
12.9. Time interval between two events	121
12.10. The time interval between cause and effect	122
12.11. The relation between relativistic and Newtonian mechanics	123
Chapter 13 MASS, MOMENTUM AND FORCE IN THE THEORY OF RELATIVITY	125
13.1. Relativistic mass	125
13.2. The fundamental law of dynamics in the theory of relativity	126
13.3. The relation between Newtonian and relativistic dynamics	128
Chapter 14 EQUATIONS OF MOTION AND THE UNCERTAINTY RELATION	130
14.1. Initial conditions and measuring apparatus	130
14.2. The uncertainty relation	134
14.3. The uncertainty relation and classical mechanics	134

Part two CONSERVATION LAWS

Chapter 15 LAW OF CONSERVATION OF LINEAR MOMENTUM	138
15.1. Closed system of bodies	138
15.2. Law of conservation of linear momentum	139
15.3. Recoil phenomena	141
15.4. Measurement of mass	142
15.5. Jet propulsion (propulsion by reaction)	143
15.6. Rocket fuel calculations	144
15.7. Centre of mass	145
15.8. Motion of the centre of mass	146
Chapter 16 TOTAL AND KINETIC ENERGY	148
16.1. Total energy of a body	148
16.2. Kinetic energy	149
16.3. Energy and linear momentum	150
16.4. Kinetic energy and work	151
16.5. Power	153
16.6. Units of energy, work and power	154
16.7. Momentum, and energy of a localized particle	155
Chapter 17 ELEMENTARY COLLISION THEORY	157
17.1. What is a collision?	157
17.2. Completely inelastic collision	158
17.3. Elastic collision	160
17.4. Neutron moderation	162
17.5. Pressure of a stream of particles on a wall	163
Chapter 18 CONSERVATIVE FORCES AND POTENTIAL ENERGY	165
18.1. Work done by a variable force	165
18.2. Work done by an elastic force	167

18.3. Work done by a Coulomb force	168
18.4. Work done by a gravitational force	170
18.5. Conservative forces	171
18.6. Potential energy of elastic, Coulomb and gravitational interactions	172
18.7. Potential of an electrostatic field	174
18.8. Electric potential of the field set up by a point charge	175
18.9. Energy of an electric field	176
Chapter 19 LAW OF CONSERVATION OF ENERGY IN NEWTONIAN MECHANICS	177
19.1. Mechanical energy and its conservation	177
19.2. Mechanical energy and friction	177
19.3. Space velocities	178
19.4. Looping the loop	179
19.5. Potential energy curves	180
19.6. Potential energy and equilibrium	183
Chapter 20 INTERNAL ENERGY	184
20.1. Internal energy of a system of particles	184
20.2. Changes in internal energy when a body is deformed	185
20.3. Changes in internal energy of a body in thermal processes	186
20.4. Changes in internal energy in chemical reactions	187
20.5. Changes in internal energy in nuclear reactions	188
Chapter 21 THE LAW OF CONSERVATION OF ENERGY	189
21.1. Work as a measure of the change in total and internal energy	189
21.2. Heat exchange	190
21.3. Quantity of heat	192
21.4. The first law of thermodynamics	193
21.5. An adiabatically isolated system	194
21.6. The law of conservation of energy	194
21.7. The law of conservation of mass	195
21.8. More about relativistic mass	197
Chapter 22 THE LAW OF CONSERVATION OF ANGULAR MOMENTUM	199
22.1. Features of rotational motion	199
22.2. Kinetic energy and moment of inertia	199
22.3. Dependence of the moment of inertia on the location of the axis of rotation	201
22.4. Moment of force	204
22.5. Equilibrium conditions for a body having an axis of rotation	205
22.6. Angular momentum and the fundamental law of dynamics	205
22.7. The law of conservation of angular momentum	207
22.8. Analogies between quantities and their relations in translational and rotational motion	209
Chapter 23 SYMMETRY IN NATURE AND THE CONSERVATION LAWS	211
23.1. The conservation laws are nature's principal laws	211
23.2. The conservation laws are forbiddenness principles	212
23.3. The conservation laws and space-time symmetry	213
23.4. Uniformity of time and the conservation of energy	214

Chapter 24 NONINERTIAL FRAMES OF REFERENCE AND GRAVITATION	214
24.1. Phenomena in an accelerated reference frame	214
24.2. Inertial forces	216
24.3. Features of inertial forces	218
24.4. Space and time in noninertial reference frames	219
24.5. The principle of equivalence	222
24.6. An idea of Einstein's theory of gravity	224
24.7. The twin paradox	228
Part three	
MOLECULAR-KINETIC THEORY OF GASES	
Chapter 25 MOLECULAR MOTION	231
25.1. How molecular speeds were measured	231
25.2. Molecular speed distribution	233
25.3. Mean free path of molecules	235
25.4. Diffusion	238
25.5. Law of diffusion	239
25.6. Separation of gas mixtures	240
Chapter 26 AN IDEAL GAS	242
26.1. Gas pressure	242
26.2. Units of pressure	244
26.3. An ideal gas	246
26.4. Temperature	248
26.5. Absolute temperature and the equation of state of an ideal gas	249
26.6. The constant-volume gas thermometer	250
26.7. The degree and the kelvin. The practical and absolute temperature scales	251
26.8. Absolute zero	253
26.9. Avogadro's number and Boltzmann's constant	254
26.10. Molecule distribution in a force field	256
26.11. Barometric distribution	258
Chapter 27 AN IDEAL GAS AND THE FIRST LAW OF THERMODYNAMICS	260
27.1. Internal energy of a monatomic ideal gas	260
27.2. Work done in the expansion of an ideal gas	261
27.3. The first law of thermodynamics and the specific heat of a gas	262
27.4. Isochoric processes	264
27.5. Isobaric processes	265
27.6. Isothermal processes	266
27.7. Adiabatic processes	267
27.8. Specific heat of a diatomic gas	269
27.9. Quantum theory of the specific heats of gases	272
Chapter 28 THE SECOND LAW OF THERMODYNAMICS	276
28.1. Quasi-static processes	276
28.2. Reversible processes	277
28.3. Irreversibility of real thermal processes	278
28.4. Irreversibility and statistics	279
28.5. Diffusion and thermodynamic probability	282

28.6. Thermodynamic probability in other thermal processes	283
28.7. Thermodynamic probability and entropy	284
28.8. Entropy and heat exchange	286
28.9. The second law of thermodynamics	288
28.10. The statistical sense of the second law of thermodynamics. Fluctuations	289
28.11. Brownian motion and fluctuations	290
28.12. Brownian motion and Boltzmann's constant	291
Chapter 29 HEAT ENGINES AND REFRIGERATORS	295
29.1. Heat engines and the advancement of engineering	295
29.2. The heat engine	295
29.3. Principle and energy balance of a heat engine	296
29.4. The heat engine and the second law of thermodynamics	298
29.5. The Carnot cycle	299
29.6. The efficiency of a real engine	300
29.7. The reverse Carnot cycle	301
29.8. Refrigerators and heat pumps	302
Chapter 30 FUNDAMENTALS OF FLUID DYNAMICS	304
30.1. Thermodynamic parameters of moving fluids	304
30.2. The equation of continuity	305
30.3. The momentum equation	306
30.4. Bernoulli's equation	306
30.5. Rate of propagation of disturbance waves in elastic media	308
30.6. Compressible fluid dynamics. Mach number	311
30.7. The Mach cone	312
30.8. The bow shock wave	313
30.9. Phenomena of a normal compression shock	314
30.10. Wave drag	316
30.11. Nozzles	318
30.12. Analogy between a nozzle and a heat engine	320
30.13. The Laval nozzle	320
30.14. The jet engine	321
30.15. The airplane wing	323
30.16. Measuring the pressure and velocity in a stream of fluid	324
30.17. Viscous fluid dynamics. Fluid friction in pipes	325

Part four

MOLECULAR FORCES AND STATES OF AGGREGATION OF MATTER

Chapter 31 MOLECULAR FORCES	327
31.1. The density and compressibility of substances	327
31.2. Molecular forces	328
31.3. The electrical origin of molecular forces	329
31.4. The molecular force graph	331
31.5. The potential energy curve of molecular interaction	333
31.6. Thermal expansion of solids and liquids	334
Chapter 32 LONG-RANGE ORDER	336
32.1. The monocrystal	336
32.2. The polycrystal	338
32.3. The crystal lattice. Long-range order	339

32.4. Defects in packing and the block structure of crystals	340
32.5. Motion of defects and diffusion	341
32.6. Dislocation motion and the deformation of the crystal	342
Chapter 33 CLOSE PACKING OF PARTICLES	344
33.1. Types of crystalline bonds	344
33.2. Closest packing of identical spheres	346
33.3. Closest packing of spheres of different radii	348
33.4. Lattices which cannot be represented as the packing of spheres	348
33.5. The structure of ice	351
33.6. Polymers	351
Chapter 34 SHORT-RANGE ORDER	354
34.1. Features of the liquid state	354
34.2. The structure of a liquid, and its properties	355
34.3. Mean residence lifetime	356
34.4. Diffusion in liquids	358
34.5. The viscosity of liquids	359
34.6. Amorphous bodies	360
34.7. Energy of the surface layer and the surface tension of liquids	361
34.8. Pressure caused by the curved surface of a liquid	363
34.9. Capillary phenomena	364
34.10. Adsorption. The Rebinder effect	365
Chapter 35 VAPOURS	367
35.1. Vaporization	367
35.2. Saturated vapour	369
35.3. The pressure of saturated vapour	370
35.4. Vapour isotherms	373
35.5. The critical state of a substance	374
35.6. Air humidity	376
Chapter 36 PHASE TRANSITIONS	378
36.1. Changes in the state of aggregation	378
36.2. Liquid-gas transition diagram	379
36.3. Crystal-gas transition diagram	379
36.4. Crystal-liquid transition diagram	380
36.5. Crystal-crystal transition diagram	381
36.6. The triple point	383
36.7. Changes in internal energy in first-order phase transitions	383
36.8. Metastable states	385
36.9. Condensation. Supersaturated vapour	387
36.10. Boiling. Superheated liquid	388
36.11. The liquefaction of gases	391

Part five ELECTRODYNAMICS

Chapter 37 A FIELD OF FIXED CHARGES IN A VACUUM	393
37.1. Lines of force	393
37.2. Equipotential surfaces	394
37.3. The relation between field strength and potential	396
37.4. A dipole in an electric field	397
37.5. The parallel-plate capacitor	399

37.6. Capacitance	400
37.7. The energy of the electric field. Energy density	401
37.8. The force of interaction between capacitor plates	401
37.9. A conductor in an electric field	402
37.10. Determining the charge of the electron	404
Chapter 38 DIELECTRICS	406
38.1. An electric field with a dielectric	406
38.2. The polarization vector	407
38.3. Electric susceptibility	408
38.4. The field energy in a dielectric	409
38.5. Deformation polarizability	410
38.6. Orientational polarizability	411
Chapter 39 DIRECT CURRENT	414
39.1. Nonelectrostatic fields. Voltage and emf	414
39.2. Current and current density	416
39.3. Ohm's law for a uniform segment of a circuit	417
39.4. Resistance	418
39.5. Ohm's law in differential form	419
39.6. Ohm's law for a nonuniform segment of a circuit and for a closed circuit	419
39.7. The Joule-Lenz law	420
39.8. Charging and discharging capacitors	421
Chapter 40 A MAGNETIC FIELD IN A VACUUM	423
40.1. The interaction of currents. Magnetic forces	423
40.2. The transformation law for the sideways momentum and the sideways force	424
40.3. Interaction between moving charges	425
40.4. Magnetic induction. Lines of induction	427
40.5. The magnetic field of a current-carrying conductor	428
40.6. The magnetic moment	430
40.7. Magnetic field strength	433
40.8. The magnetic field of a solenoid	434
40.9. Invariance of the electric charge	434
Chapter 41 CHARGES AND CURRENTS IN A MAGNETIC FIELD	436
41.1. The Lorentz force	436
41.2. The motion of charged particles in a uniform magnetic field	437
41.3. Determining the sign of the charge of elementary particles	438
41.4. The cyclotron	440
41.5. Energy of the particle and the synchronization condition	442
41.6. Proton synchrotron	444
41.7. The electron charge-to-mass ratio	447
41.8. The ion charge-to-mass ratio	448
41.9. A current-carrying conductor in a magnetic field	449
41.10. A current-carrying loop in a magnetic field	450
Chapter 42 MAGNETIC MATERIALS	451
42.1. Three types of magnetic materials	451
42.2. The magnetic moment of the atom	453
42.3. Quantities characterizing the magnetic field in matter	454
42.4. Diamagnetism	455
42.5. Paramagnetism	458
42.6. Ferromagnetism. The Curie temperature	459
42.7. Hysteresis	461

42.8. The domain structure of ferromagnetic materials	463
42.9. The Einstein and de Haas experiment	466
42.10. The Stern-Gerlach experiment	467
42.11. Electron spin	469
42.12. Antiferromagnetism	470
Chapter 43 ELECTROMAGNETIC INDUCTION	472
43.1. Faraday's discovery	472
43.2. Electromagnetic induction and the Lorentz force	473
43.3. Induced electromotive force	474
43.4. The induction phenomenon in a stationary conductor	475
43.5. Strength of an induced field	476
43.6. The electromagnetic field and the relativity principle	477
43.7. Faraday's law of induction	477
43.8. Lenz's law	479
43.9. Electromagnetic induction and the law of conservation of energy	479
43.10. Self-induction	480
43.11. Energy of an electromagnetic field	481
43.12. Closing a circuit with inductance	483
Chapter 44 ELECTRICAL CONDUCTION IN SOLIDS	484
44.1. Experimental basis for the electron theory of conduction in metals	484
44.2. The Hall effect	486
44.3. Electron gas	489
44.4. Derivation of Ohm's law from electronic theory	490
44.5. The conductivity of metals and semiconductors	492
44.6. Derivation of the Joule-Lenz law	493
44.7. Contact potential difference	495
44.8. Thermoelectricity	496
44.9. The work function	498
Chapter 45 HEAT CAPACITY AND THERMAL CONDUCTIVITY OF SOLIDS	500
45.1. Heat capacity	500
45.2. Heat capacity of metals	502
45.3. Thermal conductivity of insulators	503
45.4. Thermal conductivity of metals	506
Chapter 46 ELECTRICAL CONDUCTIVITY OF ELECTROLYTES	508
46.1. Electrolytic dissociation	508
46.2. Ohm's law and the electrical conductivity of electrolytes	509
46.3. Faraday's laws	511
46.4. The galvanic cell	512
Chapter 47 CURRENT IN A VACUUM	513
47.1. Thermionic emission	513
47.2. The diode and its characteristics	515
47.3. The triode and its characteristics	517
47.4. The cathode-ray tube	518
Chapter 48 CURRENT IN GASES	519
48.1. Ionization and recombination	519
48.2. Nonself-maintaining discharges	520

48.3. Collision ionization	523
48.4. The Geiger-Müller counter	524
48.5. Self-maintaining discharges. Plasma	526
48.6. Glow discharges	527
48.7. Plasma in a magnetic field	529
48.8. Pinching and confining a plasma	532
48.9. The magnetohydrodynamic generator	534
Index	536

FOREWORD

Physics in our time is the leading natural science. This is due, primarily, to the wealth and diversity of the ideas and methods of investigation of modern physics, and to the fundamental significance of these ideas and techniques for epistemology (the theory of knowledge) and for the development of the world outlook given by dialectical materialism. No less important is the influence of physics on the development of the productive forces of society. Many fields of modern engineering, such as electronics (including semiconductor and quantum electronics), nuclear engineering, rocket design, and radio engineering are so closely interwoven with physics that they have become inseparable. At the same time the application of new physical ideas and research techniques to long established classical branches of science and engineering has led in many cases to essentially new solutions to certain problems.

New boundary sciences have been founded and are being vigorously developed. They are located at the junctions of several other sciences and are based on physics (e.g. cybernetics, radio-astronomy, bionics, biophysics, etc.). Opportunities have been substantially extended for a further interpenetration of various branches of knowledge. Typical of our times, moreover, is an enormous reduction in the time that elapses between scientific discoveries and their application to the national economy, everyday life and regular practice.

These and other considerations inevitably led to much higher requirements for modern physics courses, both for secondary school and higher (college) education. In the Soviet Union a new physics syllabus has been introduced for secondary schools. New textbooks and teacher's aids have been written. Elective courses are offered in the upper grades of secondary schools. Their aim is to acquaint interested students with some of the problems of up-to-date physics. In specialized classes and in schools with a bias towards physics and mathematics, physics is studied according to syllabi drawn up on a higher level than previously. In recent years, the scientific and the-

oretical level of physics teaching has been considerably raised in universities and especially in engineering institutes.

This higher level of physics teaching has posed many new methodological problems. The most complex of these is the writing of suitable textbooks and teacher's aids. This involves a task which we formulate (without any claims as to completeness or strictness) as follows: it is necessary to expound the basic ideas of modern physics, their physical meaning, investigation methods, and the most important results of research without resorting to a complicated mathematical apparatus. It is necessary, furthermore, to clearly demonstrate the deep-rooted continuity of modern and classical physics, and to establish the permanent value and significance of the "classics", which must be mastered before we can gain an insight into the scope of the ideas of modern physics, understand its techniques, and interpret its achievements.

Years of experience in the teaching of physics in institutions for secondary and higher education have led the authors to the conviction that a textbook could be written to meet the new requirements only by rejecting the traditional sequence of presenting the topics. The basic methodological scheme followed in writing this textbook was *to reveal, from the very beginning, the inherent bond between classical and modern physics.*

This led, among other things, to the necessity of carrying the ideas of the special theory of relativity through all the chapters devoted to mechanics and electrodynamics, and of using them, wherever possible, in moulding the basic concepts. (The methodological ideas for expounding fundamental concepts of the theory of relativity and its numerous applications in this textbook were proposed by A. Pinsky.) Some of the most important ideas of quantum mechanics are also employed throughout the physics course.

The authors encountered considerable methodological difficulties in realizing their project. Chief among these was the introduction, from the very beginning of the systematic course, of concepts that are explained by stages all through the course. But this difficulty exists as well in the traditional structure of the course.

One common drawback, moreover, of the traditional sequence of presentation, in which the concepts and ideas of modern physics are given at the end of the course, is that there is neither space in the textbook nor classroom time left for them. Besides, it is difficult to fix these ideas and concepts because the main part of the instructional matter is already far in the past and it is inexpedient to return to it for a "second round". Finally, the traditional physics teaching techniques develop a certain psychological barrier in the student's mind that impedes his reception of the ideas of modern physics. It seems to him that these ideas contradict the common-sense approach so typical of traditional physics teaching practice.

This book is intended for students who have thoroughly learned the material of the physics course for the incomplete secondary (eight-year) school, and wish to master the fundamentals of classical and modern physics. We hope that it will also be useful to the students of the upper classes of secondary schools, both in their study of the regular physics course and, especially, of optional subjects.

Students of higher educational institutions can undoubtedly gain definite benefits from working on this textbook.

It is, however, wrong to suppose that the absence of higher mathematics in a physics book makes it easy to read and master. Physics is a complex and sufficiently difficult subject, and its study requires serious efforts. Hence, the study of any book, by means of which physics can be learned, requires time and thoughtful, intense brain work. It cannot be simply read; it must be worked over.

The authors hope that the time and efforts spent by students in studying the fundamentals of physics from this textbook will be repaid in their further studies and in their occupation.

Moscow, USSR

*Boris M. Yavorsky
and Arkady A. Pinsky*

PART ONE

MOTION AND FORCES



Chapter 1

VELOCITY

1.1 MECHANICAL MOTION

1. All bodies that surround us, from the stars and planets to such minute particles as atoms and their components, are in a state of continuous motion. The simplest form of motion is the displacement of bodies with respect to one another. It is called *mechanical motion*.

The operation of most mechanisms, machines, and all kinds of transportation are based on mechanical motion. It is, moreover, a component of more complex, nonmechanical processes. Heat phenomena, for instance, are associated with the chaotic motion of molecules; the radiation of light, with the motion of electrons in atoms; and nuclear reactions, with the motion and interaction of elementary particles (protons, neutrons, and mesons). Such examples are legion.

2. To describe the motion of a body it is necessary to specify the motion of all points of the body.

One kind is *translational motion*. Here all the points of the body travel in an identical manner, i.e. a straight line connecting any two points is displaced parallel to itself (Fig. 1.1). It is evidently sufficient in describing the translational motion of a body to specify the motion of one of its points.

Another simple form of motion is *rotational motion*, in which all points of the body describe circles lying in parallel planes. The centres of all the circles lie on a single straight line called the rotational axis (Fig. 1.2).

3. In solving many problems in mechanics it is found to be expedient to disregard the size of a body and to treat it as if it were a material point, particle, or mass point. A *mass point* is a body whose dimensions can be neglected in the given problem. Natu-

rally, this concept is an abstraction: there are no material points in nature. But the formulation of many problems in mechanics is such that this abstraction can be made use of to advantage.

In fact, if a passenger wishes to find out how long his air liner will take to fly from Moscow to Novosibirsk, he need not concern himself with the type of motion of the various parts of the liner. But the dimensions and shape of the air liner cannot be disregarded in

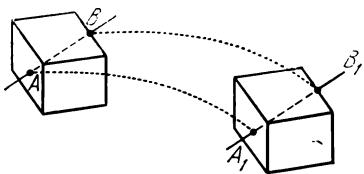


Fig. 1.1

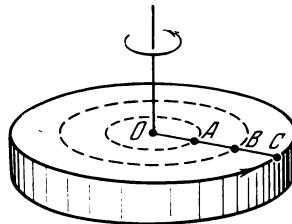


Fig. 1.2

studying such phenomena as take-off, landing, air resistance, etc. In a similar way we can treat the earth and other planets as particles if we are interested in the nature of their motion about the sun. If, however, our task is to find the reasons for the alteration of day and night or of the seasons of the year, we can no longer treat the earth as a particle and must take into account its size, revolution about its axis, the inclination of this axis to the plane of its orbit and many other factors.

Thus, in some problems a body can be dealt with as a particle without extent, while in other problems it cannot.

1.2. FRAMES OF REFERENCE. PATHS

1. When we deal with phenomena occurring near the earth's surface, we find that directions in space are not equivalent. A body released from the hand always falls vertically (approximately toward the centre of the earth); the free surface of a liquid lies in a horizontal plane; an initial velocity must be imparted to a body to make it rise vertically but none is required to make it fall, etc. This lack of equivalence of different directions is due to the fact that the earth attracts other bodies. At a vast distance from the earth, from other planets, and from the stars, i.e. *in space free of large bodies*, we would find that *all directions are equivalent*. We say that *free space is isotropic*, i.e. it has no select directions possessing special properties.

In exactly the same way *all points in space are equivalent* if there are no large bodies, such as planets or stars, near them. Consequently, we say that *free space is uniform*, i.e. it has no points possessing special properties.

2. *Time is also uniform.* This means that any event, occurring under the same conditions but at different moments of time, proceeds in exactly the same way. If a small ball drops today from a height of 6 m in 1.1 s, it must have taken the same time to drop from the same height in the same laboratory a month ago or a year ago, and its fall will require the same time a thousand years from now.

3. As we shall see further on, a number of important consequences follow from the fact that time is uniform and that space is uniform and isotropic (see Ch. 23). One can be mentioned here. Since space is uniform (all its points are equivalent) and isotropic (all directions in it are equivalent), it is impossible to determine the position of a particle with respect to space.

But it is quite possible to determine the position of one body with respect to another. The position of a lamp in a room, for instance, is fully specified by its distance from the floor and from two mutually perpendicular walls. Three such numbers determine the position of any other body, either inside or outside the room.

The *frame of reference* is a body or group of bodies which is regarded as fixed for a given problem and with respect to which the positions of all the other bodies are specified.

In principle any body can serve as a reference frame but not all reference frames turn out to be equally convenient. The motion of an automobile, for instance, can be more conveniently dealt with in a reference frame linked to the earth, and not one linked to the sun or moon. The motion of the planets, on the contrary, is most conveniently treated in a reference frame linked precisely to the sun and not to the earth or another planet. This will enable the laws of planetary motion to be more simply stated. Certain criteria for the choice of frames of reference will be discussed below.

A reference frame is usually associated with three mutually perpendicular straight lines, the *axes of coordinates* (Fig. 1.3). The position of a point is specified by three coordinates: the *x*-coordinate (abscissa), *y*-coordinate (ordinate), and the *z*-coordinate.

4. The line described by a moving particle in a given reference frame is called the *path*, or *trajectory*. If we light one end of a stick and then wave it in the air, describing various figures, we will clearly see, especially in a dark room, the path of motion of the glowing end.

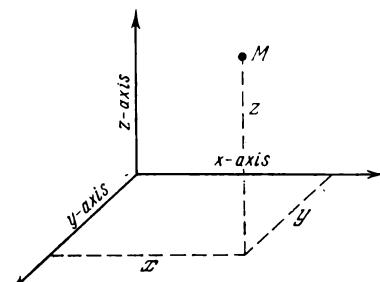


Fig. 1.3

The shape of the path depends upon the chosen reference frame. Let us consider a body dropped in a train that is in motion with respect to the earth. With respect to the railway car the path of the body is a straight line; with respect to the earth, it is a curve (a parabola if there is no air resistance). This is likewise true of any point on the propeller of an airplane in flight. In a system of coordinates linked to the airplane the point describes a circle; in one linked to the earth, the point describes a helix.

Hence, we can speak of the shape of a path only in the relative sense. There is no such thing as the shape of a path in general. We can only speak of the shape of a path, or trajectory, in a given reference frame (coordinate system).

1.3. RECTILINEAR MOTION. MOTION EQUATIONS

1. Let us consider a particle travelling in a straight line in a certain reference frame. For convenience we can direct one of the coordinate axes, for example, the x -axis, along the path of the particle. Then at each instant of time the particle will have a quite definite coordinate. This means that the coordinate of the moving particle is a function of time: $x = f(t)$. The form of this function is the *equation of motion* of the given particle.

2. To determine motion equations experimentally we must make marks in some way on the coordinate axis.

Let us consider, for instance, the motion of an escalator in a subway station. We could conduct an experiment by stretching a paper strip along the wall and having the experimenter hold an instrument that would make marks on the strip after equal intervals of time (for example, after each second). The results of our experiment could be tabulated as follows:

Instant of time t , s	1	2	3	4	5	6	7	8
Coordinate x , m	1.75	2.50	3.25	4.00	4.75	5.50	6.25	7.00

This experiment does not give us reliable information about the points in the intermediate instants of time. We can, however, reason as follows. First we try to find a function $x = f(t)$ for which the substituted values $t = 1, 2, 3$, etc. correspond exactly to the measured values $x = 1.75, 2.50, 3.25$, etc. It can readily be seen that this is the linear function $x = 1 + 0.75 t$. Assuming that the character of the motion does not change in the process of our investigation, we can further assume that the x -coordinate can be calculated by the same equation for all instants of time within the limits we are interested in (from the first to the eighth second).

We may note here that the shorter the interval of time between two consecutive marks the more sure we can be that our equation of motion is close to the truth. Hence, in any experimental investigation of the motion of a body the volume of information available plays a vital role. The more the marks we have and the shorter the time intervals between successive marks, the more accurately we can determine the motion equations.

3. The assumption that the character of motion does not change in the course of the experiment is a matter of principle. Suppose that the experimenter who had been standing on a step of the escalator suddenly decided at the end of the tenth second to speed up his descent and began walking down the steps. Evidently, the character of his motion is essentially changed and we have no right to use the motion equation we found for instants of time after the tenth second (from the beginning of the experiment).

It should be clear then that the experimentally determined motion equation can be employed only for the interval of time during which the investigation was conducted. Extrapolation, i.e. the application of the equation for a longer time interval, is permissible only after supplementary investigations to establish the fact that the character of motion remained unchanged.

1.4. UNIFORM MOTION

1. The motion of a particle is *uniform* if its coordinate is a linear function of time:

$$x = vt + b \quad (1.1)$$

where v and b are certain constants whose physical meaning will be discussed further on.

All nonuniform motion is said to be *variable* motion.

2. The motion equation can be most informatively illustrated by plotting a *graph*. This is done by constructing a coordinate system in a plane, plotting the time t on the axis of abscissas and the variable coordinate of the particle (for example, x) on the axis of ordinates. Plotting the values of the variables taken from the table along the two axes, we then construct perpendiculars to the axes at these points. The intersections of these perpendiculars constitute a series of points. When a smooth line is drawn through these points we obtain a *graph of the motion*. Since in uniform motion the coordinate is a linear function of time, the graph is observed to be a straight line (Fig. 1.4).

3. Setting $t = 0$ in the equation of motion, we obtain $x_0 = b$. This is the *initial coordinate*, i.e. the distance from the moving particle to the origin of coordinates at the initial instant of time.

Especial attention should be paid to the meaning of the *initial instant of time*. It is by no means the beginning of the motion, but

is the instant of time we started the stopwatch, i.e. the *beginning of the experiment for investigating the given motion*.

As concerns the instant the motion began, it is frequently of no interest to us, and is sometimes impossible to establish. An astronomer, for instance, investigating the motion of a meteorite, first discovers it when it has entered the comparatively dense layers of the atmosphere and has heated up and begun to glow. This then is the initial instant, the beginning of observation,

but it is obviously not the beginning of motion of the meteorite. The same is true for the motion of celestial bodies (planets and comets), the travel of cosmic particles through measuring apparatus, the travel of an airplane or rocket when detected by a radar set, etc.

1.5. VARIABLE MOTION

1. We shall deal variable motion in two examples: bodies falling from a certain height and bodies thrown vertically upward. In both cases we have rectilinear motion. We can readily establish, however, that we are dealing with nonuniform, i.e. variable, motion.

Let us drop a small and sufficiently heavy ball along the z -axis. The marks on the coordinate axis can be obtained, for instance, by snapshot photography, opening the shutter of the camera after equal intervals of time. Then we tabulate the results:

Instant of time t , s	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Coordinate z , m	0.049	0.196	0.431	0.784	1.22	1.76	2.40	3.14	3.97	4.90

Upon plotting the graph of this motion we find that it is not a straight line but a parabola, and that the equation of motion has the form $z = 4.9 t^2$ (Fig. 1.5).

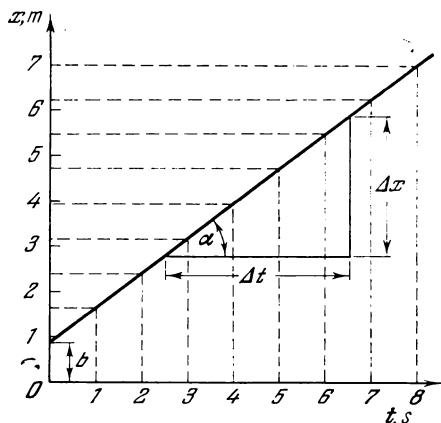


Fig. 1.4

2. If we use the same method to investigate the motion of a body thrown vertically upward, we shall obtain the following table of data:

Instant of time t , s	0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0
Coordinate z , m	0	1.74	3.14	4.12	4.70	4.90	4.70	4.12	3.14	1.74	0

This is also rectilinear variable motion. The motion equation is of the form $z = 9.8 t - 4.9 t^2$. The graph of the motion equation is also a parabola (Fig. 1.6).

3. Note should be made of the essential difference between the graph of motion and the path. The path is the curve along which a particle travels in the reference frame; the graph is the curve passing through representative points in the coordinate-time plane.

A ball falling from a certain height and a ball thrown vertically

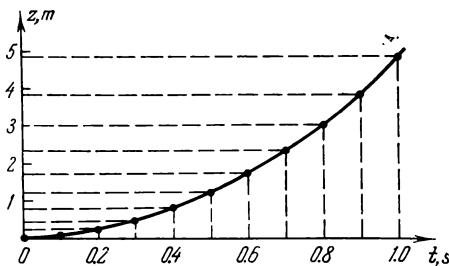


Fig. 1.5

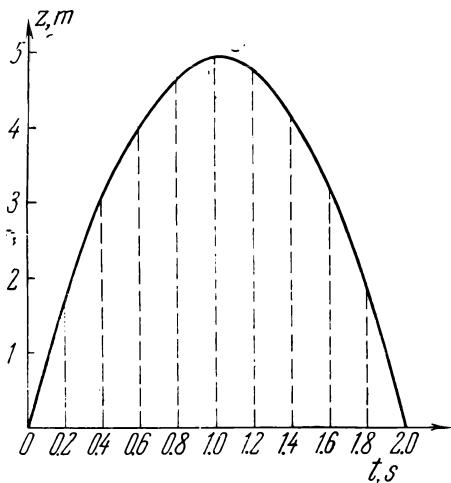


Fig. 1.6

upward travel along paths of the same shape, a straight line, but their graphs differ because their motions are not of the same nature.

1.6. AVERAGE VELOCITY

1. Consider the case in which a particle travels continuously in a single direction along one of the coordinate axes.

The distance $\Delta x = x_2 - x_1$ is called the *displacement* of a moving particle in the time interval $\Delta t = t_2 - t_1$.

Note that if the particle travels continuously in the same direction, the distance travelled is equal to the magnitude of the displacement: $\Delta s = |\Delta x|$. If, instead, the particle travels first in one direction and then stops and begins to travel in the opposite direction (as for

oscillating motion or the motion of a body thrown vertically upward), the distance travelled equals the sum of the magnitudes of the displacements in both directions: $\Delta s = |\Delta x_1| + |\Delta x_2| + \dots$.

As an example we can refer again to the motion of a body thrown vertically upward. We select the instants of time $t_1 = 0.8$ s and $t_2 = 1.4$ s. They correspond to the coordinates $x_1 = 4.70$ m and $x_2 = 4.12$ m (see the preceding table). Then the displacement is

$$\Delta x = x_2 - x_1 = 4.12 - 4.70 = -0.58 \text{ m}$$

The minus sign indicates that the displacement is in the direction opposite to that of the coordinate axis. The distance travelled is considerably greater than this displacement because the moving body reaches its maximum height $x_{\max} = 4.90$ m at the instant $t = 1.0$ s and then begins to fall. Therefore,

$$\begin{aligned}\Delta s &= |x_{\max} - x_1| + |x_2 - x_{\max}| = |4.90 - 4.70| + |4.12 - 4.90| \\ &= 0.20 + 0.78 = 0.98 \text{ m}\end{aligned}$$

2. The *average velocity* of motion in a given interval of time is the physical quantity numerically equal (i.e. equal in magnitude) to the ratio of the displacement to the time interval. Thus,

$$v_{av} = \frac{\Delta x}{\Delta t} = \frac{x_2 - x_1}{t_2 - t_1} \quad (1.2)$$

The average velocity is a vector. Its direction coincides with that of the displacement.

The unit of measurement of the velocity is defined as the velocity of a motion in which the displacement of a particle in unit time equals unit length. Thus

$$[v] = \frac{[x]}{[t]} \quad (1.3)$$

The square brackets indicate that we are dealing with the dimensions of the considered quantity.

In the International System of Units (abbreviated SI from the French: Système International d'Unités) the unit of displacement is the metre (m) and that of time is the second (s or sec). Therefore, the unit of velocity is m/s.

In addition, several nonsystem units are in use:

$$1 \text{ cm/s} = 10^{-2} \text{ m/s},$$

$$1 \text{ km/h} = \frac{1}{3.6} \text{ m/s} = 0.273 \text{ m/s},$$

$$1 \text{ knot} = 1 \text{ nautical mile per hour}$$

$$= 1852 \text{ m}/3600 \text{ s} = 0.514 \text{ m/s}$$

and others.

3. In variable motion the average velocity depends upon the choice of the interval of time during which the body travels. In the case of free fall, considered above, the average velocity in the time interval from $t_1 = 0.1$ s to $t_2 = 0.2$ s (see the table) is

$$v_{av} = \frac{0.196 - 0.049}{0.2 - 0.1} = \frac{0.147}{0.1} = 1.5 \text{ m/s}$$

In the time interval from $t_1 = 0.7$ s to $t_2 = 0.8$ s it is

$$v_{av} = \frac{3.14 - 2.40}{0.8 - 0.7} = \frac{0.74}{0.1} = 7.4 \text{ m/s}$$

and from $t_1 = 0.7$ s to $t_2 = 1.0$ s it is

$$v_{av} = \frac{4.90 - 2.40}{1.0 - 0.7} = \frac{2.50}{0.3} = 8.3 \text{ m/s}$$

The *average velocity is a constant value*, independent of the selected time interval, *only for uniform motion*. Here, $x_1 = vt_1 + b$, $x_2 = vt_2 + b$ and the average velocity is

$$v_{av} = \frac{\Delta x}{\Delta t} = \frac{x_2 - x_1}{t_2 - t_1} = \frac{vt_2 + b - vt_1 - b}{t_2 - t_1} = v$$

4. For practical purposes the average velocity is sometimes defined as the ratio of the distance travelled to the elapsed time the body is in motion. Note that the average velocity determined in this way is a scalar and not a vector quantity.

1.7. INSTANTANEOUS VELOCITY OF VARIABLE MOTION

1. Often we are not interested in the average velocity of a body during a definite interval of time but wish to find its velocity at a *given instant*, or *instantaneous velocity*. For example, if a body hits an obstacle, the force with which it acts on the obstacle depends upon its velocity at the instant of collision and not its average velocity; the range and the shape of the path of a projectile depends upon its initial and not its average velocity; etc.

The instantaneous velocity of a particle can be determined as follows. Let the coordinate of the particle be x at the instant of time t . At the instant $t_1 = t + \Delta t$ the coordinate will be $x_1 = x + \Delta x$, where Δx is the displacement. Then the average velocity is

$$v_{av} = \frac{x_2 - x_1}{t_2 - t_1} = \frac{\Delta x}{\Delta t} \quad (1.4)$$

Generally speaking, the longer the time interval Δt , the more the average velocity differs from the instantaneous value. On the contrary, the shorter the time interval Δt , the less the average velocity differs from the instantaneous velocity we are trying to find.

We can now define the *instantaneous velocity* v as the limiting value that the average velocity approaches for an infinitely small time interval. Thus

$$v = \lim_{\Delta t \rightarrow 0} v_{av} = \lim_{\Delta t \rightarrow 0} \frac{\Delta x}{\Delta t} \quad (1.5)$$

In mathematics such a limiting value is called a *derivative*. Consequently, the instantaneous velocity is the derivative of the coordinate with respect to time.

2. Since the average velocity of a particle in uniform motion is a constant and the limiting value of a constant is the constant itself, the *instantaneous velocity of a particle travelling with uniform motion is a constant*.

Hence, uniform motion can be defined as motion at constant velocity.

Consequently, the constant v in the equation of uniform motion $x = vt + b$ is the velocity of the motion and constant b is the initial coordinate.

As is evident from the graph in Fig. 1.4, the velocity of uniform motion is equal in magnitude to the tangent of the angle of inclination of the graph to the time axis:

$$v = \frac{\Delta x}{\Delta t} = K \times \tan \alpha$$

where K is the dimension factor which equals the ratio of the scales of the units of displacement (axis of ordinates) and of time (axis of abscissas).

Motion at a higher velocity corresponds to a graph with a steeper inclination.

Chapter 2 INERTIA

2.1. THE PRINCIPLE OF INERTIA

1. The basic aim of mechanics is the study of the motion of bodies in some reference frame and of the causes that determine the nature of this motion. It is necessary to find out under what conditions a body travels along a rectilinear or curvilinear path, with uniform or nonuniform motion, with acceleration or deceleration.

We know by experience that the interaction of bodies changes the nature of their motion. When, for instance, a falling body reaches the earth, it either stops (its motion ceases) or it bounces back upward, in which case its velocity is reversed in direction. If we consider a body at rest with respect to the earth, we know that it will never

begin moving by itself. The action of some other body is required to take it out of the state of rest so that it begins to move. A great many similar examples could be cited.

2. From the quite correct view that the interaction of bodies changes their velocity certain philosophers of antiquity, for instance Aristotle (384-322 B.C.), came to the false conclusion that the motion of bodies is the result of their interaction with other bodies.

One of the reasons for this erroneous belief was the fact that they regarded the earth as the absolutely fixed centre of the universe. Consequently, they considered a body to be in its natural state when it was at rest with respect to the earth. They believed motion to be a forced, temporary state of a body due to the action of external forces. It is true that they ran into certain difficulties in trying to explain the motion of a stone that has been thrown or an arrow flying out of a bow. Here there were no bodies that continuously pushed the stone or arrow in their flight. Then to explain these phenomena the ancient scientists proposed various hypotheses with the purpose of saving their basic assumption about the forced nature of motion.

3. At the end of the 14th century two problems again raised the question of motion. Firstly, the development of artillery made it necessary to find the laws of motion of cannon balls. Then, in connection with the heliocentric theory of the solar system, proposed by Nicolaus Copernicus (1473-1543), it became clear that the earth was not the centre of the universe, but only an ordinary planet, revolving with the other planets about the sun. The inevitable conclusion was that the planets (and, likewise, other bodies) are in motion by themselves, because it was entirely incredible to have the earth and other planets being pushed continuously by something for billions of years.

Moreover, it was necessary to explain why we do not sense the revolution of the earth and also why, when we jump upward, leaving the surface of the earth, we land again in the same spot.

The first correct, though incomplete, solution of the problem of motion was put forth by Galileo Galilei (1564-1642) at the beginning of the 17th century. Fifty years more passed before Sir Isaac Newton (1643-1727) clearly formulated his three basic laws of motion that became the foundation of classical mechanics.

4. To properly solve the problem of motion it was necessary first of all to disengage oneself from all ideas of external action and put the problem as follows: what will happen to a body if it ceases to interact with other bodies? We can proceed further by means of a *thought experiment* whose idea was proposed by Galileo. Evidently, we cannot free a body of the influence of the forces of gravity and friction in any real experiment conducted on the earth. We can, however, ask ourselves: what will happen if we begin, at least in our imagination, to reduce these forces?

Assume that a cart travels in loose sand; it quickly slows down and stops. On a horizontal glass surface it travels considerably farther but will, nevertheless, gradually slow down and stop. What happens if the glass is absolutely smooth? Evidently, the cart will not stop but will continue to move for an indefinitely long time.

It was this thought experiment involving a body free of all external influences that enabled Galileo to conceive the idea of inertial motion of a body. Even though Galileo was mistaken in thinking that a body can move by inertia not only uniformly in a straight line but uniformly in a circle as well (the latter is wrong!), his idea enabled Newton to correctly formulate the *law of inertia*. This law can be stated as follows:

If a given body does not interact with surrounding bodies, the velocity of its motion does not change in either magnitude or direction, i.e. it continues to move uniformly in a straight line.

This is also known as *Newton's first law of motion*.

Inertial motion is motion along the shortest distance because in free space the shortest distance between two points is a straight line.

If we state that a body is at rest we do not imply an absolute state of rest but only rest in the given reference frame, which is in motion with respect to other bodies.

5. It is precisely due to inertia that we land in the same spot when we jump upward in a moving train. When we stand on the floor of a railway car in uniform motion we have the same velocity as the car. Jumping vertically upward we retain our horizontal velocity unchanged because nothing acts on us in this direction. Hence, after travelling in the horizontal direction together with the car, we return to the same spot we left at the instant we jumped upward. Exactly the same can be said of our motion with respect to the earth.

Thrown bodies travel by inertia. In throwing a stone we impart a certain velocity to it. If the stone was not subject to air resistance or to the attraction of the earth, it would continue to travel, maintaining this velocity unchanged in magnitude and direction. The interaction of the stone with the air and earth slows down its motion, bends its path and, finally, stops it.

2.2. INERTIAL REFERENCE FRAMES

1. The statement of the law of inertia given in the preceding section is certainly incomplete. The motion of bodies was mentioned there but nothing was said about the reference frame in which the motion takes place. We already know that we can speak of the shape of a path, as well as the velocity, only with respect to some definite reference frame because a path that is rectilinear in one reference frame may turn out to be curvilinear in some other frame.

Hence it is necessary to change the statement of the law of inertia, explicitly specifying the reference frame in which the given motion is being treated. The law of inertia is more correctly stated as follows:

There are frames of reference with respect to which all bodies which do not interact with other bodies move uniformly in a straight line. Reference frames satisfying this condition are called *inertial reference frames*.

2. The concept of an inertial reference frame is an abstraction and is found in nature only with a certain degree of approximation. All reference frames are obviously associated with a definite body and all bodies in nature interact to some extent, so that it is impossible to point to strictly inertial reference frames. We can therefore only indicate reference frames that can be regarded as inertial in dealing with a given range of problems. Of course, any questions as to whether a given reference frame is inertial or not can be answered only by conducting an experiment to find out.

Experiment shows that as a first approximation reference frames fixed on the surface of the earth can be regarded as inertial. This means, in essence, that there is a whole range of phenomena which is independent of the earth's rotation. Such a reference frame is said to be *geocentric* (from the Greek *gē* meaning the earth).

The earth's rotation has practically no effect on the operation of all kinds of mechanisms and machinery and on the travel of various kinds of transportation facilities. It does not affect the course of thermal, chemical, and nuclear reactions, as well as the electromagnetic processes occurring in generators, motors, transformers, radio transmitters and receivers. Nor does it influence the propagation of electromagnetic waves or light and sound phenomena. Hence, the geocentric reference frame can be regarded as inertial to a high degree of accuracy in describing these physical processes.

3. There are many phenomena, however, caused by the rotation of the earth on its axis and its revolution about the sun. Since these motions are neither rectilinear nor uniform, they exhibit effects that are impossible in inertial reference frames.

Thus, when we observe the sky at night, we find that the whole system of stars seems to revolve about an axis passing through the North Star and the centre of the earth. If we use the geocentric reference frame, we must search for some mechanism of interaction between the earth and all the stars that could explain why their paths are curvilinear. Next we observe that the planets have quite complex motions in the sky. Moving in some direction, a planet suddenly stops, moves in the reverse direction and, after describing a loop, begins to move in the initial direction again.

The numerous attempts made by Aristotle, Ptolemy (2nd century A.D.) and other investigators to explain the observed motion of celestial bodies in the geocentric reference frame proved unsuccessful. The correct theory of the motion of celestial bodies developed by

Copernicus, Galileo, and Johannes Kepler (1571-1630) when they linked the reference frame with the sun rather than the earth. This reference frame is called *heliocentric* (from the Greek *hēlios* meaning the sun). It is inertial to a considerably greater degree than is the geocentric reference frame.

2.3. THE PRINCIPLE OF RELATIVITY

1. There is an infinitely great number of inertial reference frames with the degree of approximation mentioned above. *Any reference frame moving with uniform velocity in a straight line with respect to a certain inertial reference frame is also inertial.*

Let us assume that a body moves by inertia in a certain reference frame. Consequently, its velocity changes neither in magnitude nor direction. In another reference frame, moving with uniform velocity in a straight line with respect to the first frame, our body will have a different velocity, but which again is constant in magnitude and direction. Consequently, the second reference frame is inertial to the same degree as the first.

Hence, reference frames fixed in a railway car or ship which travels at uniform velocity in a straight line with respect to the earth are as inertial as a reference frame fixed relative to the earth.

2. Experience shows that on a ship travelling at uniform velocity in a straight line with respect to the earth, it is equally easy to move in any direction as it is on the earth. If we shoot a revolver, the bullet requires the same time to fly from the bow to the stern as it does from the stern to the bow. A body released by the hand falls vertically downward (with respect to the ship, of course; with respect to the earth it has a parabolic path). Water poured into a vessel has a horizontal free surface, as it has on the earth.

It follows from these and many other phenomena that no mechanical experiment will allow us to single out any one preferred reference frame of the many available. It follows further that there can be no question of the absolute rest or absolute motion of bodies but only of their relative motion in some inertial reference frame.

This basic law of nature is called *Galileo's relativity principle*:

The laws of mechanics are equally valid in all inertial reference frames; or, more concisely, all inertial reference frames are equally justified.

An analysis of the relativity principle, carried out by Albert Einstein (1879-1955), led him in 1905 to the conclusion that this principle is one of the most fundamental of nature's laws and one that is applicable to any other phenomena—thermal, electromagnetic, optical, etc.—and not only mechanical phenomena. This enabled him to formulate his general principle of relativity which is often called *Einstein's principle of relativity*:

All the laws of nature are the same in all inertial reference frames.

This principle is one of the foundations of the world outlook given by modern physics.

2.4. GALILEAN TRANSFORMATIONS

1. In Newtonian mechanics we deal with relatively low velocities. The highest velocities at which macroscopic bodies travel are the velocity of planets, meteors, and spaceships. They do not exceed several dozens of kilometres per second, which is very much less than the velocity of light $c = 300\,000$ km/s.

If the coordinates of a certain particle are specified in one inertial reference frame, we can calculate the coordinates of the same particle

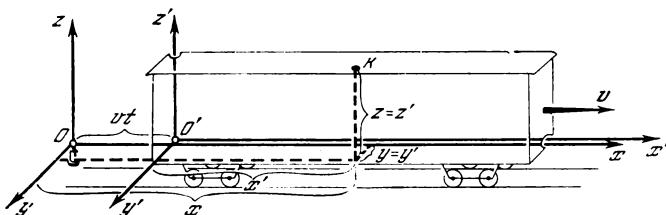


Fig. 2.1

in any other inertial reference frame moving at uniform velocity in a straight line with respect to the first frame. The exact form of the coordinate transformations will be derived in Sec. 12.7. Here we shall consider the so-called *Galilean transformations*, which enable us to find the relations between the coordinates of a certain particle in two different inertial reference frames at velocities of motion much less than the velocity of light.

2. For the sake of definiteness we shall link one reference frame with the earth and the other with a railway car that is travelling with uniform motion in a straight line at the velocity v with respect to the earth. To simplify the problem we shall arrange one of the coordinate axes (for instance, the axis of abscissas) along the direction of travel of the car (Fig. 2.1). It is evident from the drawing that the y - and z -coordinates of point K coincide in the two systems of coordinates (in the two reference frames). Thus $z' = z$ and $y' = y$. The abscissas, however, differ by the length $OO' = vt$. Hence the Galilean transformations are of the form

$$z' = z, \quad y' = y \quad \text{and} \quad x' = x - vt \quad (2.1)$$

3. To check whether the Galilean transformations satisfy the relativity principle we shall find their form for the case when the

railway car is assumed to be fixed and the earth to be travelling in the opposite direction at the velocity V . The magnitudes of velocities V and v coincide but they are of opposite direction, hence $V = -v$. Reasoning, identical to that above, leads to the equations:

$$z = z', \quad y = y' \quad \text{and} \quad x = x' + vt \quad (2.2)$$

Thus, we find the Galilean transformations to be reciprocal: any of the inertial reference frames can be taken conditionally as the fixed one, and transformation to any other inertial frame is carried

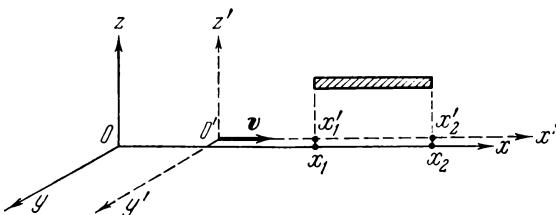


Fig. 2.2

out by the same equations, taking into account the sign of the relative velocity of the frame. And this demonstrates that these transformations agree with the relativity principle.

4. The displacement of a particle differs in different reference frames. Let us suppose that at the instant t_1 the particle had the coordinate x_1 , and at the instant t_2 it had the coordinate x_2 . Then its displacement in a reference frame fixed on the earth is

$$\Delta x = x_2 - x_1$$

The displacement of the particle in a reference frame linked to the railway car is

$$\Delta x' = x'_2 - x'_1 = x_2 - x_1 - v(t_2 - t_1) = \Delta x - v\Delta t \quad (2.3)$$

Accordingly,

$$\Delta x = \Delta x' + v\Delta t \quad (2.4)$$

5. But distances remain the same in the two reference frames. Evidently, *any distance or length* (Fig. 2.2) equals the difference in the coordinates for its end and beginning measured simultaneously. Thus $l = x_2 - x_1$ at $t_2 = t_1$, i.e. $\Delta t = 0$. Then from equations (2.3) and (2.4) it follows that

$$l = l' \quad (2.5)$$

This result would seem to be entirely obvious and it is not clear, on the face of it, why it had to be derived. The meaning of this derivation will become clear further on (in Sec. 12.8).

2.5. CLASSICAL LAW OF THE ADDITION OF VELOCITIES

1. Suppose a particle travels uniformly along the axis of abscissas at the velocity $u = \Delta x / \Delta t$. We wish to find the velocity of the same point in another inertial reference frame: $u' = \Delta x' / \Delta t$. We can use equation (2.3). Dividing both sides of the equation by Δt , we obtain

$$u' = u - v \quad (2.6)$$

or, correspondingly,

$$u = u' + v \quad (2.7)$$

This is the *classical law of addition of velocities* for the special case being considered.

2. The classical law of addition of velocities holds well at velocities of motion much less than the velocity of light.

Thus, if the railway car is travelling with respect to the earth with a velocity of 80 km/h and a passenger is walking along the car in the same direction at a velocity of 6 km/h, the velocity of the passenger with respect to the earth is 86 km/h. If a river flows with a velocity of 2 m/s and the velocity of a steamboat with respect to the water is 8 m/s, the velocity of the steamboat with respect to the earth will be 6 m/s when it travels against the current and 10 m/s when it travels downstream.

3. The logical considerations underlying the Galilean transformations appear to be self-evident and their consequences, such as invariance of distances in different reference frames and the law of addition of velocities, held excellently in experiments at the velocities that scientists dealt with for nearly three centuries after Galileo and Newton. The idea gradually evolved that these transformations were valid for any physical phenomena. Only at the turn of the century was it discovered that the application of the ideas of Newton's mechanics to phenomena associated with the propagation of light led to a number of contradictions. We shall see (Ch. 12) how contradictions arose and how they were eliminated by the theory of relativity.

Chapter 3

SCALARS AND VECTORS. VELOCITY VECTOR

3.1. SCALAR QUANTITIES

1. A *scalar* is a quantity which in any reference frame corresponds to a definite number that depends on the system of units being used. In writing a scalar we set down its numerical value and unit of measurement next to it.

For example: $l = 3 \text{ m} = 300 \text{ cm} = 3000 \text{ mm}$.

The numerical value, or magnitude, of a scalar is inversely proportional to the unit of measurement. This can be represented symbolically as

$$a [A] = a' [A'] \text{ or } \frac{a}{a'} = \frac{[A']}{[A]} \quad (3.1)$$

where $[A]$ and $[A']$ are the dimensions of the scalar and a and a' are its numerical values.

2. Examples of scalar quantities are time, length, area, volume, temperature, mass, work, and energy.

The sum or product of scalars is also a scalar. In general, any algebraic operation involving scalars yields a scalar.

It should be noted that if a certain expression made up of scalar quantities follows the sign of a nonalgebraic (transcendental) function, it is an abstract number. Such functions include exponential, logarithmic, and trigonometric functions. For example, if coordinate x is expressed in metres ($[x] = m$), then the expression $y = a^{kx}$ is meaningful only if $[k] = m^{-1}$. Similarly, the expression $x = A \cos \omega t$ is meaningful only if $[\omega] = \text{s}^{-1}$.

3.2. VECTOR QUANTITIES

1. A *vector* is a quantity which in any reference frame corresponds to a *line segment of definite direction*. Hence, all vectors are characterized by their absolute value (magnitude), direction, and point of application of their tail. Examples of vector quantities are velocity, acceleration, and force.

A vector whose tail coincides with the origin of coordinates and whose head determines the position of a certain particle is called a *radius, or position, vector* of the particle (Fig. 3.1).

Vectors are usually represented in handwriting by a lower-case letter with an arrow put above it (\vec{a} , \vec{b} , etc.). In printing they are represented by boldface lower-case letters (**a**, **b**, . . .). A radius vector is commonly represented by the letter **r**.

2. Any vector can be projected onto the axes of coordinates.

The projections of vector **a** on the coordinate axes are denoted by a_x , a_y , and a_z , where the subindex indicates the axis onto which the vector is projected (Fig. 3.2). It can be seen in the drawing that

$$a_x = x_2 - x_1, \quad a_y = y_2 - y_1, \quad \text{and} \quad a_z = z_2 - z_1 \quad (3.2)$$

where x_1 , y_1 , and z_1 = coordinates of the tail of the vector
 x_2 , y_2 , and z_2 = coordinates of its head.

Projections of a radius vector **r** are coordinates of its head since the coordinates of its tail are zero. Thus

$$r_x = x, \quad r_y = y, \quad \text{and} \quad r_z = z \quad (3.3)$$

3. The *absolute value* of a vector, or its *magnitude*, is a scalar equal in length to the line segment representing the vector. It is designated

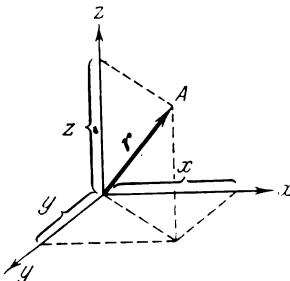


Fig. 3.1

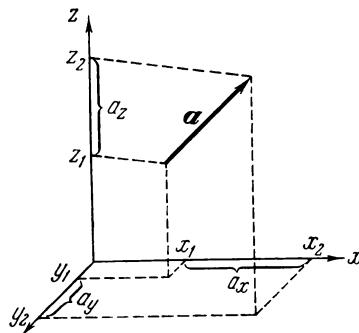


Fig. 3.2

by the symbol $| \mathbf{a} |$, or simply a . Applying the Pythagorean theorem in geometry, we obtain

$$\begin{aligned} |\mathbf{a}| &= a = \sqrt{a_x^2 + a_y^2 + a_z^2} \\ &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \end{aligned} \quad (3.4)$$

The magnitude of a radius vector is

$$|\mathbf{r}| = r = \sqrt{x^2 + y^2 + z^2} \quad (3.5)$$

3.3. CERTAIN OPERATIONS ON VECTORS

1. The sum of two vectors is a new vector each of whose projections is the sum of the respective projections of the vectors that are added. If the projections of vectors \mathbf{a} and \mathbf{b} are (a_x, a_y, a_z) and (b_x, b_y, b_z) , then it follows from the definition that

$$\mathbf{c} = \mathbf{a} + \mathbf{b} \text{ if } \begin{cases} c_x = a_x + b_x \\ c_y = a_y + b_y \\ c_z = a_z + b_z \end{cases} \quad (3.6)$$

This definition leads to many important consequences. The *commutative law of addition* is

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a} \quad (3.7)$$

We find that the commutative law is always valid for vectors. We can add them in any order and obtain the same result.

The *associative law of addition*

$$\mathbf{a} + \mathbf{b} + \mathbf{c} = (\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c}) \quad (3.8)$$

can be proved in the same way.

Polygon method of vector addition: to add several vectors, for instance \mathbf{a}_1 , \mathbf{a}_2 , \mathbf{a}_3 , \mathbf{a}_4 and \mathbf{a}_5 , the tail of the second vector is placed at the head of the first, then the tail of the third at the head of the second, etc. Finally, a new vector is constructed whose tail coincides with

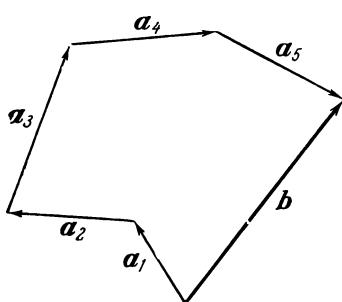


Fig. 3.3

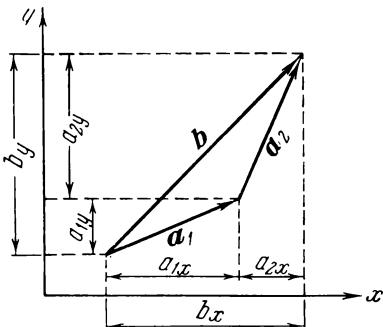


Fig. 3.4

the tail of the first vector and whose head coincides with the head of the last vector (Fig. 3.3). This closing side of the polygon is the resultant vector \mathbf{b} , which is the sum being sought. Thus

$$\mathbf{b} = \mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3 + \mathbf{a}_4 + \mathbf{a}_5$$

This can be proved by placing two vectors, \mathbf{a}_1 and \mathbf{a}_2 , in the xy -plane with the tail of vector \mathbf{a}_2 at the head of vector \mathbf{a}_1 (Fig. 3.4).

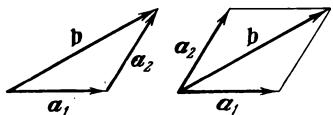


Fig. 3.5

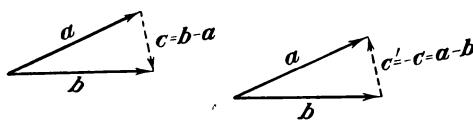


Fig. 3.6

Joining the tail of \mathbf{a}_1 with the head of \mathbf{a}_2 , we obtain vector \mathbf{b} . This, according to the definition, is the sum of the two vectors because, as is evident from the drawing, the projections of vector \mathbf{b} on the coordinate axes are the sums of the projections of the vectors \mathbf{a}_1 and \mathbf{a}_2 on the same axes.

If several vectors are to be added, the associative law of addition can be applied.

We should note that the sum of two vectors can be obtained by the *parallelogram method* (Fig. 3.5) as well. The polygon method proves more convenient, however, for adding more than two vectors.

2. The *difference* of two vectors is a vector constructed as follows: the tail of the minuend vector \mathbf{b} is placed at the tail of the subtrahend vector \mathbf{a} (Fig. 3.6). Then vector \mathbf{c} , joining the head of the subtrahend

vector to the head of the minuend vector, is the difference being sought. Thus

$$\mathbf{c} = \mathbf{b} - \mathbf{a} \quad (3.9)$$

As to the projections of the vectors

$$c_x = b_x - a_x, c_y = b_y - a_y, \text{ and } c_z = b_z - a_z$$

The proof follows directly from the polygon method since $\mathbf{a} + \mathbf{c} = \mathbf{b}$.

3. The product of vector \mathbf{a} multiplied by scalar k is a new vector \mathbf{b} whose projections are k times as long as the corresponding projections of vector \mathbf{a} . It follows from this definition that

$$\mathbf{b} = k\mathbf{a} \quad \text{if} \quad \begin{cases} b_x = ka_x \\ b_y = ka_y \\ b_z = ka_z \end{cases} \quad (3.10)$$

For the case when k is a positive integer this follows directly from the law of vector addition.

If a vector is multiplied by a scalar, its direction remains the same when the scalar is positive and is the opposite when the scalar is negative. The magnitude of the new vector is k times as large (Fig. 3.7):

$$\begin{aligned} b &= \sqrt{b_x^2 + b_y^2 + b_z^2} = \sqrt{k^2 a_x^2 + k^2 a_y^2 + k^2 a_z^2} \\ &= k \sqrt{a_x^2 + a_y^2 + a_z^2} = ka \end{aligned}$$

4. The quotient obtained by dividing vector \mathbf{a} by scalar $k \neq 0$ is a new vector \mathbf{b} whose projections are $1/k$ as long as the corresponding projections of vector \mathbf{a} . Hence, from the definition, it follows that

$$\mathbf{b} = \frac{\mathbf{a}}{k} \quad \text{if} \quad b_x = \frac{a_x}{k}, \quad b_y = \frac{a_y}{k}, \quad \text{and} \quad b_z = \frac{a_z}{k} \quad (3.11)$$

It will be left as a problem for the student to show that if a vector is divided by a scalar its direction remains the same while its magnitude is reduced to $1/k$ of its previous value.

3.4. RESOLVING A VECTOR INTO TWO COMPONENTS

1. To resolve vector \mathbf{a} into two components means to find two vectors, \mathbf{b} and \mathbf{c} , for which $\mathbf{a} = \mathbf{b} + \mathbf{c}$. Here \mathbf{b} , \mathbf{c} , and \mathbf{a} form a *closed triangle* (polygon method of vector addition).

The problem as stated above is indefinite because an infinite number of triangles can be constructed on a line of length a ; the other

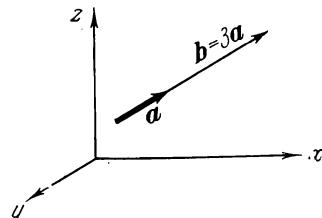


Fig. 3.7

two sides of the triangle being equal to the magnitudes of the vectors being sought (Fig. 3.8). The problem becomes definite only after certain additional conditions are specified.

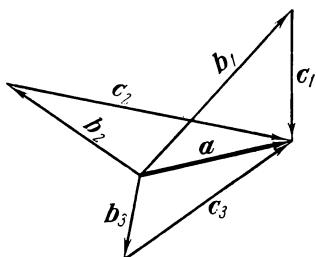


Fig. 3.8

2. A problem that is frequently met with in practice involves the resolution of a given vector \mathbf{a} into two components whose directions are parallel to certain straight lines. This problem is quite definite.

Through the ends of the line segment \mathbf{a} we draw two straight lines parallel to the given straight lines (Fig. 3.9). In the triangle thus obtained we construct the two vectors so

that the tail of the first is at the tail of vector \mathbf{a} and its head is at the vertex of the triangle opposite \mathbf{a} . The tail of the second vector is at the vertex of the triangle and its head is at the head of \mathbf{a} . These will

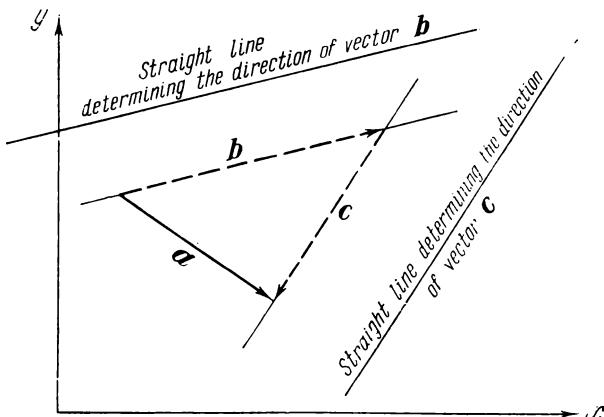


Fig. 3.9

be the two required vectors, \mathbf{b} and \mathbf{c} . They are parallel, by construction, to the given straight lines and their sum, according to the polygon method, is equal to the initial vector: $\mathbf{b} + \mathbf{c} = \mathbf{a}$.

3.5. VELOCITY IS A VECTOR

1. Let a particle travel along a rectilinear path that does not coincide with the coordinate axes. At the instant of time t_1 the position of the particle is described by position (radius) vector \mathbf{r}_1 and at

the instant t_2 by position vector \mathbf{r}_2 (Fig. 3.10). We determine the *displacement vector* as follows:

$$\Delta \mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1 \quad (3.12)$$

Then the average velocity is evidently the vector

$$\mathbf{v}_{av} = \frac{\Delta \mathbf{r}}{\Delta t} = \frac{\mathbf{r}_2 - \mathbf{r}_1}{t_2 - t_1} \quad (3.13)$$

As we know, the division of vector $\Delta \mathbf{r}$ by the scalar Δt results in a new vector whose direction coincides with that of the displacement

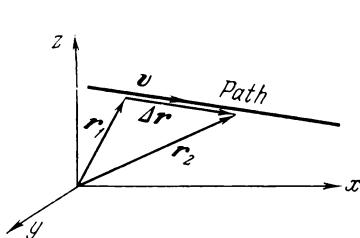


Fig. 3.10

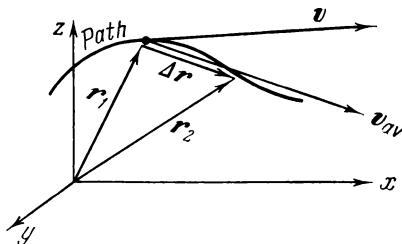


Fig. 3.11

vector. Hence, the velocity and displacement vectors have the same direction.

The instantaneous velocity

$$\mathbf{v} = \lim_{\Delta t \rightarrow 0} \frac{\Delta \mathbf{r}}{\Delta t} \quad (3.14)$$

is also a vector, coinciding in direction with the displacement vector.

Thus, the *velocity of a particle moving in a straight line is a vector in the direction along its path*.

2. The definitions of the average and instantaneous velocities (3.13 and 3.14) are also valid for the motion of a particle along a curvilinear path. Here the displacement vector and the average velocity vector are in the direction of the chord joining the two points of the path that correspond to the instants of time t_1 and t_2 (Fig. 3.11). For the case when $\Delta t \rightarrow 0$ the direction of the chord approaches that of the tangent. At the limit the infinitely small displacement vector and the instantaneous velocity vector coincide in direction with the tangent to the path.

Thus, the *instantaneous velocity vector of a particle moving along a curvilinear path has the direction of the tangent to the path*.

3. The projections of the displacement vector are

$$\left. \begin{aligned} \Delta r_x &= x_2 - x_1 = \Delta x \\ \Delta r_y &= y_2 - y_1 = \Delta y \\ \Delta r_z &= z_2 - z_1 = \Delta z \end{aligned} \right\} \quad (3.15)$$

The corresponding projections of the instantaneous velocity are

$$v_x = \lim_{\Delta t \rightarrow 0} \frac{\Delta x}{\Delta t}, \quad v_y = \lim_{\Delta t \rightarrow 0} \frac{\Delta y}{\Delta t} \quad \text{and} \quad v_z = \lim_{\Delta t \rightarrow 0} \frac{\Delta z}{\Delta t} \quad (3.16)$$

The magnitude v of the instantaneous velocity

$$v = \sqrt{v_x^2 + v_y^2 + v_z^2} \quad (3.17)$$

is called the *speed* and is simply the absolute value of \mathbf{v} .

3.6. ADDITION OF VELOCITIES

1. Since velocity is a vector, the addition of velocities in Newtonian mechanics is reduced to the addition of the corresponding projections of the velocities. Consequently, velocities are added by the polygon (or parallelogram) method.

Suppose a man is moving with the velocity \mathbf{u} on a flatcar in a direction that makes the angle α with a side of the car. The flatcar moves

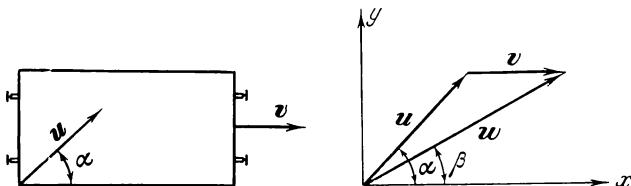


Fig. 3.12

with the velocity \mathbf{v} with respect to the earth (Fig. 3.12). Let us find the magnitude and direction of the velocity \mathbf{w} of the man with respect to the earth.

This problem can be solved as follows. The projections of vector \mathbf{u} are

$$u_x = u \cos \alpha \quad \text{and} \quad u_y = u \sin \alpha$$

and the corresponding projections of vector \mathbf{v} are

$$v_x = v \quad \text{and} \quad v_y = 0$$

Hence

$$w_x = u_x + v_x = u \cos \alpha + v \quad \text{and} \quad w_y = u_y + v_y = u \sin \alpha$$

Then the magnitude of the man's velocity with respect to the earth is

$$\begin{aligned} w &= \sqrt{w_x^2 + w_y^2} = \sqrt{u^2 \cos^2 \alpha + 2uv \cos \alpha + v^2 + u^2 \sin^2 \alpha} \\ &= \sqrt{u^2 + v^2 + 2uv \cos \alpha} \end{aligned} \quad (3.18)$$

and its direction is determined by

$$\sin \beta = \frac{w_y}{w} = \frac{u \sin \alpha}{w} \quad (3.19)$$

The same result is obtained if the law of cosines is used.

2. The motion of the man with respect to the flatcar is said to be *relative*, that of the flatcar with respect to the earth is *bulk*, or *transport, motion* and that of the man with respect to the earth is *compound motion*.

Thus, we find that in Newtonian mechanics the *velocity of compound motion equals the vector sum of the velocities of the relative and transport motions*:

$$\mathbf{w} = \mathbf{u} + \mathbf{v} \quad (3.20)$$

Equation (3.20) is the *classical law of the addition of velocities in vector form*. As we shall see in Sec. 12.5, velocities are added according to more complex rules in the theory of relativity, and equation (3.20) cannot be applied.

Chapter 4 ACCELERATION

4.1. AVERAGE AND INSTANTANEOUS ACCELERATION

1. The *average acceleration* during a given time interval is a physical quantity equal to the change of velocity divided by the time interval.

Suppose a particle has the instantaneous velocity \mathbf{v}_1 at the instant of time t_1 and velocity \mathbf{v}_2 at time t_2 . Then, according to the definition,

$$\mathbf{a}_{av} = \frac{\mathbf{v}_2 - \mathbf{v}_1}{t_2 - t_1} = \frac{\Delta \mathbf{v}}{\Delta t} \quad (4.1)$$

The definition obviously implies that acceleration is a vector. The direction of the acceleration vector depends upon the character of motion of the particle. This will be dealt with in more detail below.

The *instantaneous acceleration* is a physical quantity equal to the limit which the average acceleration approaches in an infinitely short time interval. Thus

$$\mathbf{a} = \lim_{\Delta t \rightarrow 0} \frac{\Delta \mathbf{v}}{\Delta t} \quad (4.2)$$

2. The dimensions of acceleration are

$$[a] = \frac{[v]}{[t]}$$

Since the SI unit for velocity is m/s, the unit for acceleration in this system is the metre per second squared (m/s^2). Also used is the centimetre per second squared: $1 \text{ cm/s}^2 = 10^{-2} \text{ m/s}^2$.

4.2. RECTILINEAR VARIABLE MOTION

1. If a particle moves along a straight line, its velocity and acceleration vectors are in the direction along this line. Consequently, only the magnitude of the velocity can change in rectilinear motion. Here the average and instantaneous accelerations are

$$a_{av} = \frac{v_2 - v_1}{t_2 - t_1} = \frac{\Delta v}{\Delta t} \quad (4.3)$$

$$a = \lim_{\Delta t \rightarrow 0} \frac{\Delta v}{\Delta t} \quad (4.4)$$

2. Variable motion is said to be *accelerated* if the velocity of the particle continually increases in magnitude, i.e. the condition $v_2 > v_1$ holds for any $t_2 > t_1$. An example is the motion of a heavy ball falling freely from a small height.

Accordingly, the motion of a particle is said to be *decelerated* if its velocity decreases continually in magnitude, i.e. the condition $v_2 < v_1$ holds for any $t_2 > t_1$. An example is the motion of a heavy ball thrown vertically upward (until it reaches its highest point).

It follows from definitions (4.3) and (4.4) that in the case of accelerated motion of a particle the acceleration is expressed by a positive number and in the case of decelerated motion by a negative number. This means that in the first case the vectors of velocity and acceleration are in the same direction and in the second the direction of acceleration is opposite to that of the velocity.

4.3. UNIFORMLY ACCELERATED RECTILINEAR MOTION

1. The motion of a particle is said to be *uniformly accelerated* if its instantaneous velocity is a linear function of time:

$$v = v_0 + at \quad (4.5)$$

where v_0 and a are constants.

Here v_0 means the *initial velocity*. At $t = 0$ the velocity $v = v_0$, where by $t = 0$ we mean the instant of time we began to observe the motion of the particle. This instant does not, in general, coincide with the instant the motion began, which is often unknown (see Sec. 1.4). Cases are possible, however, in which the observation and motion begin at the same instant (for example, if we start a stopwatch at the instant we release a stone and it begins to fall). In this special case $v_0 = 0$ and $v = at$.

2. Let us find the meaning of a in equation (4.5). According to the definition, the average acceleration of uniformly accelerated motion is

$$a_{av} = \frac{v_2 - v_1}{t_2 - t_1} = \frac{v_0 + at_2 - (v_0 + at_1)}{t_2 - t_1} = \frac{a(t_2 - t_1)}{t_2 - t_1} = a$$

The limit of a constant value is the constant value itself. Therefore, the instantaneous acceleration in uniformly accelerated motion is the quantity a in equation (4.5) for the instantaneous velocity.

Thus, another way of defining *uniformly accelerated motion* is to say that it is motion at constant acceleration.

4.4. VELOCITY GRAPH FOR UNIFORMLY ACCELERATED MOTION

To plot a velocity graph we lay off the time on the axis of abscissas and the corresponding instantaneous velocities on the axis of ordinates. The line joining the representative points thus obtained is the velocity graph.

For uniformly accelerated motion the velocity graph is a straight line that intercepts the value of the initial velocity on the axis

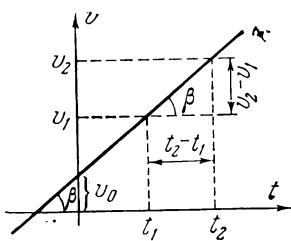


Fig. 4.1

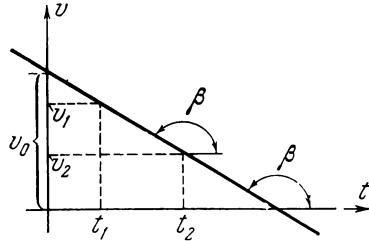


Fig. 4.2

of ordinates (Fig. 4.1). The tangent of the angle of inclination of the graph is equal in magnitude to the acceleration. Thus

$$\tan \beta = \frac{v_2 - v_1}{t_2 - t_1} = a \quad (4.6)$$

In accelerated motion the graph makes an acute angle with the axis of abscissas (Fig. 4.1); in decelerated motion it is an obtuse angle (Fig. 4.2).

4.5. GRAPHICAL CALCULATION OF DISPLACEMENT

1. The relation between the displacement and the distance travelled was dealt with in Sec. 1.6 using the example of a particle moving along the axis of abscissas. This relationship holds as well for the case when the particle moves along an arbitrary path.

In a velocity graph for uniform motion (Fig. 4.3) the product vt is the area of the hatched rectangle. Consequently, the displacement $l = vt$ is equal in magnitude to the area of this rectangle (in a properly selected scale).

This conclusion can be generalized for the case of arbitrary variable motion. Suppose the velocity of a particle is an arbitrary function $v = f(t)$, the graph of this function being that shown in Fig. 4.4.

Let us divide the time interval $t = t_2 - t_1$ into several parts and assume that the velocity does not change during a small interval

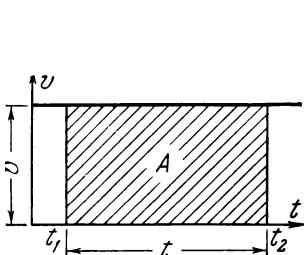


Fig. 4.3

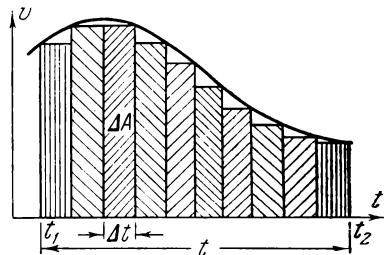


Fig. 4.4

Δt , but then suddenly takes a new value. This motion with "stepwise" velocity represents, to some degree of accuracy, the true motion in which the velocity changed smoothly, as shown by the graph, rather than in steps. The displacement during the small interval Δt is equal in magnitude to the area of the hatched strip since we assumed the motion to be uniform during this interval. The displacement during the time t is approximately equal (in magnitude, of course) to the area of the stepped figure. The shorter the intervals Δt , the more accurately this "stepwise" motion represents the true motion. We shall obtain an exact result if we consider the limit when $\Delta t \rightarrow 0$.

2. To sum up, we can state that in a velocity graph the displacement in arbitrary variable motion of a particle is equal in magnitude to the area of the figure bounded below by the axis of abscissas, at the right and left by the ordinates of the instantaneous velocity at the instants of time t_2 and t_1 and above by the curvilinear velocity graph.

We should note that the calculation of such areas is a topic taken up in the branch of mathematics called integral calculus.

4.6. DISPLACEMENT AND AVERAGE VELOCITY IN UNIFORMLY ACCELERATED MOTION

1. Using the conclusion reached in the preceding section, we can readily calculate the displacement of a particle travelling with uniformly accelerated motion. Here the velocity is a linear function

of time, the velocity graph is a straight line, and the required area of the trapezoid is equal to one half the sum of the bases multiplied by the height (see Fig. 4.1). It follows that at $t_1 = 0$ and $t_2 = t$ the displacement can be calculated by the equation

$$l - l_0 = \frac{v_0 + v}{2} t \quad (4.7)$$

Substituting the value $v = v_0 + at$, we obtain

$$l = l_0 + v_0 t + \frac{at^2}{2} \quad (4.8)$$

Hence, the coordinate of a particle travelling with uniformly accelerated motion is a quadratic function of time. Equation (4.8) is the required motion equation. The graph of the distance travelled is a parabola. Shown in Fig. 4.5 is the simplest case of such motion when

$$l_0 = 0, v_0 = 0 \text{ and } l = at^2/2$$

The student should have no difficulty in showing that the examples of variable motion discussed in Sec. 1.5 concern uniformly accelerated motion.

2. Using the definition (3.13) of the average velocity it can readily be shown that the magnitude of the average velocity in uniformly accelerated motion of a particle is one half the sum of the initial and final velocities:

$$v_{av} = \frac{v_1 + v_2}{2} \quad (4.9)$$

To prove this we can write by definition that

$$v_{av} = \frac{l_2 - l_1}{t_2 - t_1} \quad (4.10)$$

Substituting their values for l_2 and l_1 , after simple transformations we obtain

$$\begin{aligned} v_{av} &= \frac{l_0 + v_0 t_2 + (at_2^2/2) - [l_0 + v_0 t_1 + (at_1^2/2)]}{t_2 - t_1} \\ &= \frac{v_0 (t_2 - t_1) + (1/2) a (t_2 - t_1) (t_2 + t_1)}{t_2 - t_1} \\ &= \frac{2v_0 + at_1 + at_2}{2} = \frac{(v_0 + at_1) + (v_0 + at_2)}{2} = \frac{v_1 + v_2}{2} \end{aligned}$$

Note that this derivation is valid only for uniformly accelerated motion. It does not hold for the general case of variable motion. This can be demonstrated, for instance, by the travel of a bus that starts moving at one stop and slows down and stops at the next. Then $v_1 = v_2 = 0$ but the average velocity obviously does not equal zero.

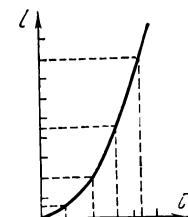
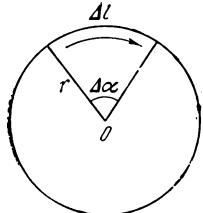


Fig. 4.5

4.7. UNIFORM CIRCULAR MOTION OF A PARTICLE

1. Let a particle move with constant speed in a circle of radius r . The displacement of the particle during the time Δt is the arc $\Delta l = r \cdot \Delta\alpha$, where $\Delta\alpha$ is the angle of rotation of the radius (Fig. 4.6).



In uniform circular motion of a particle the quantity

$$\omega = \frac{\Delta\alpha}{\Delta t} \quad (4.11)$$

is called the *angular velocity* of the particle.

Fig. 4.6 The unit of measurement of the angular velocity is the radian per second (rad/s).

The dimensions of the angular velocity are

$$[\omega] = \frac{[\alpha]}{[t]}$$

From the definition of the velocity of uniform motion we can write

$$v = \frac{\Delta l}{\Delta t} = \frac{r \cdot \Delta\alpha}{\Delta t} = r\omega \quad (4.12)$$

2. The time T required for the particle to make one complete revolution is called the *period*. The quantity v , the reciprocal of the period, indicates the number of revolutions the particle makes in unit time. It is called the *frequency*:

$$\nu = 1/T \quad (4.13)$$

3. During a time interval equal to one period the particle travels a full circle, i.e. when $\Delta t = T$, $\Delta l = 2\pi r$.

Hence

$$v = \frac{2\pi r}{T} = 2\pi r\nu \quad (4.14)$$

Combining equations (4.12) and (4.14), we obtain

$$\omega = \frac{2\pi}{T} = 2\pi\nu \quad (4.15)$$

4.8. ACCELERATION IN UNIFORM CIRCULAR MOTION OF A PARTICLE

1. In the uniform circular motion of a particle the magnitude of its velocity, i.e. its speed, does not change: $|v| = v = \text{const}$. This does not imply, however, that the particle is moving without acceleration. On the contrary, the *direction* of the velocity, which is a vector tangent to the path, is changing continuously, and this means that the particle is moving with acceleration.

2. Let us construct the vector of the change in velocity, $\Delta v = v' - v$, according to the rule given in item 2 of Sec. 3.3 (Fig. 4.7). For a small value of Δt the arc AB approaches the chord AB . In the similar triangles AOB and BMN

$$\frac{\Delta v}{v} = \frac{\Delta l}{r}$$

It follows that the magnitude of the average acceleration is

$$a_{av} = \frac{\Delta v}{\Delta t} = \frac{v}{r} \cdot \frac{\Delta l}{\Delta t}$$

and the magnitude of the instantaneous acceleration is

Fig. 4.7

$$a = \lim_{\Delta t \rightarrow 0} a_{av} = \lim_{\Delta t \rightarrow 0} \frac{\Delta v}{\Delta t} = \lim_{\Delta t \rightarrow 0} \left(\frac{v}{r} \times \frac{\Delta l}{\Delta t} \right)$$

Transposing the constant factor v/r to the left of the limit symbol, we obtain

$$a = \frac{v}{r} \lim_{\Delta t \rightarrow 0} \frac{\Delta l}{\Delta t} = \frac{v}{r} \times v = \frac{v^2}{r} \quad (4.16)$$

But according to equation (4.12), $v = \omega r$. Substituting into equation (4.16), we obtain

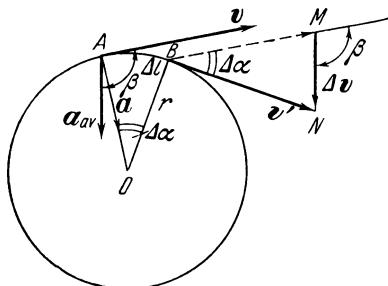
$$a_n = \frac{v^2}{r} = \omega^2 r \quad (4.17)$$

It remains to determine the direction of the vector of instantaneous acceleration.

The average acceleration vector makes the angle $\beta = (\pi + \Delta\alpha)/2$ with the velocity vector. But when $\Delta t \rightarrow 0$, angle $\Delta\alpha \rightarrow 0$ as well. Consequently, the angle between the instantaneous acceleration vector and the velocity vector is

$$\varphi = \lim_{\Delta t \rightarrow 0} \frac{\pi + \Delta\alpha}{2} = \frac{\pi}{2} \quad (4.18)$$

Thus, a particle in uniform circular motion has an acceleration perpendicular to the velocity vector, i.e. instantaneously along a radius inward toward the centre of the circle along which the particle is travelling. Hence, this acceleration is said to be *normal*, *radial*, or *centripetal* ("seeking a centre").



Chapter 5

FORCE

5.1. FORCE, A MEASURE OF THE INTERACTION OF BODIES

1. We know by experience that all bodies in nature interact in some way with one another. The atmospheric air, for instance, exerts pressure on the earth's surface and on all objects on the earth. The interaction between the molecules of water and the surface of a bather's body makes the drops of liquid stick to his body. The strong interaction between the parts of an atomic nucleus (protons and neutrons) is the reason why it is so difficult to break up (or "smash") the nucleus.

Force is a measure of the interaction of bodies or of the particles of which the bodies consist.

The word is from the Latin "*fortis*" meaning strong. This is also the origin of the word "effort". The idea of force first arose from the assessment of muscular stress. It is necessary to stress certain muscles to lift a stone, push a boat, or draw a bowstring. Different muscles may be stressed to different degrees for each case. The degree of stress was used to evaluate the effort, or force, exerted. Many examples can be given in which the word "force" has its initial meaning: "with all one's force", "the force of arms", etc.

Later people found that the concept of force could serve to describe the action of certain bodies on others. This led to the idea of "the rising force of the wind", "to force a door" and "the force of a blow" as characteristics of external actions equivalent to muscular effort.

Still later, the term "force" acquired wider and looser connotations. New expressions such as "his force of character", "the force of an argument", and "the force of habit" came into use. Such an expanded concept of the basic idea of force has been and still is a source of much misunderstanding. It should be avoided in physics. Unfortunately, there are many terms still in use such as "electromotive force", "demagnetization force", and "magnifying force", though these concepts have nothing in common with a force as we have defined the term.

In mechanics we shall employ the term "force" only in its direct sense: as a measure of the interaction of bodies.

2. Distinction is made in modern physics between the following kinds of interaction:

(a) gravitational interaction between bodies due to universal gravitation;

(b) electromagnetic interaction between stationary or moving charged particles or bodies;

(c) strong, or nuclear, interaction of elementary particles, for instance, those of the atomic nucleus;

(d) weak interaction resulting in the decay of certain elementary particles.

Mechanics deals with the forces that appear in the direct contact of bodies: friction and elastic forces. Gravitational forces are also studied in mechanics.

3. The result of the interaction of bodies may be either deformation (change in the size or shape of bodies) or acceleration (change in the magnitude or direction of their velocity).

Of course, simultaneous deformation and acceleration are also possible.

Each of these manifestations of force can be used for its measurement. It is much simpler, however, to measure the degree of deformation rather than the acceleration. Hence, the main component of the instrument employed for measuring forces, called a *dynamometer* (from the Greek *dynamis* meaning power or strength), is a spring, the deformation of which depends upon the magnitude of the force being measured (Fig. 5.1).

5.2. ELASTIC AND PLASTIC DEFORMATION

1. The deformation of a body is said to be *elastic* if the dimensions and shape of the body are completely restored as soon as the load is removed. *Plastic deformation* is one in which the changes in the size and shape of the body do not disappear when the force ceases to act. After plastic deformation, the body retains (either partly or completely) its newly acquired shape and changed dimensions.

The behaviour of bridges, beams, walls and machine tool components should be within the region of elastic deformation when they are subject to various forces. Only in this way are they of reliable design and guaranteed against failure. In the mechanical working of materials (forging, pressworking, etc.), on the contrary, they are subjected to plastic deformation in such a way that the workpiece obtained in the working operations is of the required size and shape, which it retains when the forces are removed.

2. The kind of deformation that occurs depends upon the magnitude of the load applied and the length of time it acts, as well as on the material of which the body is made and the state of this material (its temperature, previous working, etc.).

For instance, if we bend a steel strip and immediately release it, it will resume its original shape. Here we are obviously dealing with

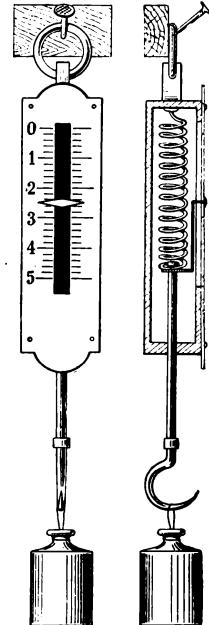


Fig. 5.1

elastic deformation. But if the strip is clamped in the bent state for a long time, it will not return completely to its original shape when we remove the external force by unclamping the strip. A considerable increase in the time the external force is applied transforms elastic deformation into plastic.

Temperature greatly affects the kind of deformation that occurs. A steel strip heated to yellow heat becomes plastic at comparatively small external forces, which at room temperature could only result in elastic deformation of the strip. On the other hand lead, which is quite plastic at room temperature, becomes elastic at low temperatures. If a lead bell is struck after being cooled by liquid air it gives a clear, resonant sound. This means that it is subject to elastic deformations. The increased plasticity of metals with temperature is used in the working of metals.

Thus, we see that there are no sharply defined boundaries between elastic and plastic deformation. Experience shows, however, that small deformations of short duration can, with sufficient accuracy, be regarded as elastic ones.

3. We know from practice that the greater the deformation we wish to obtain, the greater the force we must apply to the body being deformed. Consequently, we can assess the magnitude of the force applied by the amount of deformation it produces. Analyses of many experiments indicate that the *magnitude of the elastic deformation is proportional to the applied force*. This is a statement of *Hooke's law*, discovered by Robert Hooke (1635-1703).

In deformation leading to the elongation (or compression) of a body

$$F = k(l - l_0) = k \Delta l \quad (5.1)$$

where F = magnitude of the force applied

l_0 = initial length of the body

l = length of the deformed body

k = proportionality factor, called the *coefficient, or modulus, of elasticity*.

The coefficient of elasticity is equal in magnitude to the force causing a deformation of unit length.

Hooke's law indicates that the dynamometer scale should have uniform graduations (see Fig. 5.1). This considerably simplifies the graduation of the dynamometer.

5.3. FORCE IS A VECTOR

1. Every force has a definite direction, and the result of its action depends on this direction and the magnitude of the force. The friction force, for instance, is in the direction opposite to that of the velocity

vector. If a spring is stretched by a force, it will be compressed when the force is reversed. When a soccer player kicks the ball he imparts an acceleration to it in the same direction as that of the force which his foot applies to it. Finally, Hooke's law indicates that force is a vector because the coefficient of elasticity is a scalar and the elongation $\Delta l = l - l_0$ is a vector quantity. Their product is a vector (see Sec. 3.3).

Therefore, *force is a vector*.

2. Hooke's law can be written in vector form, taking into account the directions of the displacement and force vectors. Here distinction should be made between the external force and the elastic force.

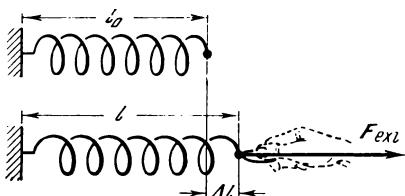


Fig. 5.2

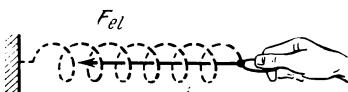


Fig. 5.3

Suppose an external body, for example, the hand, stretches a helical spring (Fig. 5.2). The force with which it acts on the spring is called the *external force*. The direction of the external force coincides with that of the displacement. Consequently, these vectors have the same direction. Thus

$$\mathbf{F}_{ext} = k\Delta l \quad (5.2)$$

We know by experience that a deformed spring acts on the hand with a force equal to the external force but opposite in direction. The force with which a deformed spring acts on an external body is called the *elastic force* (Fig. 5.3). Since its direction is opposite to the elongation vector,

$$\mathbf{F}_{el} = -k\Delta l \quad (5.3)$$

3. After analyzing the different kinds of interactions of bodies Newton came to the conclusion that *two bodies always interact with forces that are equal in magnitude but opposite in direction*:

$$\mathbf{F}_{12} = -\mathbf{F}_{21} \quad (5.4)$$

where \mathbf{F}_{12} is the force with which the first body acts upon the second and \mathbf{F}_{21} is the force with which the second body acts upon the first. This equation is a statement of *Newton's third law of motion*.

As an example we can consider an experiment on the interaction between a magnet and a piece of iron. Dynamometers attached to the



Fig. 5.4

two bodies will register forces which are equal in magnitude but opposite in direction (Fig. 5.4).

Examples of the application of Newton's third law will be taken up in the following chapters.

5.4. VECTOR ADDITION AND RESOLUTION OF FORCES APPLIED TO A PARTICLE

1. If several forces are applied to a particle, they can be replaced by the *resultant force*. This resultant is the vector sum of the forces and can be found by the polygon method (Fig. 5.5). The resultant of two for-

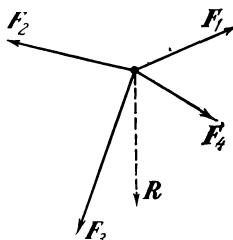


Fig. 5.5

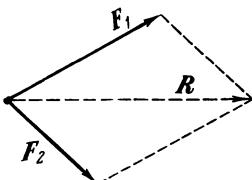
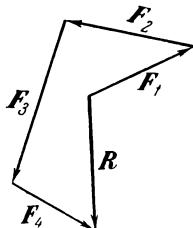


Fig. 5.6

ces can be determined by the parallelogram method as well (Fig. 5.6), but this method is less convenient than the polygon method if more than two forces are involved.

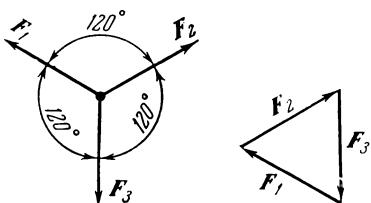


Fig. 5.7

If in constructing a force polygon we find that it is closed, we must conclude that the resultant of the given system of forces equals zero. Such a system of forces is said to be *balanced*, or *in equilibrium*. An example is a system of three forces equal in magnitude arranged in a single plane at angles of 120° to one another (Fig. 5.7).

A system of forces, applied at a single point, can be balanced by applying a *balancing*, or *equilibrium*, force to the point. The balan-

cing force is equal in magnitude to the resultant force, but is opposite in direction.

2. A force, like any other vector, can be resolved into two components (see Sec. 3.4), and it is frequently necessary in practice to resolve a force into two components in given directions. An example follows.

Suppose that a certain force P , for instance the weight of some load, is applied to a bracket (Fig. 5.8). The problem is to find the forces acting in the bars. Here the required forces are directed along

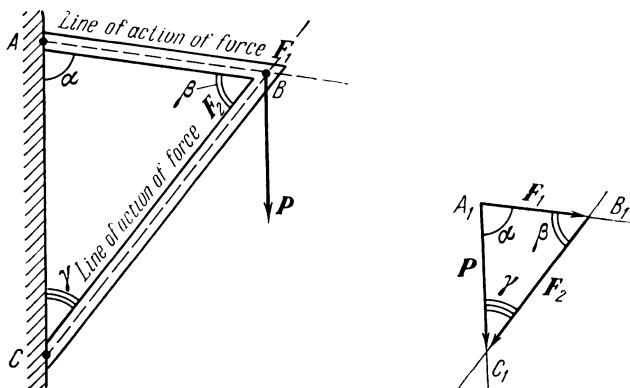


Fig. 5.8

the bars. We construct triangle $A_1B_1C_1$ so that its sides are parallel to force P and to the bars. The length of side A_1C_1 is equal to P (in some convenient scale). Then vector $F_1 = \overrightarrow{A_1B_1}$ is the force stretching bar AB , and vector $F_2 = \overrightarrow{B_1C_1}$ is the force compressing bar BC . In the scale used forces F_1 and F_2 are equal to the sides A_1B_1 and B_1C_1 of the triangle. They can also be calculated from the following proportion:

$$\frac{F_1}{AB} = \frac{F_2}{BC} = \frac{P}{AC}$$

Chapter 6

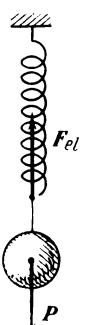
WEIGHT AND MASS

6.1. FORCE OF GRAVITY. WEIGHT

1. We know from our everyday experience that all bodies fall to the earth if this motion is not hindered by other bodies. What is the cause of this falling motion?

Let us carry out the following experiment. We attach a weight with a string to a spring and release the weight. The weight will move downward a certain distance and then stop. The stretched spring acts on the weight with a certain force F_{el} , directed vertically upward, but the weight remains in equilibrium. This is possible only if the weight is subject to the action of another force P , equal in magnitude to the first force, but directed vertically downward, toward the earth (Fig. 6.1).

If we cut the string joining the weight to the spring, the interaction between them disappears. The spring resumes its initial shape and size again. But the interaction between the weight



and the earth remains; the earth continues to attract the weight with the force P . Owing to this force, the weight is moved from its state of rest and begins to fall to the earth.

The force with which the earth attracts bodies is called the force of gravity, the gravitational force, or simply gravity.

2. The force of gravity acts upon all bodies near the earth. If they do not fall to the earth, then their motion is restricted by certain other bodies: a support, string, spring, wall, etc.

Bodies that restrict the motion of other bodies are called *constraints*. The origin of this term is obvious. These bodies constrain the motion of the given bodies, impeding free motion. Thus, the surface of the table is the constraint for all objects lying on it; the floor serves as the constraint for the table; etc.

Constraints are deformed by the action of gravity, and, according to Newton's third law, the reaction of the deformed constraint balances the force of gravity. While it is comparatively easy to observe the deformation of a spring, the observation of the deformation of other constraints often requires instruments that register very small displacements.

3. *The weight is the force with which a body acts on a constraint (presses on a horizontal support or stretches a string) owing to the attraction of this body to the earth.*

It will be shown in Sec. 7.2 that the weight of a body is equal to the force of gravity if the body and its constraint are at rest with respect to the earth (or are travelling with uniform motion in a straight line).

The reaction of the deformed constraint can be used to measure weight. This can be done by means of a dynamometer (spring balance) calibrated in units of force.

We can, however, proceed in an entirely different way. We can select a *standard weight*, a body whose weight equals unit force, and

then make weights which are several times larger or are a fraction of this standard. We will then have a set of weights. Now we can place a body whose weight we wish to determine on one pan of an equal-armed balance and then counterbalance it with weights from our set placed in the other pan (Fig. 6.2). Since equal-armed beam balances are in equilibrium when equal forces act on the pans, we conclude that the weight of the body equals that of the weights counterbalancing it.

Hence, weight can be determined either with a spring balance or with a beam balance.

6.2. FREE FALL

1. *Free fall* is the motion of bodies subject only to the force of gravity. To investigate the free fall of bodies we must eliminate the influence of all other forces and, in particular, of the air resistance.

It can be readily seen that air resistance substantially affects the falling of bodies. Suppose we drop a stone and a piece of cotton wool. The stone rapidly drops to the earth while the cotton wool descends more slowly. We could come to the mistaken conclusion that heavier objects fall faster than lighter ones. But if we roll the cotton wool into a tight ball it will fall much faster to the earth. What changes have we made in this experiment? We have reduced the size of the ball of cotton wool and, consequently, the air resistance. In exactly the same way a sheet of paper crumpled and rolled into a tight wad is subject to much less air resistance and, therefore, falls much faster than a flat sheet of the same weight.

Naturally, the question arises as to how bodies fall in a vacuum where there is no air resistance. Will they fall with the same acceleration? Further on we shall see that at a given point on the earth all bodies fall with the same acceleration regardless of their weight. We shall test this statement with the following experiment (Fig. 6.3). Let us take a glass tube more than a metre long, sealed at one end and closed at the other by a cover with a pipe connection. The tube contains a piece of lead and a feather. If we suddenly overturn the tube, the lead will fall rapidly while the feather will flutter slowly to the bottom. Then we connect the tube to a vacuum pump and exhaust all the air. Now, if we overturn the tube, we shall see that both bodies fall together, and that the feather is not in the least behind the lead.

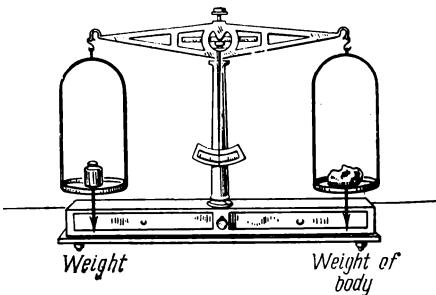


Fig. 6.2

Therefore, the *free fall of all bodies is exactly the same regardless of their weight.*

2. The first scientist to reach this conclusion was Galileo at the end of the 16th century. Prevalent up to that time was Aristotle's teaching that heavy bodies fall proportionately faster than lighter ones. To prove his viewpoint Galileo is said to have conducted a number of experiments, dropping various objects from the famous Leaning Tower of Pisa and measuring the time of their fall.

Galileo showed that free fall is uniformly accelerated motion. To prove this statement, it is sufficient to show that the distance travelled by a body in free fall is proportional to the square of the time required for the motion (see Sec. 4.6). Galileo had no instruments that could precisely measure very short intervals of time. Hence, he imitated free fall by the motion of a ball rolling down an incline.

3. Today we can measure the time it takes a body to fall from any height by using an electric stopwatch. We hold a small steel ball at a certain height with an electromagnet. A special relay enables us to de-energize the electromagnet and simultaneously start the stopwatch. The ball, dropping alongside a vertical rule, travels a certain distance and switches off the stopwatch at the end of its path. By changing the distance and measuring each time of fall, we can readily establish the equation of motion. Air resistance can be excluded by placing the whole installation in a closed tube and exhausting all the air. If a small heavy ball is dropped from a small height, the air resistance can even be neglected (see Sec. 11.9).



Fig. 6.3

The results of these experiments prove the following:

- (a) *free fall is uniformly accelerated motion;*
- (b) *all bodies fall freely with the same acceleration g ;*
- (c) *free-fall acceleration (the acceleration of gravity) varies slightly at various points on the earth.* This difference (which does not exceed 0.6 per cent) can be neglected for practical purposes not involving too precise measurements, and the value can be taken equal to $g = 9.81 \text{ m/s}^2$.

6.3. MASS OF A BODY

1. Experiments show that the force of gravity acting upon a given body is different at various points on the earth; it varies with the geographical latitude and the height above sea level. But the ratio of the force of gravity to the free-fall acceleration for a given body at

any point on the earth is a constant value:

$$\frac{P_1}{g_1} = \frac{P_2}{g_2} = \frac{P_3}{g_3} = \dots = \text{const}$$

Therefore, the ratio P/g is a certain characteristic of the body. This quantity is called the *mass* of the body.

The *mass of a body is the ratio of the force of gravity acting on the body to the free-fall acceleration*. Thus

$$m = \frac{P}{g} \quad (6.1)$$

The concept of mass is one of the most important ones in physics. It will be dealt with many times in later chapters of this book: in studying Newton's laws of motion and the law of universal gravitation; in connection with such vital concepts as momentum and energy. This will enable us to considerably extend our ideas on mass.

2. Mass is a scalar quantity; the force of gravity and free-fall acceleration are vectors having the same direction. It is better to write equation (6.1) in vector form:

$$\mathbf{P} = m\mathbf{g} \quad (6.2)$$

3. Since free-fall acceleration is the same for all bodies, it follows from equation (6.1) that

$$\frac{m_1}{m_2} = \frac{P_1}{P_2} \quad (6.3)$$

i.e. the *ratio of the masses of two bodies equals the ratio of their weights*, because here the weight equals the force of gravity.

This is the basis for determining the mass by means of a beam balance. As mentioned above (see Fig. 6.2), a beam balance is in equilibrium when the weight of the body is equal to the weight on the other pan. *But then the mass of the body is equal to the mass of the weights*. After making a standard weight and accepting its mass as unity, we can compare the mass of any other body with that of the standard by using a beam balance. Consequently, the verb "to weigh" is actually used in two different senses: to determine the weight, and to determine the mass.

Therefore, the *mass of a body can be determined by a beam balance*.

6.4. THE DENSITY OF SUBSTANCES

If bodies of different volumes are made of the same substance they will have different masses. Experience shows, however, that the ratio of the masses of these bodies to their volumes is a constant value. Thus

$$\frac{m_1}{V_1} = \frac{m_2}{V_2} = \dots = \text{const}$$

This ratio can serve as a characteristic of a substance: its *density*.

The *density* of a substance is the ratio of mass to volume:

$$\rho = m/V \quad (6.4)$$

As will be shown further on, the density of a substance depends upon its temperature, the external pressure, and the state of aggregation. Hence, handbooks always specify the conditions under which tabulated density data are correct. Without this information the data is meaningless.

Chapter 7

FUNDAMENTAL LAW OF DYNAMICS

7.1. FORCE AND ACCELERATION

The law of inertia implies that a body cannot change its velocity by itself without interaction with surrounding bodies. Every change in the magnitude or direction of a body's velocity is caused by the action of external bodies upon it. This action, as we know, is characterized by forces (see Sec. 5.1). *The fundamental law of dynamics* (otherwise called *Newton's second law of motion*) expresses the relation between force and the change in velocity of interacting bodies.

The simplest form of the fundamental law of mechanics is valid for inertial reference frames. Noninertial frames will be dealt with further on (see Ch. 24).

We shall make our first acquaintance with this fundamental law given the condition that the velocity of the particle with respect to the selected reference frame is considerably less than the velocity of light. We shall show that many problems of mechanics can be solved in this approximation with an accuracy sufficient for all practical purposes. The more general relativistic case will be taken up later (see Ch. 13).

2. The essence of the fundamental law of dynamics can be made clear by the following experiment. A smooth, solid rod is fixed in the front and back walls of a railway car. A body of mass m slides with negligible friction along this rod. A dynamometer linked to the body has its other end secured to the front wall of the car (Fig. 7.1). Suppose the car is travelling with a certain acceleration a with respect to an inertial reference frame (for instance, the earth).

If the body is stationary with respect to the car, then with respect to the earth (an inertial reference frame) it moves with the same acceleration a as the car. The experiment will show that in this case the dynamometer spring is in the deformed state. This means that the spring acts on the body with a certain force, which imparts the acceleration to the body. Thus, if we terminate the interaction

between the spring and the body (for instance, by cutting the string that holds them together), the spring will return to its undeformed state. As for the body, it will begin to travel with uniform motion with respect to the earth, retaining by inertia the velocity it acquired before interaction was discontinued.

If we repeat this experiment with bodies of different masses and at different accelerations of the car, we shall find that the degree of elongation of the spring is proportional to the mass of the body (for

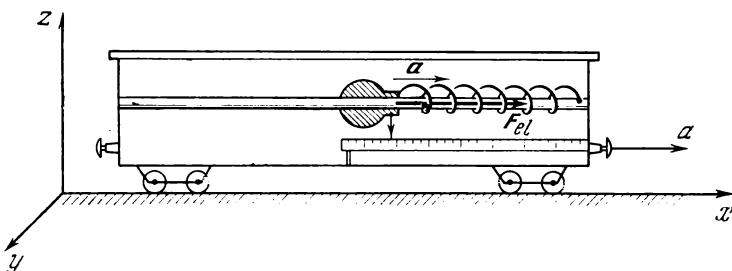


Fig. 7.1

constant acceleration) and to the acceleration (for a body of given mass). Moreover, the elongation of the spring does not depend on the velocity of the body and car. By definition, however, the elastic force is proportional to the elongation of the spring (see Sec. 5.2). Hence, the elongation is a measure of the acting force.

The results of this experiment can be stated as follows:

(a) *If a body is travelling with acceleration with respect to a certain inertial reference frame, it is subject to the action of a force.*

(b) *A force produces an acceleration that is proportional to the force and coincides with it in direction. For a given acceleration, the force is proportional to the mass of the body to which acceleration is imparted.*

(c) *At velocities considerably less than that of light the force does not depend upon the velocity of the body being accelerated.*

3. A similar result will be obtained if we investigate the curvilinear motion of a particle. This is shown by the following experiment. A rod with a weight and spring is secured by one end to the vertical spindle of a centrifugal machine and is rotated about the spindle axis (Fig. 7.2). We shall find that the spring is stretched when the system is rotated, and that it acts on the body with a force which depends upon the mass of the body, its distance from the axis of rotation, and the speed of rotation.

This result is readily explained. In uniform circular motion of a body in a circle, the magnitude of the velocity is constant while the direction of the velocity vector is continuously changing. But we have already established (Sec. 4.8) that here we have *normal*,

or *centripetal*, *acceleration*. Thus

$$a_n = v^2/r = \omega^2 r$$

The stretched spring imparts this acceleration to the body.

After repeating this experiment with various bodies, at various speeds of rotation, and at various distances from the body to the axis of rotation, we shall find that the force acting on the body is proportional to the mass of the body and to the centripetal acceleration. The product of the mass of the body by the centripetal acceleration is called the *centripetal force*.

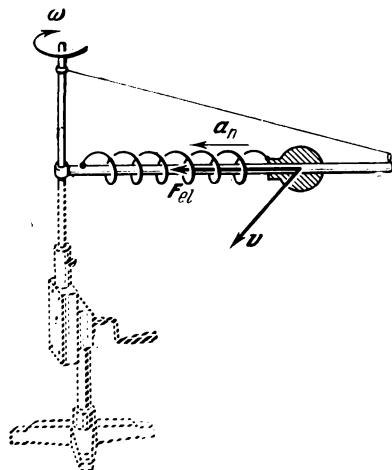


Fig. 7.2

4. Generalizing the results of the two series of experiments we come to the following conclusion: the *force equals the product of the mass of a body by the acceleration imparted by the force to the body*:

$$\mathbf{F} = m\mathbf{a} \quad (7.1)$$

This is Newton's second law of motion (or the fundamental law of dynamics) for the case when the velocity of the body is much less than that of light in a vacuum.

5. The fundamental law of dynamics can be written in a somewhat different form which often proves convenient. By definition, the acceleration is

$$\mathbf{a} = \frac{\mathbf{v}_2 - \mathbf{v}_1}{t_2 - t_1}$$

Substituting in equation (7.1) we obtain

$$\mathbf{F} = \frac{m\mathbf{v}_2 - m\mathbf{v}_1}{t_2 - t_1} = \frac{\Delta(m\mathbf{v})}{\Delta t} \quad (7.2)$$

Momentum is a vector equal to the product of the mass of a body by its velocity:

$$\mathbf{p} = m\mathbf{v} \quad (7.3)$$

Then the fundamental law of dynamics (7.2) can be stated as follows: the *force is equal to the change in the momentum in unit time*. Thus

$$\mathbf{F} = \frac{\Delta\mathbf{p}}{\Delta t} \quad (7.4)$$

This is the most general statement of the fundamental law of dynamics.

6. Finally, if a body is subject to the action of a set of forces, with the resultant equalling \mathbf{R} , Newton's second law is written as follows:

$$\mathbf{R} = m\mathbf{a} = \frac{\Delta(m\mathbf{v})}{\Delta t} \quad (7.5)$$

We should note that when a particle travels with uniform circular motion, the resultant force is toward the centre. It is the centripetal force.

7.2. APPLYING THE FUNDAMENTAL LAW OF DYNAMICS

1. Making use of the fundamental law of dynamics we can determine the forces acting on a body, or the character of its motion if the forces are given. Thus, if the motion equation is given, we can determine the acceleration of the body. When we know the acceleration and the mass of the body, it is easy to find the acting force.

The following rules should be observed in setting up the motion equation:

first, *all* forces acting on the particle (including the reactions) should be found;

then the resultant of these forces should be determined;

finally, on the basis of the fundamental law of dynamics we equate the resultant force with the product of the mass by the acceleration.

The motion equation obtained in this way is then solved to find the unknown quantities.

2. **Example 1.** A passenger of mass m is standing on the floor of a lift (elevator). Determine the force exerted by the passenger on the floor if the lift moves:

- (a) vertically upward with acceleration;
- (b) vertically upward with deceleration;
- (c) vertically downward with acceleration;
- (d) vertically downward with deceleration;
- (e) at uniform velocity.

Since the passenger is at rest with respect to the lift, he moves with respect to the earth with the same acceleration as the lift. We know from Newton's third law that the floor of the lift exerts the same force on the passenger that he exerts on the floor, but that these forces are in opposite directions. Hence, the passenger is subject to two forces: the force of gravity \mathbf{P} and the reaction \mathbf{Q} of the floor.

We select a coordinate system with the z -axis vertically upward. Then the reaction is a positive vector and the gravity force is negative. The sign of the acceleration vector depends upon the character of the motion (Fig. 7.3.). In case (a) the acceleration vector is directed vertically upward and is therefore positive. The same is true of case (d) because the acceleration vector in deceleration is opposed

to the velocity vector. In cases (b) and (c) the acceleration vector is directed vertically downward.

In the vector form, the motion equation is written

$$\mathbf{P} + \mathbf{Q} = m\mathbf{a} \quad (7.6)$$

When the equation is changed to the scalar form, the signs of the vectors are taken into account.

For cases (a) and (d): $-P + Q = ma$, from which

$$Q = P + ma = m(g + a)$$

For cases (b) and (c): $-P + Q = -ma$, from which

$$Q = m(g - a)$$

Finally, for case (e): $-P + Q = 0$, from which

$$Q = P$$

3. Hence, if the lift moves with acceleration with respect to the earth and is thereby a noninertial reference frame, the force which a body exerts on a support (i.e. its weight) is not equal to the force

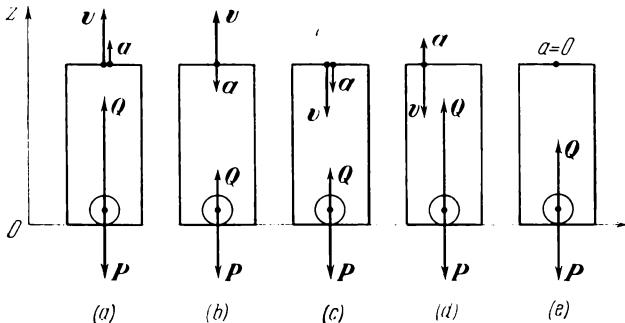


Fig. 7.3

of gravity. In particular, if the acceleration of the lift is directed opposite to the acceleration of gravity, the force exerted on the support is *greater* than the force of gravity. When the direction of the acceleration of the lift coincides with that of the acceleration of gravity, the force exerted on the support is *less* than the force of gravity. For the special case when $a = g$ it follows from equation (7.6) that $Q = 0$, i.e. the body exerts no force on the support.

Only for the case when the lift moves at uniform velocity with respect to the earth (and is, therefore, an inertial reference frame) is the force exerted on the support equal to the gravity force.

4. **Example 2.** An airplane, flying with velocity v , loops a loop of radius r in a vertical plane. With what force is the pilot pressed to

his seat? At what minimum velocity will the pilot remain in his seat without hanging on the belts? Consider the upper and lower points of the airplane's path.

Instead of looking for the force that presses the pilot to his seat, we shall determine the equal and opposing reaction of the seat. Two forces act on the pilot: the reaction of the seat and the earth's attraction. At the lower point of the loop, the reaction is directed vertically upward, and at the upper point, vertically downward. Though the airplane flies in a circle with uniform speed, the velocity vector is continuously changing in direction. Consequently, we have normal (centripetal) acceleration $a_n = v^2/r$. At the lower point of the loop it is directed vertically upward, and at the upper point, vertically downward (Fig. 7.4). For convenience we choose a coordinate system with the z-axis vertically downward, along the force of gravity.

The motion equation in the vector form is

$$\mathbf{P} + \mathbf{Q} = m\mathbf{a}_n \quad (7.7)$$

Projecting on the z-axis, we obtain for the lower point:

$$P - Q = -\frac{mv^2}{r}$$

from which

$$Q = P + \frac{mv^2}{r} = m \left(\frac{v^2}{r} + g \right)$$

Consequently, the pilot is experiencing an overload (*g*-load). For the upper point

$$P + Q = \frac{mv^2}{r}$$

from which

$$Q = \frac{mv^2}{r} - P = m \left(\frac{v^2}{r} - g \right)$$

Since Q is the magnitude of the reaction, i.e. a positive number, the value v^2/r cannot be less than g . Hence, the pilot remains in his seat without hanging on his belts if $v^2/r \geq g$. The minimum velocity required for this is found from the condition

$$v_{\min} = \sqrt{gr} \quad (7.8)$$

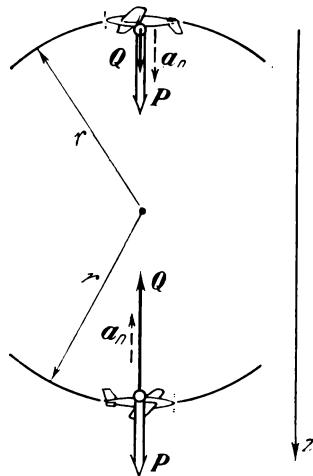


Fig. 7.4

7.3. WEIGHTLESSNESS

1. In Sec. 6.1 we defined the weight as the force with which a body acts on a constraint owing to the attraction of this body by the earth. In the preceding section we found that when a body is stationary with respect to the earth, its weight equals the force of gravity.

Analyzing the solution of Example 1 in the preceding section, we reach the conclusion that the weight also equals the force of gravity when the body is in a lift moving with respect to the earth with uniform velocity in a straight line, i.e. in an inertial reference frame. Therefore, *in any inertial reference frame weight equals the force of gravity*.

But if the body is in a lift moving with acceleration with respect to the earth, or in an airplane flying in a circle (i.e. also with acceleration with respect to the earth), the force the body exerts on its support, i.e. the weight, no longer equals the force of gravity.

A reference frame moving with acceleration with respect to an inertial reference frame is called a *noninertial reference frame*. Phenomena in noninertial reference frames will be treated in more detail in Ch. 24. Here we shall limit ourselves to an analysis of two concepts: weightlessness and *g*-load (overload).

We shall consider the force exerted by a cosmonaut (or astronaut) on his cradle in a spaceship at takeoff, during braking for re-entry, and in free flight on a geocentric orbit.

2. At takeoff the spaceship moves with acceleration owing to the action of the booster rocket. This is quite similar to item (a) in the preceding section. Here the force exerted by the cosmonaut on his cradle is greater than the force of gravity. According to Newton's third law, the force exerted on the cradle $|F_{ex}| = Q = m(g + a)$, while the force of gravity equals mg . Therefore, the cosmonaut experiences a *g*-load, i.e. it seems to him that he has become heavier.

A trained person can withstand a sixfold *g*-load, i.e. an overload of 6 *g*'s. Thus, $F_{ex} \approx 6P$. This means that the acceleration of the spaceship should not exceed the fivefold value of the acceleration of gravity because if $m(g + a) \leq 6mg$, then $a \leq 5g$.

3. It will be left to the student to show, on the basis of the solution of Example 1 (Sec. 7.2), item (d), that similar *g*-loads are experienced in braking a spaceship for re-entry.

4. When the spaceship is in orbit around the earth, its acceleration equals that of gravity, i.e. $\mathbf{a} = \mathbf{g}$. Using equation (7.6) we find that for this case $Q = 0$. Hence, the cosmonaut exerts no force on his cradle and he will experience a "loss of weight".

We can state then that *weightlessness occurs in a noninertial reference frame (the spaceship in our case) that moves with acceleration $\mathbf{a} = \mathbf{g}$ with respect to the earth*.

Note that such acceleration does not necessarily imply some definite kind of motion. The motion may be rectilinear if the rocket engines are turned off in vertical takeoff or landing, or it may be curvilinear if the spaceship is orbiting the earth. In the latter case, this acceleration is centripetal. Weightlessness does not depend upon the shape of the path. There is only one indispensable condition: the spaceship must move with an acceleration $\mathbf{a} = \mathbf{g}$.

7.4. SYSTEM OF UNITS

1. As mentioned above, the units of measurement of any physical quantity can be chosen arbitrarily. The only necessary condition is the *consistency* of the units of measurement and the quantity being measured. Thus, a solid rod can serve as the unit for measuring length, the velocity of some uniform motion for measuring velocity, etc. But such arbitrariness in the selection of units of measurement for physical quantities may lead to inconveniences. It will be necessary to introduce various numerical factors into equations relating certain quantities.

We know, for instance, that the area of a rectangle is proportional to its length and height:

$$A = \alpha l h$$

where A = area

l = length

h = height

α = factor depending upon the choice of the units. If the length and height are given in metres and the area in hectares then $\alpha = 0.0001$ and $A = 0.0001 lh$. It proves more convenient, however, to use as the unit of area the area of a square with sides equal to one unit of the length measurement. Hence, if the length is measured in metres (m), the area is measured in square metres (m^2) and the area equation is in its simplest form:

$$A = lh$$

Therefore, *it is expedient to choose the units of measurement of physical quantities so that they are related together in the same way as the quantities to be measured*. This will considerably reduce the number of numerical factors in equations expressing the relationships between physical quantities.

A set of units, worked out on the basis of this rule, is called a *system of units*.

2. All units in a given system are divided into two classes:

Fundamental units are established arbitrarily and independently of one another. Fundamental units are usually determined by means of standards.

Derived units are expressed in terms of the fundamental units by means of the corresponding physical laws.

A length unit, for instance, is fundamental, while an area unit is derived.

The dimensions of a physical quantity B express the relationship between the unit of measurement $[B]$ of this quantity and the fundamental units $[A_1]$, $[A_2]$, $[A_3]$, ... of the given system. Dimensional equations are of the form

$$[B] = [A_1]^{n_1} \times [A_2]^{n_2} \times [A_3]^{n_3} \dots$$

An effort is usually made to reduce the number of fundamental units, but this does not always prove convenient in practice. Therefore, in working out a system of units we must consider the requirements imposed by available measuring techniques and apparatus and also tradition.

7.5. THE INTERNATIONAL SYSTEM OF UNITS

1. In 1960 the Eleventh General Conference of Weights and Measures adopted the International System of Units (SI). It was adopted as preferable to all other systems of units in the USSR in 1963, and has been adopted by most countries. Three fundamental units are used in this system for measuring mechanical quantities:

metre (m)—the unit of length;
second (s or sec)—the unit of time;
kilogram (kg)—the unit of mass.

2. The derived units for measuring mechanical quantities are defined as follows:

the unit of velocity, metre per second (m/s), is the velocity of the uniform motion of a particle travelling a distance of 1 m in 1 s;

the unit of acceleration, metre per second squared (m/s^2), is the acceleration of the uniformly accelerated motion of a particle whose velocity changes 1 m/s in 1 s;

the unit of force, newton (N), is the force which, when applied to a particle with a mass of 1 kg, will give it an acceleration of 1 m/s^2 .

The units of the other mechanical quantities will be given when these quantities are defined.

3. The fundamental units—length, time, and mass—are assigned the dimension symbols L , T and M . Derived units have the following dimensions:

$$[v] = LT^{-1}, \quad [a] = LT^{-2}, \quad [F] = MLT^{-2}, \quad [\rho] = ML^{-3}, \text{ etc.}$$

7.6. THE CGS AND MK(FORCE)S SYSTEMS OF UNITS

1. In extensive use, in addition to the SI units, are the cgs (*physical*) and mk(force)s (*engineering*) systems of units.

The fundamental units of the cgs system are the *centimetre* (cm), the *second* (s or sec), and the *gram* (g).

The derived units are defined in the same way as the corresponding SI units. The unit of force is the *dyne*, the force which, when applied to a mass of 1 g, will give it an acceleration of 1 cm/s². We shall leave it to the student to show that

$$1 \text{ N} = 10^5 \text{ dynes}$$

2. The mk(force)s system is of somewhat different structure. Its kinematic units (length, time, velocity and acceleration) coincide with the SI units. The third fundamental unit, however, is the *kilogram-force* (kgf). It is defined as the *weight of the standard of mass at standard acceleration of gravity*.

It can readily be shown that

$$1 \text{ kgf} = 9.80665 \text{ N}$$

because it follows from the definition of mass that $P = mg$. Then, substituting $m = 1 \text{ kg}$ and $g = 9.80665 \text{ m/s}^2$, we find that a body with a mass of 1 kg weighs 9.80665 N at standard acceleration of gravity. But this, according to our definition, is the kilogram-force.

Mass in the mk(force)s system is a derived unit. This engineering unit of mass has been variously called the metric slug, the mug, the par, or the TME*. It is defined as the mass of a particle whose acceleration is 1 m/s² when the force on it is 1 kgf. Thus

$$1 \text{ TME} = 9.80665 \text{ kg}$$

It is assumed for approximate calculations that 1 kgf = 9.81 N and 1 TME = 9.81 kg.

Chapter 8

EQUATIONS OF MOTION AND INITIAL CONDITIONS

8.1. THE BASIC PROBLEM OF DYNAMICS

1. Very many problems in astronomy, transportation, artillery, and other fields of science and engineering are stated as follows: given a body that can be treated as a material point, or particle (for instance, a planet, projectile, rocket, etc.), and the forces acting

* From *Technische Mass Einheit*, the German for Engineering Mass Unit. This unit is used in German-speaking countries.

on this body. Find the equation of motion of the particle, i.e. express its coordinates in the form of definite functions of time.

This is the *basic problem of dynamics*, which can be concisely stated thus: *determine the equations of motion of a particle when the forces acting on it are known.*

To solve this problem we begin by determining the acceleration of the particle by means of the fundamental law of dynamics (Newton's second law). Then, using the formulas of kinematics, we find the expressions for the velocities and coordinates.

2. Mathematical difficulties that arise in solving the problem in its general form may be formidable. At the same time a particular solution of any problem can be obtained by numerical methods. Such solutions may be approximate, but the calculations can be carried out to any degree of accuracy specified beforehand.

Two simple problems will be dealt with below. One will be solved in its general form, the other by numerical methods. We shall try to find what is needed to solve the basic problem of dynamics.

We can obtain exact solutions only in the simplest cases, for instance in calculating the motion of a body subject to a constant force. As for numerical methods, they are suitable for solving any problem but always require a great many arithmetical operations.

Today such operations are performed by electronic computers, which readily solve such problems when properly programmed.

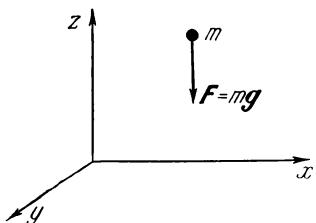


Fig. 8.1

8.2. MOTION OF A PARTICLE SUBJECT TO THE FORCE OF GRAVITY

1. Suppose that no other force but that of gravity acts on a particle. This particle could be a body on the moon, which has no atmosphere. But, with a certain degree of approximation such a solution will be suitable for the earth as well if the velocity of the particle does not exceed several metres per second. At such velocities air resistance can be neglected.

We shall regard the reference frame fixed on the earth's surface as inertial. The coordinate axes will be located as shown in Fig. 8.1. The expression for the force acting on the particle is

$$\mathbf{F} = \mathbf{P} = mg \quad (8.1)$$

or in projections on the coordinate axes

$$F_x = 0, \quad F_y = 0, \quad \text{and} \quad F_z = -mg \quad (8.2)$$

The fundamental law of dynamics is of the form

$$F_x = ma_x = 0, \quad F_y = ma_y = 0, \quad \text{and} \quad F_z = ma_z = -mg \quad (8.3)$$

Hence, the projections of the acceleration are

$$a_x = 0, \quad a_y = 0, \quad \text{and} \quad a_z = -g \quad (8.4)$$

2. There should be no difficulty now in finding the law of motion. The particle moves without acceleration, i.e. uniformly, along the axis of abscissas (the x -axis). Its velocity along the x -axis is constant and its x -coordinate is a linear function of time (see Sec. 1.4). Thus

$$\left. \begin{array}{l} v_x = u_0 = \text{const} \\ x = x_0 + u_0 t \end{array} \right\} \quad (8.5)$$

Similarly, for the axis of ordinates (the y -axis)

$$\left. \begin{array}{l} v_y = w_0 = \text{const} \\ y = y_0 + w_0 t \end{array} \right\} \quad (8.6)$$

But along the z -axis the particle moves with constant acceleration. Then (see Sec. 4.3 through 4.6) its velocity is a linear function of time and its coordinate is a quadratic function of time:

$$\left. \begin{array}{l} v_z = v_0 + a_z t = v_0 - gt \\ z = z_0 + v_0 t + \frac{a_z t^2}{2} = z_0 + v_0 t - \frac{gt^2}{2} \end{array} \right\} \quad (8.7)$$

3. We have thus found the motion equation:

$$\left. \begin{array}{l} x = x_0 + u_0 t \\ y = y_0 + w_0 t \\ z = z_0 + v_0 t - \frac{gt^2}{2} \end{array} \right\} \quad (8.8)$$

We obtained a minus sign for the acceleration in equations (8.8) while in equation (4.8) the sign was plus. This is due to the direction of the z -axis. In equation (4.8) the acceleration is in the direction of the z -axis; in (8.8) the acceleration vector is opposite in direction.

The motion equation contains six arbitrary constants: the *initial coordinates* x_0 , y_0 and z_0 and the *initial velocities* u_0 , w_0 and v_0 . They specify the position and velocity of the particle at the instant of time $t = 0$, which is taken as the initial instant.

4. The system of equations (8.5) through (8.8) describes *all possible cases* of the motion of a particle subject to the force of gravity. For any particular solution we must indicate the initial conditions.

Suppose that a body is thrown upward from the origin of coordinates at an angle α to the horizon at a velocity V_0 , and that V_0 lies in the xz -plane (Fig. 8.2). Then the initial conditions are of the form

$$w_0 = 0, \quad u_0 = V_0 \cos \alpha \quad \text{and} \quad v_0 = V_0 \sin \alpha \quad (8.9)$$

The motion equation for our case is

$$y = 0, \quad x = V_0 t \cos \alpha \quad \text{and} \quad z = V_0 t \sin \alpha - \frac{gt^2}{2} \quad (8.10)$$

We shall show that the body moves along a parabola lying in the vertical xz -plane. From the second equation of (8.10) it follows that $t = x/V_0 \cos \alpha$. Substituting into the third equation we obtain

$$z = \frac{V_0 x \sin \alpha}{V_0 \cos \alpha} - \frac{gx^2}{2V_0^2 \cos^2 \alpha}$$

or

$$z = x \tan \alpha - \frac{gx^2}{2V_0^2 \cos^2 \alpha} \quad (8.11)$$

This is the general form of the equation of a parabola. To find the horizontal range we assume in equation (8.11) that $z = 0$ and $x = L$.

Then

$$L = \frac{2V_0^2 \tan \alpha \cos^2 \alpha}{g} = \frac{V_0^2 \sin 2\alpha}{g} \quad (8.12)$$

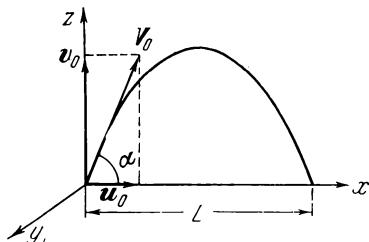


Fig. 8.2

The maximum range is reached when the body is thrown upward at an angle of 45° because in this case $\sin 2\alpha = \sin 90^\circ = 1$.

5. Equations for the following particular cases are to be derived by the student as homework:

(a) A body at a certain height h above the earth is thrown horizontally with an initial velocity u_0 . Find the motion equation, trajectory, and horizontal range.

(b) A body is thrown vertically upward from the origin of coordinates with an initial velocity v_0 . Find the motion equation, trajectory, maximum height reached, and time required to reach this height.

(c) A body is thrown upward at an angle α to the horizontal with a velocity V . Find the motion equation and trajectory.

8.3. NUMERICAL SOLUTION OF THE BASIC PROBLEM OF DYNAMICS

1. We had no difficulty in the preceding section in finding the general law of motion (the motion equation) only because the force applied to the particle was constant. If this force is variable, higher mathematics will be required to find the general law of motion. But a particular solution for specified initial conditions can always be obtained by numerical methods. Here the solution is reduced to quite elementary, though tedious, mathematical operations.

2. The calculations will be based on the following considerations. The average acceleration during the time interval $\Delta t = t_n - t_{n-1}$ is determined from the relationship (see Sec. 4.1)

$$a_{av} = \frac{v_n - v_{n-1}}{\Delta t}$$

Assuming (for a small Δt) that the average acceleration differs little from the instantaneous acceleration at the final instant $t = t_n$, we obtain approximately

$$v_n \approx v_{n-1} + a_n \Delta t \quad (8.13)$$

We obtain a similar equation for the average velocity

$$\bar{v}_n \approx \bar{v}_{n-1} + a_n \Delta t \quad (8.14)$$

in which the average velocity is

$$\bar{v}_n \approx \frac{v_n + v_{n-1}}{2}$$

But according to definition (see Sec. 1.6) the average velocity is

$$\bar{v}_n = \frac{\Delta x}{\Delta t} = \frac{x_n - x_{n-1}}{\Delta t}$$

from which

$$x_n = x_{n-1} + \bar{v}_n \Delta t \quad (8.15)$$

Now we have all that is required for a numerical calculation of the motion equation. If we know the coordinate and velocity of a particle at a certain initial instant we can calculate the acceleration by using the fundamental law of dynamics. Then, using equation (8.13), we calculate the velocity after Δt seconds have elapsed, and by equation (8.15) the coordinate. Assuming that the new velocity and coordinate are the initial values, and making use of the same equations, we take another step and find a newer velocity and coordinate, and so on, step by step. After a finite number of such steps we find the velocity and coordinate for the required instant of time.

Obviously, the smaller Δt , the more accurate the result. But this will require a greater amount of calculations.

8.4. MOTION OF A BODY SUBJECT TO AN ELASTIC FORCE

1. In this section we shall demonstrate the application of the numerical method for solving the principal problem of mechanics using as an example the motion of a particle subject to the action of an elastic force.

Let the elastic force $\mathbf{F} = -k\mathbf{x}$ (Fig. 8.3) act on a particle. Then the fundamental law of dynamics will take the form

$$-kx = ma \quad (8.16)$$

or, if projected on the coordinate axes,

$$-kx = ma_x, \quad 0 = ma_y \quad \text{and} \quad 0 = ma_z \quad (8.17)$$

Assume further that $y_0 = z_0 = 0$ and $w_0 = u_0 = 0$. Hence, the particle moves only along the x -axis with the acceleration

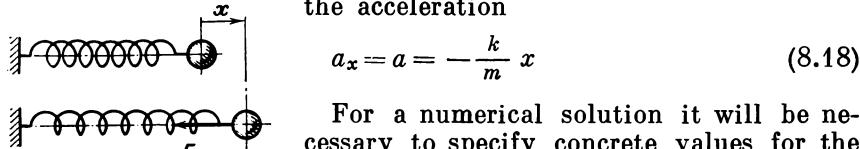


Fig. 8.3

$$a_x = a = -\frac{k}{m} x \quad (8.18)$$

For a numerical solution it will be necessary to specify concrete values for the constant k/m , initial coordinate x_0 , and initial velocity v_0 . Let

$$k/m = 2.467 = \pi^2/4, \quad x_0 = 1.000, \quad \text{and} \quad v_0 = 0 \quad (8.19)$$

The calculations will be based on a time interval $\Delta t = 0.1$ s.

It follows from equation (8.18) that the acceleration during the first time interval is approximately

$$a_0 = -2.467x_0 = -2.467$$

The final velocity after $\Delta t = 0.1$ s is

$$v_0^{fin} = v_0 + a_0 \Delta t = -2.467 \times 0.1 = -0.2467$$

The average velocity during this time interval is

$$\bar{v}_0 = \frac{v_0 + v_0^{fin}}{2} = \frac{0 - 0.2467}{2} = -0.1234$$

The coordinate of the particle at the end of the time interval is

$$x_1 = x_0 + \bar{v}_0 \Delta t = 1.000 - (0.1234 \times 0.1) = 0.988$$

These calculations can be continued in a similar way (Table 8.1).

2. There is no point in continuing the calculations beyond the value $t = 1$ because from this instant the values of the x -coordinate are repeated but with the opposite sign. Therefore, at $t = 2$ s the coordinate of the particle is $x = -1$, and at $t = 3$ s it is $x = 0$. Finally, at $t = 4$ s the coordinate and velocity are the same as at $t = 0$.

The graph showing the dependence of the coordinate of the particle on time (Fig. 8.4) greatly resembles the graph of the function $x = -A \cos \omega t$, where $\omega = \sqrt{k/m}$. Also shown is the graph of the instantaneous velocity of the particle, which is very much like the graph of the function $v = B \sin \omega t$. This is due to the fact that

Table 8.1

t	x	$\frac{a}{x} = -2.467$	\bar{v}	Calculations
0.0	1.000	-2.467	-0.123	$v_0 = 0$ $v_0^{\text{fin}} = v_0 + a_0 \Delta t = -0.246$
0.1	0.988	-2.437	-0.367	$x_1 = x_0 + \bar{v}_0 \Delta t = 1.000 - 0.012 = 0.988$ $\bar{v}_1 = \bar{v}_0 + a_1 \Delta t = -0.123 - 0.244 = -0.367$
0.2	0.951	-2.347	-0.602	$x_2 = x_1 + \bar{v}_1 \Delta t = 0.988 - 0.037 = 0.951$ $\bar{v}_2 = \bar{v}_1 + a_2 \Delta t = -0.367 - 0.235 = -0.602$
0.3	0.891	-2.199	-0.822	$x_3 = x_2 + \bar{v}_2 \Delta t = 0.951 - 0.060 = 0.891$ $\bar{v}_3 = \bar{v}_2 + a_3 \Delta t = -0.602 - 0.220 = -0.822$
0.4	0.809	-1.996	-1.022	$x_4 = x_3 + \bar{v}_3 \Delta t = 0.891 - 0.082 = 0.809$ $\bar{v}_4 = \bar{v}_3 + a_4 \Delta t = -0.822 - 0.200 = -1.002$
0.5	0.707	-1.743	-1.196	$x_5 = x_4 + \bar{v}_4 \Delta t = 0.809 - 0.102 = 0.707$ $\bar{v}_5 = \bar{v}_4 + a_5 \Delta t = -1.022 - 0.174 = -1.196$
0.6	0.587	-1.450	-1.341	$x_6 = x_5 + \bar{v}_5 \Delta t = 0.707 - 0.120 = 0.587$ $\bar{v}_6 = \bar{v}_5 + a_6 \Delta t = -1.196 - 0.145 = -1.341$
0.7	0.453	-1.120	-1.453	$x_7 = x_6 + \bar{v}_6 \Delta t = 0.587 - 0.134 = 0.453$ $\bar{v}_7 = \bar{v}_6 + a_7 \Delta t = -1.341 - 0.112 = -1.453$
0.8	0.308	-0.722	-1.525	$x_8 = x_7 + \bar{v}_7 \Delta t = 0.453 - 0.145 = 0.308$ $\bar{v}_8 = \bar{v}_7 + a_8 \Delta t = -1.453 - 0.072 = -1.525$
0.9	0.156	-0.380	-1.563	$x_9 = x_8 + \bar{v}_8 \Delta t = 0.308 - 0.152 = 0.156$ $\bar{v}_9 = \bar{v}_8 + a_9 \Delta t = -1.525 - 0.038 = -1.563$
1.0	0.000	0.000	-1.563	$x_{10} = x_9 + \bar{v}_9 \Delta t = 0.156 - 0.156 = 0.000$ $\bar{v}_{10} = \bar{v}_9 + a_{10} \Delta t = -1.563 - 0.000 = -1.563$
1.1	-0.156	0.380	-1.525	$x_{11} = x_{10} + \bar{v}_{10} \Delta t = 0.000 - 0.156 = -0.156$ $\bar{v}_{11} = \bar{v}_{10} + a_{11} \Delta t = -1.563 + 0.038 = -1.525$
1.2	-0.308	0.722	-1.453	$x_{12} = x_{11} + \bar{v}_{11} \Delta t = -0.156 - 0.152 = -0.308$ $\bar{v}_{12} = \bar{v}_{11} + a_{12} \Delta t = -1.525 + 0.072 = -1.453$

a precise solution of the problem by methods used in higher mathematics leads to the following:

$$x = \cos \frac{\pi}{2} t \text{ and } v = -\frac{\pi}{2} \sin \frac{\pi}{2} t \quad (8.20)$$

Then, expressing the variable angle $\alpha = (\pi/2) t$ in degrees instead of radians, we can find the corresponding values of the trigonometric functions in equations (8.20) from a table. After solving the equations

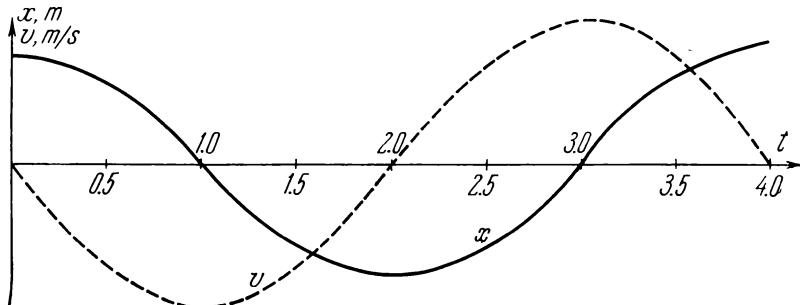


Fig. 8.4

we can compare the results with our previous calculations (Table 8.2). Note that equation (8.20) gives the *instantaneous* value of the velocity at the beginning of each time interval, while in the previous calculations we found the *approximate value of the average velocity* during this interval.

Table 8.2

t, s	$\alpha = \frac{\pi}{2} t = 90^\circ t$	Calculated by formulas		Calculated step-by-step numerically	
		$x = \cos \frac{\pi}{2} t$	$v = -\frac{\pi}{2} \sin \frac{\pi}{2} t$	x	\bar{v}
0	0	1.0000	0.0000	1.000	-0.123
0.1	9°	0.9877	-0.2457	0.988	-0.367
0.2	18°	0.9511	-0.4854	0.951	-0.602
0.3	27°	0.8910	-0.7131	0.891	-0.822
0.4	36°	0.8090	-0.9233	0.809	-1.022
0.5	45°	0.7071	-1.1107	0.707	-1.196
0.6	54°	0.5878	-1.2708	0.587	-1.341
0.7	63°	0.4540	-1.3996	0.453	-1.453
0.8	72°	0.3090	-1.4938	0.308	-1.525
0.9	81°	0.1564	-1.5515	0.156	-1.563
1.0	90°	0.0000	-1.5708	0.000	-1.563
1.1	99°	-0.1564	-1.5515	-0.156	-1.525

We should note that the calculated data were obtained by dividing the time interval into 100 rather than 10 parts, and that rounded-off values were listed in the table. As we can see, the coordinates calculated by the formula and numerically coincide fairly well.

3. The motion of a particle subject to an elastic force is called simple *harmonic motion*. This kind of motion will be discussed in detail in Chapter 49.

8.5. QUANTITIES DETERMINING THE MOTION EQUATION OF A PARTICLE

1. In the problems discussed in the preceding sections we found that the following data are required to determine the motion equation of a particle:

(a) we must know the *force* acting on the particle; it can be given as a function of either time or the coordinates;

(b) we must know the *initial conditions*: the coordinates and velocity at a certain instant of time. (The initial momentum $p_0 = mv_0$ may be specified instead of the initial velocity.)

If these quantities are known we can always find the motion equation of the particle, i.e. express its coordinates as a function of time. This in turn enables us to determine the behaviour of the particle at any previous or subsequent instant of time.

Thus, if we know the forces of interaction between the sun and the planets, as well as the coordinates and velocities of the planets at some instant of time (i.e. the initial conditions), we can determine their state of motion many centuries ago and predict their position for many centuries to come. Solar and lunar eclipses, favourable oppositions of Mars (i.e. the instants of time when the distance between Mars and the earth is least), and other such phenomena can be predicted by these methods.

If we know the velocity of a space rocket at the instant the fuel burns out and its coordinates at some instant of time as well as the forces acting on it, we can calculate its trajectory, determine its location at any instant of time, and predict the time and coordinates of its landing.

2. If the true motion of a particle does not coincide with its calculated motion, then either the initial conditions have been incorrectly determined or the force has been set incorrectly. It is usually possible to determine the degree of error from the disparity between the calculated and true paths. As an example let us consider the history of the discovery of the planets Neptune and Pluto.

In 1781 the English astronomer Sir William Herschel (1738-1822) designed and built what was a huge telescope for his time and discovered a new, the seventh, planet, which was named Uranus. On the basis of the action exerted on Uranus by the sun and the other

planets of the solar system known at that time (Mercury, Venus, Earth, Mars, Jupiter and Saturn), the law of motion and path of Uranus were determined. The calculated path proved to be wrong: Uranus travelled along a different path. Since the initial conditions had been correctly determined, the only reasonable proposition was to assume that not all the forces acting on Uranus had been considered.

The French astronomer U. J. Leverrier (1811-1877) and the English astronomer J. C. Adams (1819-1892) suggested, independently of each other, that there must be another planet beyond Uranus, which nobody had yet observed. From the disparity between the calculated path of Uranus and its true (observed) path they were able to determine the law of motion of this unknown planet and predict at what point in the sky it could be found at any definite instant of time. These calculations were brilliantly confirmed by observations. In 1846 the German astronomer J. G. Galle (1812-1910) directed a telescope at a predicted point in the sky and actually discovered the planet that was later named Neptune.

At the beginning of the 20th century the American astronomer P. Lowell (1855-1916) on the basis of more precise observations came to the conclusion that the difference between the calculated path of Uranus and the true nature of its motion could not be explained only by the action of Neptune, and that beyond Neptune there was still another, ninth planet. In 1930, after Lowell's death and in the observatory he had founded, the American astronomer C. W. Tombaugh (b. 1906) discovered this planet, which was named Pluto.

Using a number of problems of Newtonian mechanics as examples, we have found that the *applied force and the initial conditions fully determine the motion equation* of a particle.

The further development of science showed that in dealing with objects in the subatomic domain it is impossible to preset the kind of initial conditions used in Newtonian mechanics (see Ch. 14).

Chapter 9

GRAVITATION

9.1. DISCOVERY OF THE LAW OF GRAVITATION

1. By the beginning of the 17th century most scientists had accepted the validity of the heliocentric theory of the solar system. According to this theory, advanced by Copernicus, the earth and all the other planets revolve about the sun, which is the centre of our planetary system. But neither the laws of motion of the planets nor the causes of the nature of their motion were clear to the scientists of his time.

The German astronomer and mathematician Johannes Kepler (1571-1630) analyzed the extensive observational data on planetary motion compiled by the Danish astronomer Tycho Brahe (1546-1601) and by himself, and deduced the laws of motion of planets about the sun. It was clear to him that these laws could be explained only after finding the forces acting on the planets. But neither he nor his contemporaries were able to do this. The problem was finally solved by the great English scientist Sir Isaac Newton and set forth in his book *Philosophiae Naturalis Principia Mathematica* (Physics was then called natural philosophy), published in 1687.

2. As a first approximation we may assume that the planets revolve with almost uniform motion in orbits that differ only slightly from circles. But in the circular motion of a planet (which we regard as a mass point) there is a normal (centripetal) acceleration directed toward the centre of the orbit, where the sun is located. We know from the fundamental law of dynamics that this acceleration is caused by a force. Hence, the sun acts on each planet with a force directed toward the centre of the sun (Fig. 9.1). From Newton's third law it follows that the planet in turn acts on the sun with an equal force but in the opposite direction.

3. We know that the moon revolves about the earth. Evidently, there is a similar force of attraction acting here: the moon is attracted to the earth and the earth to the moon.

Newton advanced the decisive proposition that the force of gravity acting on a body near the surface of the earth and the force with which the earth attracts the moon are of exactly the same origin. This he demonstrated by comparing the acceleration of free fall with the normal acceleration with which the moon revolves on its orbit.

This normal acceleration can be calculated from the following data. The distance from the earth to the moon is $r = 384\,403\text{ km} = 3.84 \times 10^8\text{ m}$ and the moon's period of revolution is $T = 27.322\text{ days} = 27.3 \times 24 \times 3600\text{ s}$. The orbital speed is $v = 2\pi r/T$ and the normal acceleration

$$a_n = \frac{v^2}{r} = \frac{4\pi^2 r}{T^2} = \frac{4\pi^2 \times 3.84 \times 10^8}{(27.3 \times 24 \times 3600)^2} = 2.72 \times 10^{-3}\text{ m/s}^2 \quad (9.1)$$

As we can see, the normal acceleration of the moon and the acceleration of gravity differ greatly in magnitude. Newton, however, explained the difference in these values by putting forward the hypothesis that the force of gravitational attraction decreases with an increase in the distance between the interacting bodies according to a definite law, which we shall now derive. We know that the free-fall acceleration equals 9.81 m/s^2 near the earth's surface, i.e. when the distance from the centre of the earth to the body being attracted is equal to the mean radius of the earth ($R = 6371\text{ km} = 6.37 \times 10^6\text{ m}$). The

normal acceleration of the moon, on the other hand, was obtained for a *point on its orbit*, when the distance between the interacting

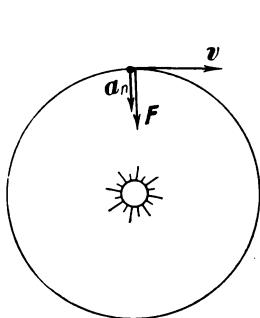


Fig. 9.1

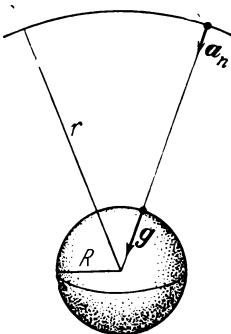


Fig. 9.2

bodies equals the radius of the lunar orbit $r = 3.84 \times 10^8$ m (Fig. 9.2). Setting up the proportion

$$\frac{g}{a_n} = \left(\frac{r}{R} \right)^m \quad (9.2)$$

we can determine the unknown power. Substituting the distances and accelerations, we obtain

$$\frac{9.81}{2.72 \times 10^{-3}} = \left(\frac{3.84 \times 10^8}{6.37 \times 10^6} \right)^m$$

or $3600 = 60^m$, from which $m = 2$.

Thus, the acceleration due to the force of gravitational attraction decreases in inverse proportion to the square of the distance. From equation (9.2) with $m = 2$ it follows that

$$a_n = \frac{gR^2}{r^2} = \frac{K}{r^2} \quad (9.3)$$

where K is a constant.

9.2. NEWTON'S LAW OF UNIVERSAL GRAVITATION

1. The results obtained by Newton from an analysis of the normal acceleration due to the revolution of the moon about the earth led him to the conclusion that all bodies in nature are attracted to one another with a certain force, called the *force of gravitational attraction*, and that this force decreases in inverse proportion to the square of the distance between the interacting bodies.

Suppose that two bodies of masses m_1 and m_2 are located at a distance r from each other. They interact with forces $F_1 = m_1 a_1$ and

$F_2 = m_2 a_2$, and $F_1 = F_2$ according to Newton's third law. Taking equation (9.3) into account, we have

$$\frac{m_1 K_1}{r^2} = \frac{m_2 K_2}{r^2} \quad (9.4)$$

This equality will still hold if we assume that $K_1 = Gm_2$ and $K_2 = Gm_1$, where G is a constant. Then

$$F = m_1 a_1 = \frac{m_1 K_1}{r^2} = G \frac{m_1 m_2}{r^2}$$

Hence

$$F = G \frac{m_1 m_2}{r^2} \quad (9.5)$$

The law of universal gravitation can be stated as follows:

The force of gravitational attraction between two mass points (particles) is proportional to the product of their masses and inversely proportional to the square of the distance between them.

2. The above discussion can in no way be regarded as a deduction of the law of universal gravitation. It can only serve to illustrate Newton's line of reasoning.

To test this law we must be able to determine the force of gravitational attraction between bodies of finite dimensions as well as between particles. In the general case this problem poses great mathematical difficulties. It cannot be solved by elementary methods. Precise calculations show that formula (9.5) can also be used to find the force of interaction between *uniform spheres* of masses m_1 and m_2 , with r as the distance between their centres.

3. Factor G in the law of universal gravitation is called the *gravitational constant*. In magnitude it equals the force of interaction of particles of unit mass located at unit distance from each other.

In SI units the gravitational constant is equal in magnitude to the force of interaction between two particles with the masses $m_1 = m_2 = 1 \text{ kg}$, located at the distance $r = 1 \text{ m}$ from each other. Its value can be determined only by experiment.

9.3. THE CAVENDISH EXPERIMENT

In 1798 the English scientist Henry Cavendish (1731-1810) performed an experiment to measure the force of gravitational attraction in the laboratory and thereby determine the gravitational constant. His apparatus consisted of a *torsion balance*, whose principle is clear from Fig. 9.3.

A light horizontal rod is suspended by a fine quartz thread to which a small mirror is attached. A beam of light falling on this mirror is reflected onto a scale. When the rod turns about the vertical axis, the reflected beam moves along the scale, indicating the angle

of twist of the thread. Two small lead balls each of mass m are attached to the ends of the rod. Two large lead balls, each of mass M , are placed symmetrically, on opposite sides, near the small balls at the ends of the rod. The attraction between the large and small balls twists the rod and thread until the elastic force of the twisted thread counterbalances the gravitational interaction between the lead

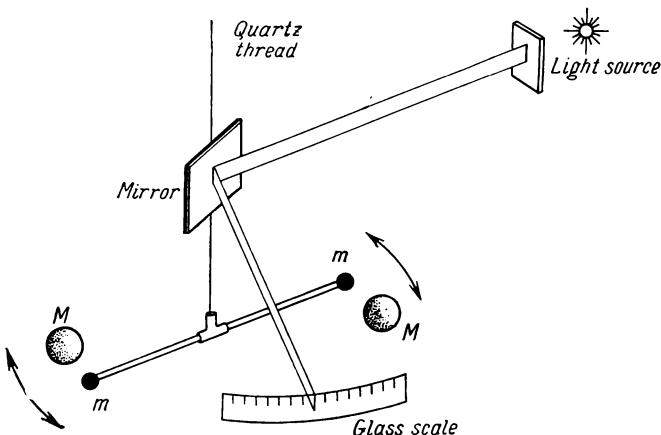


Fig. 9.3

balls. Measuring the angle of twist of the thread, the force of interaction between the balls can be determined. Then, knowing the masses of the balls and the distance between their centres, we can calculate the gravitational constant.

The most accurate measurements, made by improved techniques, yielded the following result:

$$G = (6.6732 \pm 0.0031) \times 10^{-11} \text{ N}\cdot\text{m}^2/\text{kg}^2$$

9.4. DETERMINING THE DISTANCES FROM THE SUN TO THE PLANETS

1. Let us apply the law of universal gravitation to calculate the distances from the sun to the planets. To simplify the problem we shall assume that the planets revolve in circular orbits. This will have only a slight effect on the results because the true orbits of the planets differ only slightly from circles.

The force of gravitational attraction exerted by the sun on the planets produces the normal acceleration $a_n = F/m = GM/r^2$. But, since $a_n = v^2/r = 4\pi^2r/T^2$, then

$$\frac{GM}{r^2} = \frac{4\pi^2r}{T^2} \quad (9.6)$$

where M = mass of the sun

r = radius of the planet's orbit

T = period of revolution of the planet about the sun. Then

$$\frac{r^3}{T^2} = \frac{GM}{4\pi^2} = \text{const} \quad (9.7)$$

Thus, the cubes of the mean distances of the planets from the sun are proportional to the squares of their orbital periods (Kepler's third law).

Since the periods of revolution of the planets can be measured with great accuracy, the distances from the planets to the sun can be determined by measuring the distance from any one planet. The average distance from the earth to the sun has been adopted as such a standard length. It is called the *astronomical unit* (AU) and is equal to 149.5×10^9 m.

In 1960-1962, by means of powerful radar apparatus, Soviet scientists measured the distance from the earth to Venus. This enabled them to determine more precisely the value of the astronomical unit and, thereby, the distances to all the planets of the solar system. Excellent agreement of the results of astronomical and radar measurements is good proof of the validity of Newton's law of universal gravitation.

9.5. THE GRAVITATIONAL FIELD

1. In modern physics the mechanism of gravitational interaction is conceived of as follows. Each particle of mass M sets up a field which modifies the surrounding space. When a test body of mass m is placed in this field, the field exerts a force F on it that depends upon the properties of the field at this point and on the mass of the test body. Naturally, the test body sets up its own field in the space around it and this, in turn, exerts a force on the body of mass M .

Space free of matter has a number of geometric and physical properties. For example, the shortest distance between two points is a straight line, time flows identically at all points of this space, beams of light are straight lines, etc.

According to Einstein, when a body of mass M is brought into a region of space, it changes the properties of the space. The shortest distance between two points will no longer be a straight line. It will be a curve whose shape depends upon the mass of the body and the distances from the body to the investigated points in space. Light still propagates along the shortest distance between two points, but now the light beams are curved. In addition, time is slowed down in the vicinity of massive bodies.

Thus, a gravitational field reflects the changes in the physical and geometric properties of space in the vicinity of massive bodies (see Sec. 24.5 and 24.6).

The gravitational field is material. It exists independently of our consciousness and can be detected by its action on physical objects, for instance, measuring instruments. The gravitational field is a form of matter.

9.6. GRAVITATIONAL FIELD INTENSITY

1. To provide a qualitative description of a gravitational field we introduce a physical quantity called the *intensity* I of the gravitational field. The field intensity is equal to the ratio of the force of gravitational attraction exerted on a test body to the mass of this body. Thus

$$I = \frac{F}{m} \quad (9.8)$$

Making use of the law of universal gravitation [equation (9.5)] and substituting M for m_1 and m for m_2 , we obtain

$$I = \frac{GM}{r^2} \quad (9.9)$$

2. As we can see, the gravitational field intensity is determined by the mass of the body setting up the field (source of the field) and the distance from this body to the point in the field that interests us. It does not depend upon the mass of the test body placed at this point.

Evidently, the gravitational field intensity coincides in its physical sense with the acceleration of the test body. Since the field intensity does not depend upon the mass of the test body, all bodies, regardless of their mass, move with the same acceleration at a given point of the gravitational field. This, in particular, is why all bodies fall at a given point in a gravitational field with the same acceleration (see Sec. 6.2).

3. The field intensity is a vector having the direction of the force of gravitational attraction. If a radius vector r is drawn from the source M of the field to the point being investigated in the field, the field intensity vector always has the opposite direction. To show this, equation (9.9) should be written with a minus sign. Thus

$$I = -G \frac{M}{r^2} \quad (9.10)$$

9.7. THE EARTH'S GRAVITATIONAL FIELD

1. As a body moves away from the surface of the earth, the force of gravity and the free-fall acceleration decrease. A body at sea level is attracted to the earth with the force

$$P_0 = mg_0 = G \frac{Mm}{R^2} \quad (9.11)$$

where M and R = mass and radius of the earth

m = mass of the body

g_0 = free-fall acceleration at sea level.

A body raised to the altitude h above sea level is attracted to the earth with the force

$$P = mg = G \frac{Mm}{(R+h)^2} \quad (9.12)$$

It follows then that the force of gravity and the free-fall acceleration vary with altitude to the same degree.

At low altitudes, much less than the earth's radius, these variations can be neglected and the force of gravity and the free-fall acceleration can be considered constants. In fact, since the mean radius of the earth $R = 6371$ km, even at several hundred kilometres above sea level only very sensitive instruments can detect variations in these values. Let $h = 300$ km, then

$$\frac{R^2}{(R+h)^2} = \left(\frac{6371}{6671}\right)^2 = 0.955$$

which differs from unity by 4.5 per cent. At $h = 30$ km

$$\frac{R^2}{(R+h)^2} = \left(\frac{6371}{6401}\right)^2 = 0.9957$$

which differs from unity by only 0.43 per cent.

2. If we know the free-fall acceleration and the gravitational constant, we can determine the mass of the earth. In fact, from equation (9.11) we can write

$$g_0 = \frac{GM}{R^2}$$

from which the mass of the earth is

$$M = \frac{g_0 R^2}{G} = \frac{9.8 \times 6.372 \times 10^{12}}{6.67 \times 10^{-11}} = 5.97 \times 10^{24} \text{ kg} \quad (9.13)$$

In a similar way we can find the mass of the sun. Knowing that the radius of the earth's orbit is $R_e = 149.5 \times 10^9$ m and the period of the earth's revolution about the sun is $T = 1$ year = 31.56×10^6 s, we can calculate the normal acceleration of the earth. Thus

$$a_e = \frac{4\pi^2 R_e}{T^2} = 5.9 \times 10^{-3} \text{ m/s}^2$$

By analogy with equation (9.13) we have

$$M_{\text{sun}} = \frac{a_e R_e^2}{G} = \frac{5.9 \times 10^{-3} \times 149.5^2 \times 10^{18}}{6.67 \times 10^{-11}} = 1.98 \times 10^{30} \text{ kg}$$

Thus, the sun's mass is 330 000 times that of the earth.

3. Now, when we know the mass of the earth and its radius, we can find its average density:

$$\rho_{av} = \frac{M}{V} = \frac{3M}{4\pi R^3} = \frac{3 \times 5.97 \times 10^{24}}{4\pi \times 6.378 \times 10^{18}} = 5.5 \times 10^3 \text{ kg/m}^3$$

Since the rock of which the surface layer of the earth consists has only about half this density, the earth evidently has a core whose density is approximately twice its average density. This means that the earth's core consists mainly of metals such as iron and copper.

According to a hypothesis advanced by the Soviet mathematician and geophysicist Academician O. Schmidt (1891-1956), the earth was formed from a meteoritic cloud. In colliding with one another the meteorites became heated; additional heat was evolved in the decay of radioactive substances. As a result, the earth acquired plastic properties. Denser substances, such as metals were attracted to the centre, where they formed the earth's core. The less dense substances came to the surface, where they formed the rock crust, called the *lithosphere*.

However, massive deposits of denser substances, metal ores, are frequently found in the lithosphere. Near these deposits the acceleration of gravity is greater than in regions consisting only of rock. Consequently, precise gravitational acceleration measurements can aid geologists in prospecting for ore deposits. This method, known as *gravitational prospecting*, is widely applied in practice.

9.8. EFFECT OF THE EARTH'S ROTATION ON FREE-FALL ACCELERATION

1. In the preceding section we derived the relationship between the free-fall acceleration and the altitude above sea level. But we took

no account of the effect of the daily rotation of the earth on this acceleration.

If the earth did not rotate about its axis, the force of gravity would coincide in magnitude and direction with the force of gravitational attraction. Moreover, the free-fall acceleration would coincide with the intensity of the earth's gravitational field.

Suppose a body of mass m is located at point A on the geographical latitude φ (Fig. 9.4). Since the body participates in the daily rotation

of the earth, it moves in a circle of radius $AK = r = R \cos \varphi$, where R is the radius of the earth. Any curvilinear motion has a

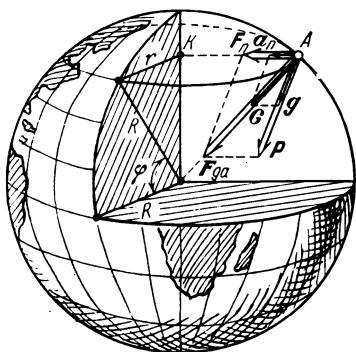


Fig. 9.4

normal acceleration. Thus

$$a_n = \frac{v^2}{r} = \frac{4\pi^2 R \cos \varphi}{T^2}$$

where T is the period of rotation of the earth.

To find the force that produces the normal acceleration we resolve the force of gravitational attraction F_{ga} into two components: the centripetal force F_n and the force of gravity P . From Fig. 9.4 we can see that at any place on the earth's surface, except the poles, the force of gravity is less than the force of gravitational attraction. Moreover, everywhere, except at the equator and the poles, the force of gravity is directed not to the centre of the earth but slightly to one side.

Owing to the fact that the force of gravity at the poles is somewhat greater than at the equator, the earth is not a true sphere but is flattened at the poles and bulges at the equator. Its polar radius equals 6356.9 km, and its equatorial radius equals 6378.4 km.

2. The free-fall acceleration acquires its maximum value at the poles. Here the normal acceleration is zero and the earth's radius has its minimum value. Consequently,

$$g_{\text{pole}} = I_{\text{pole}} = \frac{GM}{R_{\text{pole}}^2}$$

The free-fall acceleration at the equator equals the difference between the intensity of the gravitational field I and the normal acceleration. Thus

$$g_{\text{eq}} = I_{\text{eq}} - a_n^{\text{eq}} = \frac{GM}{R_{\text{eq}}^2} - \frac{4\pi^2 R_{\text{eq}}}{T^2}$$

Precise measurements show that

$$g_{\text{pole}} = 9.83210 \text{ m/s}^2 \quad \text{and} \quad g_{\text{eq}} = 9.78038 \text{ m/s}^2$$

At a latitude of 45° the free-fall acceleration of a body is $g = 9.80620 \text{ m/s}^2$.

For approximate calculations it is assumed that $g = 9.81 \text{ m/s}^2$.

Chapter 10

ELECTRIC FORCES

10.1. ELECTRIC CHARGE

1. Besides gravitational force there exists in nature a special class of *electromagnetic* interactions, the simplest example of which is the attraction or repulsion of electrified bodies.

The following experiment will demonstrate the electrification of bodies. Two plastic rods are mounted on needles so that they are free to turn as shown in Fig. 10.1. Attached to the end of one rod is a well ground and polished metal plate. The other rod carries a plate of plexiglass, also polished. First we remove the rods from their needles and bring the plates into contact. Now, when we replace the rods on their needles and release them, the plates are attracted to each other. This, of course, is not gravitational attraction because the masses of the bodies remain unchanged before and after contact, and gravitational forces depend only on the masses of bodies and the distance between them. In this experiment we observe a new class of forces known as *electric forces*.

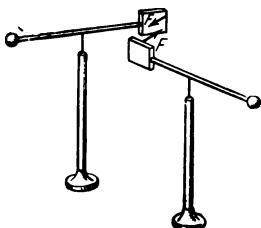


Fig. 10.1

2. The physical quantity that describes the property of bodies of showing electrical interaction under specific conditions and determining the magnitude of electric forces is called the *electric charge*. The phenomenon associated with the redistribution of charges on bodies is called *electrification*.

The electrification of bodies by contact, when they are rubbed together, is often called electrification by friction. This is wrong in principle. The rubbing plays no essential role; it only improves the contact between the bodies thus enhancing their electrification.

3. In all phenomena associated with the redistribution of electric charges in an insulated system of interacting bodies, the algebraic sum of the electric charges remains constant. This *law of conservation of electric charge* is just as basic as the other conservation laws, which will be dealt with later (see Ch. 15).

10.2. COULOMB'S LAW

1. Experiments show that the force of electrical interaction depends in a complex way on the shape of the electrified bodies and on how the charge is distributed on them. Hence, there is no one simple equation describing electrical interaction for every case. The law of interaction can be written in sufficiently simple form only for *point charges*.

The concept of a point charge is a convenient abstraction similar to that of the mass point (see Sec. 1.1). A point charge is the charge of a body whose size is much less than any distances found in the problem being considered.

2. Using a torsion balance (see Sec. 9.3), the French physicist Charles Augustin de Coulomb (1736-1806) experimentally discovered in 1785 the law of interaction of point charges. This torsion balance

(Fig. 10.2) had a light glass beam with a metal ball at one end. Coulomb touched this ball with an electrified rod that also had a ball on its end. This distributed the charge between the balls and they repelled each other. At this the beam twisted the suspension fibre until the elastic force counterbalanced the force of electrical interaction. By turning the handle on the suspension head, Coulomb could change the angle of twist of the fibre, thereby changing the elastic force and the distance between the charges. In a series of experiments Coulomb found that the force of electrical interaction between charges varies inversely proportional to the square of the distance between them.

Contrary to the Cavendish experiment, in which the masses of the interacting bodies were known, Coulomb had no way of measuring the charges on the balls. However, he devised the following procedure. He brought an uncharged ball of the same size into contact with the electrified ball and then removed the first to quite some distance from the torsion balance. Since this distributed the charge equally between the two contacting balls, the charge of the test ball was reduced by one half. Coulomb then found that the force of electrical interaction was also reduced by one half.

After repeating this operation several times, Coulomb came to the conclusion that the force of electrical interaction is directly proportional to the product of the charges of the interacting bodies.

3. Coulomb's experiments were not very exact because the balls were large and the force was measured with considerable error. Moreover, the experiments were conducted in air, which had an effect, though not very great, on the results. Nevertheless, Coulomb was able to formulate a law which was later brilliantly substantiated by all investigations of electrical phenomena. He stated that the *force of interaction between two point charges in a vacuum is directly proportional to the product of the two charges and inversely proportional to the square distance between them*. Thus

$$F = k \frac{q_1 q_2}{r^2} \quad (10.1)$$

where F = force of interaction

q_1 and q_2 = the charges

r = distance between them

k = proportionality factor depending upon the system of units used in calculations.

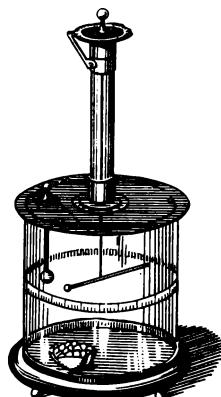


Fig. 10.2

4. The Coulomb force, like any other force (for instance, the gravitational force), obeys Newton's third law: the *forces of interaction between charges are equal in magnitude and act along the line joining the charges but in opposite directions.*

Experiments show that like charges repel and unlike charges attract each other. This is an essential difference between the Coulomb force and the gravitational force, which is always a force of attraction.

The force of repulsion exerted on a given charge has the same direction as the position vector drawn to the charge. The force of

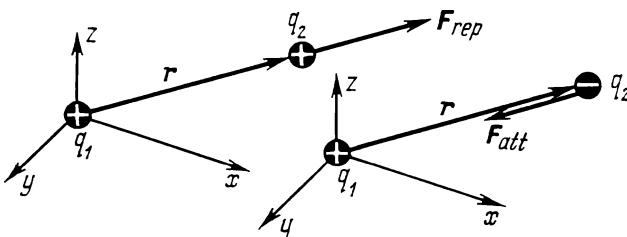


Fig. 10.3

attraction has the opposite direction (Fig. 10.3). Hence, it proves convenient to label forces of repulsion as positive and forces of attraction as negative.

Coulomb's law automatically takes the sign of the forces into account. The product of like charges is positive and, hence, the force of repulsion has the plus sign. But if the charges are unlike their product is a negative value as required for a force of attraction.

Accordingly, the equation of the law of gravitation should be written with a minus sign:

$$F = -G \frac{mM}{r^2} \quad (10.2)$$

because masses are always positive and a negative force, the force of attraction, is exerted between them.

10.3. UNITS OF CHARGE AND SYSTEMS OF UNITS

1. Coulomb's law acquires its simplest form if we assume that the proportionality factor equals unity. Then (for $q_1 = q_2 = q$)

$$F = \frac{q^2}{r^2} \quad (10.3)$$

But, in assuming that $k = 1$, we have already defined the unit of electric charge because the units of force and distance are specified.

The unit of charge in the cgs electrostatic system of units (cgse) is called the *electrostatic unit* (esu), or the *statcoulomb* (statcoul). Thus, one *electrostatic unit*, or one *statcoulomb*, is that point charge which when placed in a vacuum one cm from an equivalent charge exerts upon it a force of one dyne.

2. In SI units the unit of charge is the *coulomb* (C or coul):

$$1 \text{ C} = 3 \times 10^9 \text{ esu}$$

Thus, two point charges of one coulomb each, placed in a vacuum one m apart, interact with a force equal to

$$F = \frac{(3 \times 10^9)^2}{(10^2)^2} = 9 \times 10^{14} \text{ dynes} = 9 \times 10^9 \text{ N}$$

Since factor k in Coulomb's law (like gravitational constant G in the law of universal gravitation) is equal in magnitude to the force exerted by two unit charges at unit distance on each other,

$$k = 9 \times 10^9 \text{ N-m}^2/\text{C}^2$$

In SI units Coulomb's law is written as

$$F = \frac{q_1 q_2}{4\pi\epsilon_0 r^2} \quad (10.4)$$

The quantity ϵ_0 is called the *permittivity constant*. Combining equations (10.1) and (10.4), we obtain

$$\epsilon_0 = \frac{1}{4\pi k} = \frac{1}{36\pi \times 10^9} \text{ C}^2/\text{N-m}^2 = 8.85 \times 10^{-12} \text{ C}^2/\text{N-m}^2$$

3. It would seem natural to adopt the charge carried by the electron as the unit of charge. Then the charge of any body would be a whole number indicating the number of electrons this body had acquired (or lost) in electrification. However, since the unit of charge had already been adopted, the charge of the electron was determined experimentally. The electronic charge (elementary charge e) was found to be

$$e = 1.6 \times 10^{-19} \text{ C} = 4.8 \times 10^{-10} \text{ esu}$$

from which

$$1 \text{ C} = 6.25 \times 10^{18} e$$

4. Let us compare the forces of gravitational and electrical interaction between two electrons. We can write

$$F_{\text{Coul}} = \frac{e^2}{4\pi\epsilon_0 r^2} \quad \text{and} \quad F_{gr} = \frac{Gm^2}{r^2}$$

in which the rest mass of the electron $m = 9 \times 10^{-31} \text{ kg}$. Then

$$\frac{F_{\text{Coul}}}{F_{gr}} = \frac{e^2}{4\pi\epsilon_0 Gm^2} = \frac{1.6^2 \times 10^{-38} \times 36 \times 10^9}{4\pi \times 6.67 \times 10^{-11} \times 9^2 \times 10^{-62}} = 4.26 \times 10^{42}$$

Hence, in investigating phenomena that occur in electrically charged systems, particularly in the world of molecules, atoms and atomic nuclei, gravitational forces can be completely neglected in comparison with electric forces. It is only for cosmic bodies of huge mass (planets or stars) that gravitational forces begin to play a predominant role.

10.4. THE ELECTRIC DIPOLE

1. A dipole is an electric system of two unlike charges (positive and negative) of equal magnitude (Fig. 10.4). The electrical characteristic of the dipole is the vector

$$\mathbf{p}_e = q\mathbf{l} \quad (10.5)$$

It is called the electric-dipole moment. The vector \mathbf{l} is in the direction from the negative to the positive charge. The magnitude $|\mathbf{l}| = l$

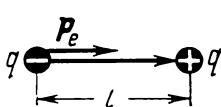


Fig. 10.4

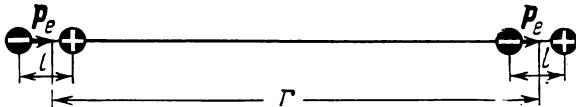


Fig. 10.5

is sometimes called the dipole arm. Vector \mathbf{p}_e has the same direction as vector \mathbf{l} .

Since the dipole is an electrically neutral system, it would seem that it is incapable of electrical interaction. This is a hasty conclusion, however. *Dipoles do interact*, because the charges forming a dipole are located at different points in space.

2. Consider the interaction of two identical dipoles located on a single axis. We shall denote the distance between the dipoles by r ; let this distance be much greater than the dipole arm: $r \gg l$ (Fig. 10.5).

The force of interaction is made up of four components: two forces of repulsion between like charges and two forces of attraction between unlike charges. Thus

$$\begin{aligned} F &= \frac{(-q)(-q)}{4\pi\epsilon_0 r^2} + \frac{(+\mathbf{q})(+\mathbf{q})}{4\pi\epsilon_0 r^2} + \frac{(-q)(+q)}{4\pi\epsilon_0 (r-l)^2} + \frac{(-q)(+q)}{4\pi\epsilon_0 (r+l)^2} \\ &= \frac{q^2}{4\pi\epsilon_0} \left[\frac{2}{r^2} - \frac{1}{(r-l)^2} - \frac{1}{(r+l)^2} \right] \end{aligned}$$

After simple transformations we obtain

$$F = -\frac{6q^2l^2}{4\pi\epsilon_0} \cdot \frac{r^2 - (l^2/3)}{r^2(r^2 - l^2)^2}$$

Substituting p_e for ql and neglecting l^2 because it is very small compared to r^2 , we can write

$$F = -\frac{6p_e^2}{4\pi\epsilon_0 r^4} \quad (10.6)$$

This expression can be readily generalized for the case of the interaction of different electric dipole moments \mathbf{p}_{1e} and \mathbf{p}_{2e} :

$$F = -\frac{6p_{1e}p_{2e}}{4\pi\epsilon_0 r^4} \quad (10.7)$$

Hence, if the dipole moments of two dipoles are located on a single straight line and have the same direction, they are attracted by each other, and the force of attraction is directly proportional to the product of the electric dipole moments and inversely proportional

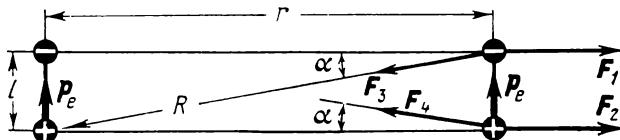


Fig. 10.6

to the fourth power of the distance between them. Consequently, dipole interaction decreases considerably faster with distance than the interaction between point charges.

We leave it to the student to show that dipoles, whose moments are located on a single straight line and have opposite directions, repel each other with a force of the same magnitude as in equation (10.7).

3. Let us calculate the force of interaction between two dipoles arranged as shown in Fig. 10.6. The resultant force is

$$F' = F_1 + F_2 - F_3 \cos \alpha - F_4 \cos \alpha = \frac{2q^2}{4\pi\epsilon_0 r^2} - \frac{2q^2 \cos \alpha}{4\pi\epsilon_0 R^2}$$

Since $\cos \alpha = r/R$ and $R^2 - r^2 = l^2$, after simple transformations we obtain

$$F' = \frac{2q^2 l^2 (R^2 + Rr + r^2)}{4\pi\epsilon_0 r^2 R^3 (R+r)}$$

Assuming, as above, that $l \ll r$ and, consequently, that $R \cong r$, we obtain

$$F' = \frac{2q^2 l^2 \times 3r^2}{4\pi\epsilon_0 \times 2r^6} = \frac{3p_e^2}{4\pi\epsilon_0 r^4} = \frac{F}{2} \quad (10.6')$$

It should be evident to the student that in the antiparallel arrangement of the dipole moments there will be a force of attraction of the same magnitude between them.

4. Comparing equations (10.6) and (10.6'), we see that the force of interaction of dipoles depends on their mutual position and separation. This is in contrast to central forces (such as gravitational and Coulomb forces), whose magnitude depends on the distance between the interacting bodies. Nuclear forces have similar properties (see Sec. 80.5).

Dipole interaction plays an exceptionally vital role in physics. It was found that any system consisting of asymmetrically arranged charges whose algebraic sum equals zero can be dealt with, as a first approximation, as a dipole. This provides an explanation for one of the kinds of molecular interaction, the *van der Waals forces* (see Sec. 31.3).

10.5. THE ELECTRIC FIELD. FIELD STRENGTH

1. As with the concept of the gravitational field (see Sec. 9.5), we shall now introduce the concept of the electric field. We shall assume that electric interaction follows the pattern "charge—field—charge", each charge setting up an electric field that affects all other charges.

The electric field is a form of matter. It exists independently of our consciousness and can be detected by its effect on physical objects, for instance, measuring instruments.

2. The quantitative force characteristic of the electric field is a vector quantity called the *electric field strength*, or *intensity*, \mathbf{E} . Thus

$$\mathbf{E} = \mathbf{F}/q \quad (10.8)$$

The electric field strength is equal in magnitude to the ratio of the force exerted at the given point of the field on a positive test point charge q to the magnitude of the charge. The test charge must be so small that its field does not distort the investigated field.

10.6. ELECTRIC FIELD OF A POINT CHARGE AND OF A DIPOLE

1. The equation for the strength of a field set up by a point charge can be derived from Coulomb's law. Denoting the charge that sets up the field by Q and the test charge by q , we can write

$$F = \frac{Qq}{4\pi\epsilon_0 r^2}$$

from which

$$E = \frac{Q}{4\pi\epsilon_0 r^2} \quad (10.9)$$

Let us place charge Q at the origin of coordinates. Next we draw a radius vector to a point M in the field. Then, if the source Q of the

field is positive, the field-strength vector has the same direction as the radius vector; if the source is negative, it has the opposite direction (Fig. 10.7).

We should note that the strength is determined by the source Q of the field and by the position of point M in the field (by radius vector \mathbf{r}), but not by the test charge q . Hence, we can speak of the field

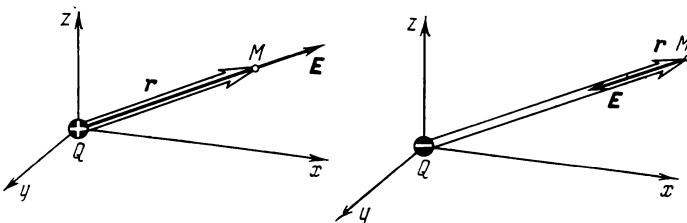


Fig. 10.7

at a certain point in space regardless of whether or not there is a test charge at that point.

2. Generally speaking, we can calculate the field of any charged body because field strength is a vector and we know the rule for adding vectors. The method presents no difficulties in principle. We

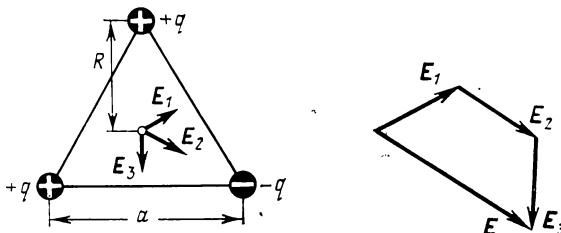


Fig. 10.8

imagine the body as broken up into such small parts that the charge of each part can be regarded as a point charge. Then we calculate the strength of the field set up by each point charge and find the *vector sum*, using the polygon method.

Let us calculate, for example, the field strength at the centre of an equilateral triangle if charges of equal magnitude, two positive and one negative, are located at its vertices (Fig. 10.8). Vectors \mathbf{E}_1 , \mathbf{E}_2 and \mathbf{E}_3 are equal in magnitude. According to equation (10.9),

$$\mathbf{E}_1 = \frac{q}{4\pi\epsilon_0 R^2} = \frac{3q}{4\pi\epsilon_0 a^2}$$

since $R = a/\sqrt{3}$. Adding the three vectors by the polygon method, we obtain an isosceles trapezoid whose larger base is the resultant vector \mathbf{E} . It is evident from Fig. 10.8 that

$$\mathbf{E} = \mathbf{E}_2 + \mathbf{E}_1 \sin 30^\circ + \mathbf{E}_3 \sin 30^\circ = 2\mathbf{E}_1 = \frac{6q}{4\pi\epsilon_0 a^2}$$

It can also be seen that vector \mathbf{E} has the same direction as vector \mathbf{E}_2 .

3. Let us calculate the field strengths E_{\parallel} on the axis of a dipole and E_{\perp} at a point on the normal to the middle of the axis (Fig. 10.9).

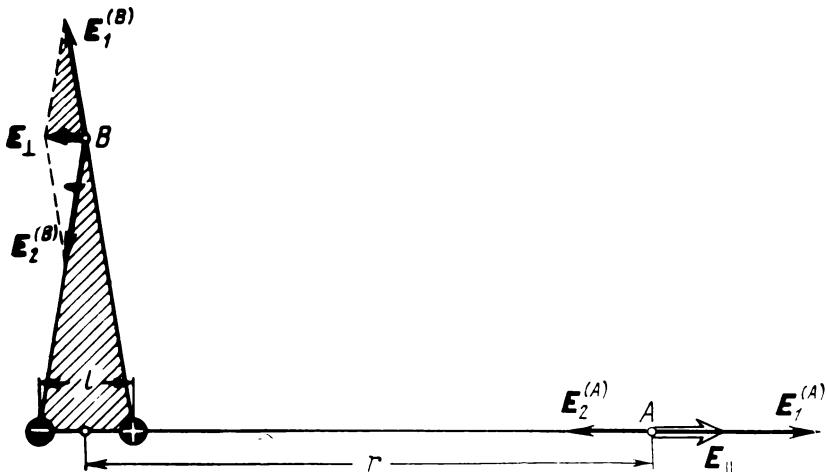


Fig. 10.9

As in Sec. 10.4 the distance from a point of the field to the axis of the dipole is considerably longer than the arm of the dipole: $r \gg l$. The field strength at point A equals the vector sum of the strengths of the fields set up by the positive and negative charges. Since \mathbf{E}_1 and \mathbf{E}_2 are directed along a single straight line,

$$\begin{aligned} E_{\parallel} &= E_1^{(A)} + E_2^{(A)} = \frac{q}{4\pi\epsilon_0 \left(r - \frac{l}{2}\right)^2} + \frac{-q}{4\pi\epsilon_0 \left(r + \frac{l}{2}\right)^2} \\ &= \frac{q}{4\pi\epsilon_0} \cdot \frac{2lr}{\left(r^2 - \frac{l^2}{4}\right)^2} \end{aligned}$$

Neglecting the small value l^2 (because $l^2 \ll r^2$) and substituting p_e for ql , we obtain

$$E_{\parallel} = \frac{2p_e}{4\pi\epsilon_0 r^3} \quad (10.40)$$

The field strength at point B equals the vector sum of the strengths $E_1^{(B)}$ and $E_2^{(B)}$. Since the strength and distance triangles are similar (both are hatched in Fig. 10.9), we can write

$$\frac{E_{\perp}}{E_1^{(B)}} = \frac{l}{R}$$

According to equation (10.9), $E_1^{(B)} = q/4\pi\epsilon_0 R^2$, hence

$$E_{\perp} = \frac{ql}{4\pi\epsilon_0 R^3} = \frac{p_e}{4\pi\epsilon_0 R^3}$$

Since $r \gg l$, we can assume without appreciable error that $R \cong r$. Then

$$E_{\perp} = \frac{p_e}{4\pi\epsilon_0 r^3} \quad (10.11)$$

Thus, the strength of the field set up by a dipole is directly proportional to the moment of the dipole and inversely proportional to the cube of the distance from the point in the field to the centre of the dipole.

Chapter 11

FRICTION

11.1 EXTERNAL AND INTERNAL FRICTION

1. The *force of friction* f is defined as the force that develops at the surfaces of contact of two bodies and impedes their relative motion. It is applied to the bodies along their surfaces of contact in the direction opposed to their *relative velocity* of motion.

Distinction is made between *external* (dry) and *internal* (fluid or viscous) friction.

External friction is the interaction between the surfaces of two solid bodies in contact. When the surfaces are at rest with respect to each other, we speak of *static friction*; when the surfaces are in relative motion, we speak of *sliding*, or *kinetic*, *friction*.

When one of the bodies rolls along the surface of the other without slipping, we have a special kind of resistance which is called *rolling friction*.

Internal friction, or *viscosity*, is the interaction between layers of a liquid or a gas that are moving with respect to one another. In contrast to external friction, here there is no static friction.

11.2. STATIC FRICTION

1. The features of external friction can be studied by means of the device shown in Fig. 11.1. A block of wood, lying on a smooth horizontal board, is attached through a dynamometer to a string that is

wound on a drum mounted on the shaft of an electric motor. By varying the speed of the motor it is possible to vary the pulling force within the required range. Experiments show that as long as the pulling force is less than a certain value F_{\min} , the block cannot be

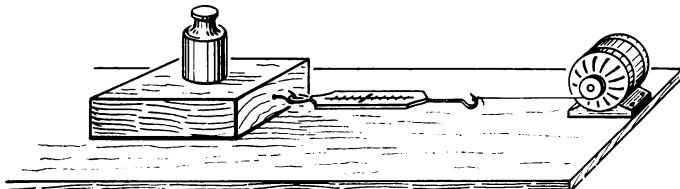


Fig. 11.1

set in motion. When the pulling force exceeds this value, the block suddenly starts moving, either uniformly or with acceleration, according to the conditions in the experiment.

2. If the block is not accelerated when a pulling force is exerted on it, this can only be due to the fact that the *pulling force is counterbalanced by the friction force*.

Here the fundamental law of dynamics has the form

$$F + f = ma \quad (11.1)$$

where F = pulling force

f = force of friction

m = mass of the block

a = its acceleration.

When $a = 0$, the block is either at rest or is moving with uniform velocity. Hence, $|F| = |f|$. From this it follows that we can determine the force of static friction or that of sliding friction simply by measuring the pulling force corresponding to this state of motion.

The force of static friction is not a single-valued quantity. Depending on the applied pulling force, the value of the force of static friction varies from zero to F_{\min} , i.e. the force at which the block suddenly begins to move. Therefore

$$f_s \leq f_s^{\max} = F_{\min} \quad (11.2)$$

The force of static friction is usually called the *maximum force of static friction*.

3. The theory of dry friction has not been sufficiently investigated, but the mechanism of static friction can be roughly expounded as follows.

The surface of a solid body, even if finely ground, is far from smooth. It has on it microscopic projections and depressions, fissures and other irregularities. Often the surface is covered with oxides, adherent

layers of gases and liquids, and foreign inclusions. When the surfaces of two bodies come into contact, the microprojections often engage the corresponding depressions (at a large magnification the contacting surfaces may resemble those in Fig. 11.2). Naturally, the meshing of these projections impedes relative motion of the contacting bodies.

At various spots on the surfaces the distance between the bodies may be of the order of the range of intermolecular forces. This leads to surface adhesion at the contact points, which also impedes relative motion.

The engagement of the microprojections and the partial adhesion of the surfaces are promoted by the normal force P_n , which presses the bodies together. This force may be the force of gravity or its normal component (if the body lies on an inclined plane) or any other force perpendicular to the contact surface of the bodies.

A pulling force less than the maximum force of static friction leads mainly to elastic deformation of the microprojections and the contact

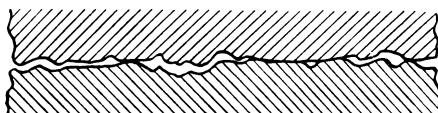


Fig. 11.2

points where the forces of intermolecular cohesion are exerted. The resulting elastic force is, in essence, the force of static friction.

4. The above is only a qualitative explanation of the origin of static friction, and is insufficient to calculate the magnitude of the friction force. We shall derive the corresponding law by resorting to an experiment.

Let us vary the normal force P_n (and, accordingly, the corresponding reaction force $N = P_n$) by loading the wooden block of the device shown in Fig. 11.1 with various weights. Measuring the pulling force each time, we find that the *force of static friction is proportional to the normal force*. Thus

$$f_s^{\max} = \mu_s N \quad (11.3)$$

The quantity μ_s is called the *coefficient of static friction*.

If we change the area of contact of the block, for example, by turning the block on its side, we shall find that the force of static friction does not depend on the area of contact. Now, dividing the two sides of equation (11.3) by the area of the contact surface of the block and denoting the quantity called the *tangential stress* by $\tau = f/A$, we can, since $p = P_n/A$ is the contact pressure, write the derived law of friction as follows:

$$\tau_s^{\max} = \mu_s p \quad (11.4)$$

5. The coefficient of static friction depends upon the quality of surface finish. The friction of ground and polished surfaces is usually less than that of rough surfaces. This is true only within limits, however. Experiments show that for a very high class of finish, when the contacting surfaces are carefully ground and polished, the bodies firmly adhere, leading to a sharp increase in the coefficient of friction. Evidently, the main cause of the development of friction forces is the meshing of the projections when the surfaces are rough, and the molecular cohesion when the surfaces are finely ground.

The force of static friction also depends upon the time the surfaces have been in contact with each other. Prolonged contact and a large normal force cause plastic deformation of the contacting surfaces. This contributes to the cohesion of the bodies and increases the force of static friction.

11.3. THE ANGLE OF FRICTION

The coefficient of friction can be conveniently determined by the device shown in Fig. 11.3. If we carefully raise the inclined plane, we find that at a certain angle φ_0 the block suddenly starts to slide down the plane.

Three forces are acting on the block: the force of gravity P , the reaction force N and the friction force f_s . In the absence of acceleration the resultant of the three forces equals zero. This means that

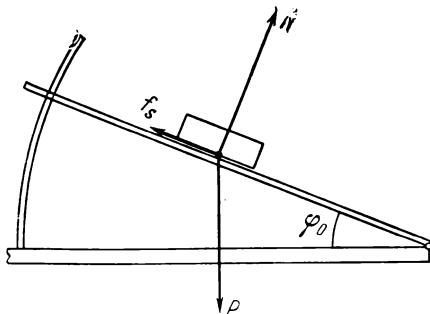


Fig. 11.3



Fig. 11.4

they form a closed triangle (Fig. 11.4). Since the force of gravity is directed vertically, the force of friction along the inclined plane and the reaction force is perpendicular to the plane, the force triangle NPf_s is a right triangle and its vertex angle is φ_0 . It follows that $f_s = N \tan \varphi_0$. But, since $f_s = \mu_s N$

$$\tan \varphi_0 = \mu_s \quad (11.5)$$

Measurements made on various materials and for various grades

of surface finish indicate that the coefficient of static friction varies within very wide limits, from several hundredths to unity. With considerable molecular cohesion the coefficient may reach values as high as 2 or 3.

11.4. SLIDING FRICTION

1. Experiments show that a pulling force must be continuously applied to the block of the device in Fig. 11.1 to maintain uniform motion. But uniform motion is inertial: it continues without the action of external forces. It is obvious, then, that the absence of acceleration of the block when a pulling force is applied can only be explained by the fact that the pulling force counterbalances the force of sliding friction (*kinetic friction*).

If we investigate the force of sliding friction under various conditions, we shall find that it depends, like static friction, on the normal force pressing the bodies together and on the quality of finish of the rubbing surfaces. It does not depend upon the area of contact and only to a small degree on the velocity of relative motion of the bodies. This enables the force of sliding (kinetic) friction to be determined from the formula:

$$f_k = \mu N \quad (11.6)$$

where μ is the coefficient of sliding friction.

2. While static friction is due mainly to the elastic deformation of the microprojections on the contacting bodies, sliding friction occurs as the result of the plastic deformation of the projections and their partial destruction. In fact, when the bodies are at rest the various microprojections are subject to different forces and have different strengths. When the pulling force is applied, some projections are immediately torn off. Then the pulling force is concentrated on the remaining projections which can no longer withstand the increasing load. This leads to an avalanche-type destruction of the microprojections and the body suddenly starts to move. The friction force decreases as motion begins because the reduced meshing and adhesion of the surfaces furthers their relative slipping action, whereby the coefficient of sliding friction μ is less than the coefficient of static friction μ_s .

3. A qualitative analysis of the mechanism of sliding friction reveals that the friction is due to *wear* and *heating* of the rubbing surfaces. The wear depends upon the condition of the rubbing surfaces, i.e. the degrees of their roughness, hardness, etc. Wear plays an important role in certain machining operations, such as lapping and polishing.

The wear and heating of machine parts subject to friction are certainly harmful factors, and many measures are taken in engineer-

ing to reduce friction. Lubrication is most often used for this purpose (see Sec. 11.7). In a great many cases sliding friction can be changed to rolling friction by installing ball or roller bearings (Fig. 11.5).

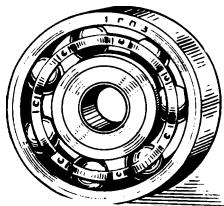


Fig. 11.5

11.5. ROLLING FRICTION

- Let us replace the sliding wooden block in our experimental device with a cylinder, which can rotate about its axis (Fig. 11.6). If the axis of the cylinder is clamped by screws in the yoke to prevent the cylinder from rotating and the weight of the cylinder is approximately the same as that of the block, there will be no appreciable difference in the forces of friction because friction is practically independent of the area of contact. If the screws in the yoke are loosened to release

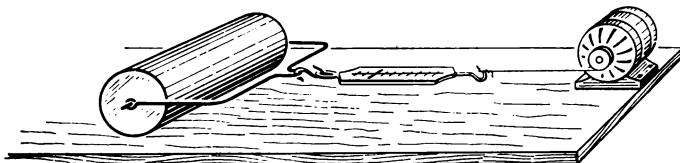


Fig. 11.6

the cylinder, it will begin to roll along the board and the force of friction will be sharply reduced. Hence, the *force of rolling friction is substantially less than that of sliding friction.*

2. It can be shown by experiments that the force of rolling friction is directly proportional to the normal force and inversely proportional to the radius of a rolling cylinder or wheel. Thus

$$f_r = k \frac{N}{r} \quad (11.7)$$

Here k is the coefficient of rolling friction. It has the dimensions of length and in SI units is measured in metres. Equation (11.7) is valid only if the cylinder rolls along the surface without slipping.

11.6. MOTION OF BODIES SUBJECT TO THE FORCE OF FRICTION

- Up till now we have dealt with the force of friction as something that impedes the motion of bodies. This is not always true, however. In fact it is static friction that enables a person or an automobile to move along the earth's surface, a train or streetcar to travel along

tracks, etc. When a person walks, a force of static friction is developed between the soles of his shoes and the ground. This is the force that causes the motion. Everybody knows how difficult it is to walk on ice, where the force of static friction is small.

2. Let us consider the role of the friction force when a bicycle rider makes a turn on a level road. We know from experience that it is necessary to lean with the bicycle toward the proper side ("into the curve"). This will automatically steer the front wheel as required.

We shall find the forces acting on the cyclist when he leans to his left (Fig. 11.7). Here the force of gravity P and the reaction force N do not act along a single straight line. These forces have a tendency to rotate the cyclist clockwise in a vertical plane (as seen in Fig. 11.7) which would lead to skidding of the wheels. But this develops a force of static friction f_s , acting in the direction in which the cyclist leans.

Since the force of friction is perpendicular to the velocity, it imparts a normal acceleration $a_n = v^2/r$ to the bicycle and the rider. It follows from the fundamental law of dynamics that the centripetal force equals the force of static friction. Thus

$$\frac{mv^2}{r} = f_s \leqslant \mu_s mg$$

$$\frac{v^2}{r} \leqslant g\mu_s$$

The angle of inclination of the cyclist (from the vertical) can be found from the condition that the resultant of the reaction and friction forces should lie on the axis of the moving body. Therefore

$$\tan \alpha = \frac{f_s}{N} = \frac{mv^2}{rmg} = \frac{v^2}{rg} \leqslant \mu_s$$

Hence, the cyclist must not lean away from the vertical more than the angle of friction $\varphi_0 = \arctan \mu_s$. Cycle race tracks are banked on curves to allow cyclists to make sharp turns at high speeds. The roadway is inclined by an angle that satisfies the condition: $\tan \alpha \cong \frac{v^2}{rg}$.

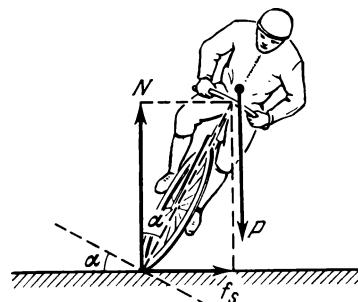


Fig. 11.7

11.7. INTERNAL FRICTION

1. Experiments show that forces of internal friction (viscous forces) are developed in flowing liquids and gases. The existence of these forces can be demonstrated by means of the device in Fig. 11.8. A

heavy metal cylinder hangs from a thin wire in an external cylindrical vessel so that the axis of the vessel coincides with the wire. A liquid is poured into the vessel. Now, if the vessel is rotated about its vertical axis, the inner cylinder will turn through a certain angle. Therefore, the rotating liquid acts on the surface of the inner cylinder with a force that is counterbalanced by the elasticity of the twisted wire.

This force depends upon the distance between the inner cylinder and the inner wall of the vessel, the speed at which the vessel is rotated, the kind of liquid used, and its temperature.

2. The phenomenon of internal friction, or viscosity, can be described as follows. Assume that there is a layer of liquid between two planes, and that the upper plane moves with a velocity v with

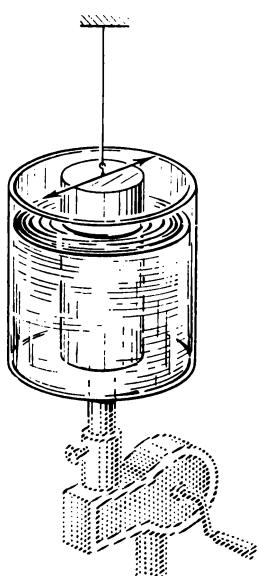


Fig. 11.8

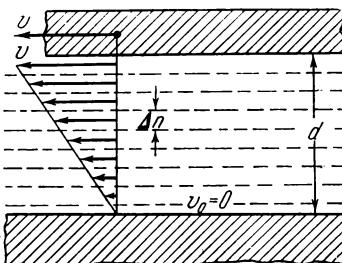


Fig. 11.9

respect to the lower (Fig. 11.9). Let us imagine that the liquid is divided into very thin layers by parallel planes spaced at intervals of Δn . The layers of liquid in contact with the solid bodies adhere to them. The intermediate layers have the velocity distribution shown in Fig. 11.9. This distribution is characterized by the velocity gradient, the quantity $\Delta v / \Delta n$, which shows how rapidly the magnitude of the velocity varies per unit length in the direction perpendicular to the velocity vector (and, correspondingly to the layers of liquid).

When there is a velocity gradient, forces are developed between the contacting layers of liquid. The forces are directed along the planes of contact and tend to impede relative motion of the layers. These are the forces of internal friction, or fluid friction, as it is sometimes called. According to Newton, the *tangential stress* of the forces of internal friction is proportional to the velocity gradient. Thus

$$\tau = \frac{f_{in}}{A} = \eta \frac{\Delta v}{\Delta n} \quad (11.8)$$

where η is the coefficient of internal friction, or *viscosity*, of the liquid.

The SI unit of viscosity is the *pascal-second*:

$$1 \text{ Pa}\cdot\text{s} = 1 \text{ N}\cdot\text{s}/\text{m}^2$$

and its dimensions are $[\eta] = ML^{-1}T^{-1}$. In the cgs system the unit of viscosity is called the *poise*:

$$1 \text{ P} = 1 \text{ dyne}\cdot\text{s}/\text{cm}^2 = 0.1 \text{ Pa}\cdot\text{s}$$

3. The forces of internal friction are manifested in gases as well as in liquids. This can be demonstrated by rapidly rotating the device shown in Fig. 11.8, with air (or some other gas) between the vessel and the inner cylinder. The viscosity of gases is very low, about 10^{-5} that of liquids.

The forces of internal friction are considerably less than those of sliding friction. Therefore, *lubrication* is used to reduce the friction between moving parts of mechanisms and machines. This consists in having a layer of viscous liquid, the lubricant, filling the space between the rubbing surfaces and holding them apart. Lubrication substantially reduces the forces of friction, as well as heating and wear of the parts due to friction. At the same time no liquids should be allowed to get into friction clutches, between the belt and sheaves of belt drives, between the driving wheels of a locomotive and the track, etc. because it is the force of friction in these cases that transmits motion. Any reduction in friction here is obviously harmful.

11.8. MOTION OF BODIES IN FLUIDS

1. Experiments show that bodies moving in liquids or gases (both are fluids) are subject to resistance. This resistance depends upon the size and shape of the body, its velocity and the properties of the fluid.

Consider a stream of liquid, moving with a velocity v in a certain reference frame that is fixed to a stationary body. From the principle of relativity it follows that the force with which the flowing stream acts on the body is exactly equal to the resistance experienced by the body if it were moving with the same velocity in stationary liquid.

2. The oncoming stream of liquid is deformed by the body and flows around it. The layer of liquid directly adjacent to the body adheres to it. Under certain conditions a *boundary layer* is formed. This is a region of the liquid in which its velocity grows very rapidly from zero almost to the velocity of undisturbed flow, i.e. it has a large velocity gradient. If the boundary breaks away (or, as we say, is separated) from the body, eddy currents are formed behind the body (Fig. 11.10). The resistance of a fluid depends to a great extent on the phenomena that occur in the boundary layer and on the kind of eddies

that are set up. But a detailed analysis of the phenomena occurring in the boundary layer and a calculation of the resistance forces are exceptionally complex tasks. We can attempt to assess only the order of magnitude of these forces and the main parameters on which they depend.

We can assume that the resistance force consists of two components: pressure resistance, or drag, and friction drag. The first is determined

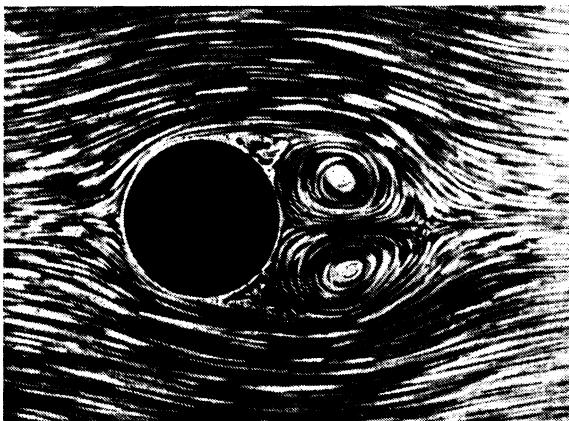


Fig. 11.10

by the *difference in pressure* at the front and rear edges of a body in a stream of fluid. The second by the *forces of internal friction* due to the large velocity gradient in the boundary layer.

We shall use the dimensional method to assess the influence of each of these components.

3. The resistance force arising from the difference in pressure at the front and rear edges of a body in a fluid stream depends upon the density of the fluid, the velocity of flow and the midsection area (the maximum area of cross-section perpendicular to the flow). Thus

$$R = B \rho^\alpha v^\beta A^\gamma \quad (11.9)$$

where B is a dimensionless factor.

In any equation the right and left sides must have the same dimensions, i.e. $[R] = [\rho^\alpha v^\beta A^\gamma]$. Substituting MLT^{-2} for $[R]$, ML^{-3} for $[\rho]$, LT^{-1} for $[v]$ and L^2 for $[A]$, we determine the unknown exponents α , β and γ . We can write

$$MLT^{-2} = M^\alpha L^{-3\alpha} L^\beta T^{-\beta} L^{2\gamma}$$

from which we find that

$$\alpha = 1, \quad -3\alpha + \beta + 2\gamma = 1 \quad \text{and} \quad -\beta = -2$$

or

$$\alpha = 1, \quad \beta = 2 \quad \text{and} \quad \gamma = 1$$

According to accepted practice we denote factor B by $C/2$. Then the equation for the pressure drag becomes

$$R = CA \frac{\rho v^2}{2} \quad (11.10)$$

Factor C in this equation depends largely on the shape of the body or, as they say, its streamlining. Thus, while $C = 1.1$ to 1.2 for a

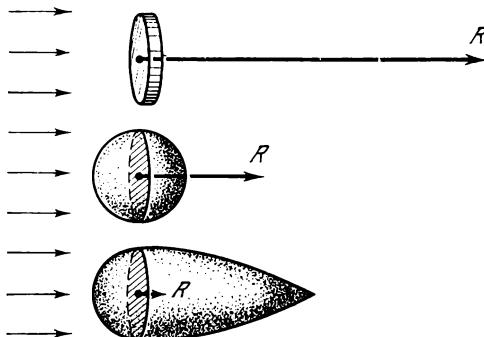


Fig. 11.11

round disk and $C = 0.4$ to 0.2 for a sphere, for a drop-shaped body $C \approx 0.04$, which is about one-tenth of the value for a sphere and one-thirtieth of that for a disk (Fig. 11.11).

4. The force of internal friction, according to equation (11.8), depends upon the viscosity, velocity and the size of the body. Thus

$$f_{in} = D\eta^m v^n S^k \quad (11.11)$$

where D is a dimensionless factor and S is a characteristic dimension of the body. Dimensional analysis of the left and right sides of this equation yields

$$MLT^{-2} = M^m L^{-m} T^{-n} L^n T^{-n} L^k$$

from which

$$m = 1, \quad -m + n + k = 1 \quad \text{and} \quad -m - n = -2$$

hence

$$m = n = k = 1$$

Then the friction drag equation is

$$f_{in} = D\eta v S \quad (11.12)$$

Like C , factor D is usually determined experimentally. Sir George G. Stokes (1819-1903) found for a sphere that $D = 6\pi$ and that the characteristic dimension S is the radius r . Thus, for a sphere

$$f_{in} = 6\pi\eta vr \quad (11.13)$$

5. The total force of resistance is a quite complicated combination of the pressure drag R and the friction drag f . However, at very low velocities, the friction drag, proportional to the first power of the velocity, will be considerably greater than the pressure drag which is proportional to the square of the velocity. The opposite will be true, evidently, at high velocities.

Since the ideas of "high" and "low" velocities are rather indefinite, we shall introduce a special criterion for their evaluation. For this purpose, we first find the ratio of the pressure drag to the friction drag:

$$\frac{R}{f_{in}} = \frac{C}{2D} \frac{\rho v^2 A}{\eta v S}$$

Neglecting the dimensionless factors C and D , and taking the cross-section area to be proportional to the square of the characteristic dimension ($A \propto S^2$), we obtain

$$\frac{R}{f_{in}} \cong \frac{\rho v S}{\eta} = Re \quad (11.14)$$

The dimensionless number Re , called the *Reynolds number*, plays an exceptionally important role in hydro- and aerodynamics. It is the quantity that determines resistance.

Hence, for "creep" when $Re < 1$, the pressure drag can be neglected and only the friction drag need be taken into account. On the contrary, only the pressure drag is taken into consideration for high values of the Reynolds number, the friction drag being neglected.

11.9. BODIES FALLING IN A FLUID

1. The free fall of bodies, i.e. the motion of bodies subject only to the force of gravity, was discussed earlier. We established that this is a uniformly accelerated motion and found the corresponding motion equation (see Sec. 8.2). Here we shall deal with the fall of bodies in a medium that exerts resistance. Such media may be either liquids or gases, i.e. fluids. At the same time we shall determine the conditions under which we can neglect air resistance and deal with the fall of a body in air as free fall.

2. Three forces act on a body falling in a fluid: the force of gravity P , the buoyant force F_b , and the force of resistance F_{res} (Fig. 11.12). The fundamental law of dynamics for this case can be written as

$$ma = P - F_b - F_{res} \quad (11.15)$$

Only one of these three forces is variable, the force of resistance, whose value increases rapidly with velocity. Gravity and the buoyant force are constants. Hence, as the body continues to fall, its velocity increases until the retarding force of resistance becomes as great as the difference between the force of gravity and the buoyant force. At this instant, the acceleration has decreased to zero. From then on the body falls with *uniform velocity*. Thus, when a body falls in a fluid, only the initial stage is accelerated motion. After a certain instant the body continues to fall with uniform velocity, which is called the *terminal velocity*.

3. The terminal velocity v_t can be determined from equation (11.15) under the condition that $a = 0$:

$$P - F_b - F_{res}^t = 0 \quad (11.16)$$

To carry out such a calculation it is necessary to assess the order of magnitude of the Reynolds number and to choose the proper expression for the force of resistance.

Consider, for example, a steel ball ($\rho = 7.8 \times 10^3 \text{ kg/m}^3$) with the radius $r = 2 \text{ mm}$ falling in glycerine. The viscosity of glycerine is $\eta = 0.83 \text{ Pa}\cdot\text{s}$ and its density is $\rho_l = 1.2 \times 10^3 \text{ kg/m}^3$. The Reynolds number is

$$\text{Re} = \frac{\rho_l r v}{\eta} = \frac{1.2 \times 10^3 \times 2 \times 10^{-3} v}{0.83} \cong 3v$$

Let $\text{Re} < 1$ then the force of resistance is calculated by Stokes' formula [equation (11.13)].

As we know, the force of gravity is $P = mg = \frac{4}{3} \pi r^3 \rho g$. The buoyant force, according to *Archimedes' principle*, equals the weight of the fluid displaced by the body: $F_b = \frac{4}{3} \pi r^3 \rho_l g$. After substituting into equation (11.16) and transforming, we obtain

$$v_t = \frac{2r^2 g (\rho - \rho_l)}{9\eta} = \frac{2 \times 4 \times 10^{-6} \times 9.8 \times 6.6 \times 10^3}{9 \times 0.83} \cong 0.07 \text{ m/s}$$

We can now see that we were justified in using the Stokes formula for calculating the force of resistance because we find that $\text{Re} = 0.21 < 1$.

In gases, for instance in air, such creep is observed only for very small bodies, such as dust specks, droplets of fog, etc.

4. For large Reynolds numbers the force of resistance is due mainly to the pressure drag R . Hence, the fundamental law of dynamics

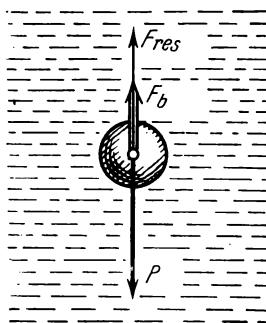


Fig. 11.12

(neglecting the buoyant force for gases since $F_b \ll P$) becomes

$$ma = mg - CA \frac{\rho_0 v^2}{2} \quad (11.17)$$

where ρ_0 is the density of the gas.

If a heavy body of small size falls from a small height, the force of resistance may prove to be much less than the force of gravity. Then the resistance can be neglected and free fall can be assumed. If the body falls from a great height, however, the resistance cannot be neglected and from a certain instant the body falls with a constant terminal velocity. This can be readily determined from equation (11.17), setting $a = 0$. Thus

$$v_t = \sqrt{\frac{2mg}{CA\rho_0}} \quad (11.18)$$

Hence, droplets of rain of spherical shape ($C = 0.4$) with a radius of about $r \approx 1$ or 2 mm fall through air of the density $\rho_0 = 1 \text{ kg/m}^3$ with a terminal velocity that is independent of the altitude of the clouds above the earth's surface (the density of water is $\rho = 10^3 \text{ kg/m}^3$)

$$v_t = \sqrt{\frac{2 \times 4\pi r^3 \rho g}{3C\pi r^2 \rho_0}} = \sqrt{\frac{8r\rho g}{3C\rho_0}} = \sqrt{\frac{8 \times 10^{-3} \times 10^3 \times 9.8}{3 \times 0.4 \times 1}} \approx 8 \text{ m/s}$$

5. An analysis of equation (11.18) explains the principle of the parachute. An airman falling from a considerable altitude without opening his parachute reaches a terminal velocity of several dozens of metres per second. It is fatal, of course, to crash into the earth at such a velocity.

The force of resistance is increased tremendously when he opens his parachute. In fact, the maximum cross-section against which the air flows is about a hundred times more for a chute than for a falling man, and factor C is several times greater for the chute. When the factor CA is increased by several hundred times, the terminal velocity is greatly reduced, approximately to 3 or 4 m/s. It is quite safe to land at this velocity.

Chapter 12

THE THEORY OF RELATIVITY

12.1. VELOCITY OF LIGHT AND THE LAW OF ADDITION OF VELOCITIES

1. By the middle of the 19th century methods had been devised for a quite accurate measurement of the velocity of light. It was found to be $c = 3 \times 10^8 \text{ m/s}$ in a vacuum or, more accurately,

$$c = (2.9979250 \pm 0.0000010) \times 10^8 \text{ m/s}$$

A question that naturally arises is: what reference frame is implied for the given velocity of light? It is, in general, meaningless to speak of a velocity without specifying the reference frame.

Proceeding from the classical law of the addition of velocities, we come to the conclusion that the velocity of light should be different in different reference frames. Consequently, the given value

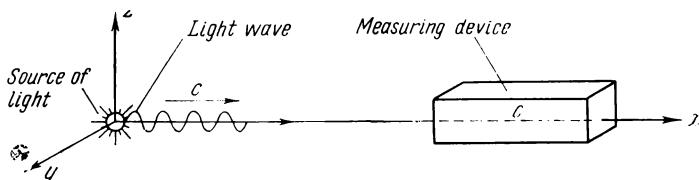


Fig. 12.1

of the velocity of light holds only for some single reference frame, for instance, the one containing the source of the light. This reasoning can be tested by the following thought experiment.

Assume that an instrument for measuring the velocity of light with sufficiently high accuracy is at rest with respect to the source-

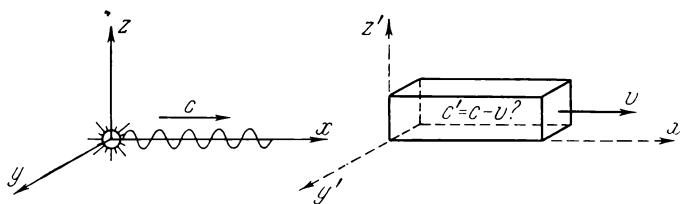


Fig. 12.2

of light which is in reference frame xyz. The instrument will register a velocity of light equal to c (Fig. 12.1).

Next we put the instrument into another reference frame which moves with respect to frame xyz as shown in Figs. 12.2, 12.3 and 12.4. On the basis of the classical law of the addition of velocities, we would expect that $c' = c - v$, $c'' = c + v$ and $c''' = \sqrt{c^2 + v^2}$, i.e. that the velocities of light c' , c'' and c''' are different. But now we shall see whether this is so.

2. The first experiment for measuring the velocity of light in a moving reference frame was conducted by Albert A. Michelson (1852-1931) in 1881. Similar experiments have been conducted by many scientists down through the years almost to the present time and, naturally, the accuracy of measurement steadily increased. But all these experiments yielded a negative result: to a very high:

degree of accuracy, in all inertial reference frames (regardless of the magnitude and direction of their velocity) the velocity of light in a vacuum is exactly the same as in the reference frame linked to the light source, i.e. $c' = c'' = c''' = c$.

This result indicates that the classical law of the addition of velocities has only a limited field of application. It is invalid, in particular, for describing phenomena associated with the propagation of

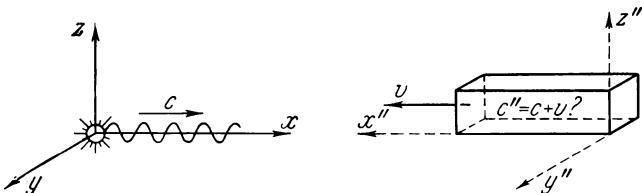


Fig. 12.3

light. But the classical law of the addition of velocities is based on the Galilean transformations. Therefore, the latter also have only a limited field of application.

3. Since the invariance of the velocity of light in all inertial reference frames is a firmly established experimental fact, and this fact disagrees with the Galilean transformations, the need arose for

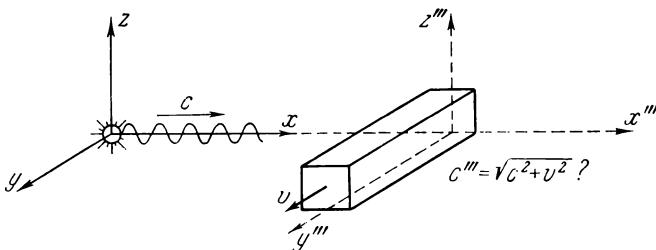


Fig. 12.4

a critical revision of the ideas on which the transformations are based. This was done by Einstein in 1905, and it required a complete reconsideration of our concepts of space and time.

Einstein showed that in deriving the Galilean transformations, two propositions were introduced in the implicit form, which seemed so obvious that no one thought it necessary to give grounds for them. They were the following:

(a) It was assumed that the simultaneity of two events is an absolute concept: two events simultaneous in one reference frame are simultaneous in any other frame. From this it followed that clocks

located in arbitrary inertial reference frames indicate the same time. In short, that the course of time is exactly the same in all inertial reference frames.

(b) It was assumed that the length of a rod is the same in all inertial frames of reference (see Sec. 2.4).

In fact, it was found that these propositions are not universal, but only a first approximation, valid only for Newtonian mechanics.

12.2. BASIC POSTULATES OF THE SPECIAL THEORY OF RELATIVITY

The special theory of relativity is the foundation of relativistic mechanics. It is based on two principles which are postulates. These postulates have been reliably confirmed by experimental tests.

1. *Principle of relativity.* All inertial reference frames are equally valid; not only mechanical, but all other phenomena of nature proceed in the same way in all inertial reference frames.

2. *Principle of the invariance of the velocity of light.* The velocity of light in vacuum is the same and equal to c in all inertial frames of reference.

Now we shall discuss the consequences following from these basic propositions.

12.3. SIMULTANEITY OF EVENTS

1. To determine the instant of time when a certain event occurs at a given point A in space, it is necessary to read off the time on a clock located near this point. The same must be done for any other event at point B . There is no sense in comparing the time indicated on the two clocks unless they have been synchronized beforehand. This proposition seems so obvious that for many years no one asked themselves the question: what, exactly, is the physical procedure for synchronizing clocks?

Einstein proposed that clocks be synchronized by means of light signals on the basis of the following thought experiment. By definition we assume that the time during which a light signal travels from A to B is equal to the time required for it to travel from B to A . Let an outgoing signal leave point A at the instant of time t_A^{out} , reach point B at instant t_B where it is reflected and returns to point A at instant t_A^{in} . Then, by definition,

$$t_B - t_A^{out} = t_A^{in} - t_B$$

In other words, the clock at point B should indicate the time

$$t_B = \frac{1}{2} (t_A^{out} + t_A^{in})$$

Two conditions are complied with when clocks are synchronized in this way:

(a) condition of symmetry: if clock A runs synchronously with clock B then clock B runs synchronously with clock A ;

(b) condition of transitivity: if clock A runs synchronously with clock B , and clock B with clock C , then clock A runs synchronously with clock C .

2. It follows from the two basic postulates of the theory of relativity that *two events simultaneous in one reference frame are not simultaneous in another reference frame*.

To prove this statement, which appears paradoxical at first glance, we shall consider the following experiment. Let one reference frame be fixed on the surface of the earth and another on a railway car

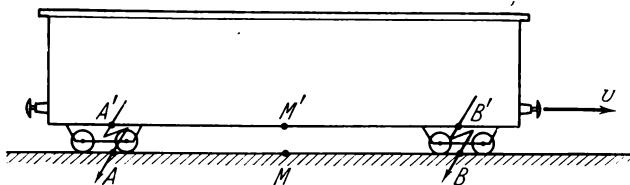


Fig. 12.5

travelling with respect to the earth at uniform velocity in a straight line (Fig. 12.5). We mark off the points A , B and M on the earth so that $AM = BM$. In the car we mark off the corresponding points A' , B' and M' , so that $A'M' = B'M'$. Assume, now, that at the instant the points marked off on the earth and in the car coincide, certain events occur at points A and B , for instance, two bolts of lightning strike these points.

Evidently, the light from the two flashes will reach point M simultaneously, but that from B' will reach M' before the flash from A' . Let us analyze these events from the points of view of observers in the two reference frames.

On the earth: the distance $AM = BM$. The velocity of light is the same in all directions. Signals of the two events arrived simultaneously at point M . Therefore, the *events at points A and B occurred simultaneously*.

In the car: the distance $A'M' = B'M'$. The velocity of light is the same in all directions. The signal from point B' arrived before that from point A' . Therefore, the *event at point B' occurred earlier than that at point A'*.

If the car had been travelling from right to left, the opposite result would be obtained: the event at B' occurred later than the one at point A' .

Hence, the concept of simultaneity is relative, and the course of time differs, or as they sometimes say, clocks run differently, in different inertial reference frames.

3. Note that in the above discussion we used only two principles: we considered both reference frames to be equally valid (principle of relativity), and we took the velocity of light to be the same in all

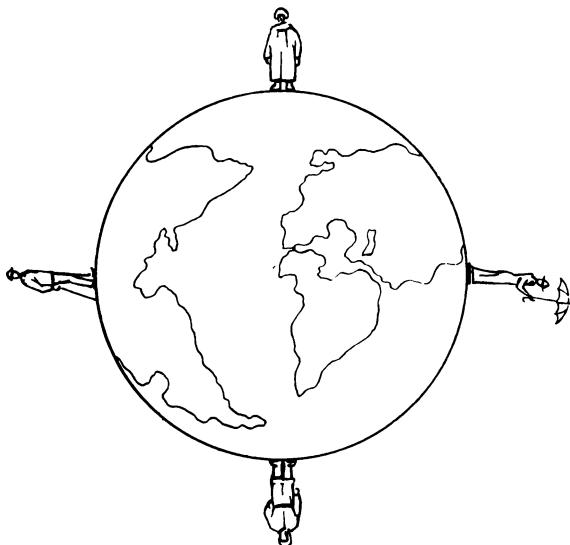


Fig. 12.6

directions (principle of the invariance of the velocity of light). We did not introduce any additional hypotheses.

In classical physics, it was assumed without any proof that the course of time is identical in all reference frames. This led to the classical law of the addition of velocities which contradicts the experimentally established principle of the invariance of the velocity of light.

4. Students often ask, are the events at points *A* and *B* actually simultaneous or not? This question is evidently without meaning. To answer the question of what *actually* happens means merely to indicate some *preferred* reference frame. There simply is no such reference frame in nature.

In general, it is not always possible to answer the question of what happens in reality. When it is morning in Moscow, it is evening in Vladivostok and night-time in San Francisco. It is absurd to ask what the time of day is in reality at this moment: morning, evening or night? These are relative concepts and the answer depends on what place on the earth the observer is located. The concepts of "up" and

"down" are just as relative for people living on opposite sides of the globe (Fig. 12.6).

In exactly the same way, the concept of simultaneity is relative. It is impossible to answer the question of whether two events took place simultaneously or not without specifying the reference frame with respect to which the given problem is to be solved.

5. Why, then, do we neglect this property of time in classical physics, and consider time to be invariant, i.e. a quantity independent of the reference frame?

The fact is that in classical physics we deal with velocities that are so much less than the velocity of light that we can neglect the time lag in the signals and assume that light propagates practically instantaneously, with an infinitely high velocity. Then the observers on the ground and in the railway car would detect no difference in the arrival of the signals from points A and B (Fig. 12.5). Hence, they would reach the same conclusion on the simultaneity or non-simultaneity of events.

Therefore, the relativity of simultaneity is a consequence of the finiteness of the velocity with which light propagates and, in general, of the finiteness of the velocity with which any interaction can be transmitted. Einstein based his theory of relativity on the experimental fact that no processes whatsoever are propagated at an infinitely high velocity.

12.4. SIMULTANEITY AND LENGTH

We defined the length of a rod previously as the difference in the coordinates of its beginning and end, *measured simultaneously* (see Sec. 2.4). However, the idea of simultaneity is relative and events that are simultaneous in one reference frame turn out to be nonsimultaneous in another frame. It follows that the *length of a rod will be different in different reference frames* (see Sec. 12.8). We cannot speak of the length of a rod without indicating the reference frame with respect to which this length is measured. This is just as meaningless as to speak of the velocity of a passenger without specifying the reference frame; his velocity with respect to the railway car is not equal to his velocity with respect to the earth.

Thus, according to the theory of relativity, the intervals of time between events and lengths or distances are relative concepts having different values in different inertial reference frames.

12.5. RELATIVISTIC LAW FOR THE ADDITION OF VELOCITIES

1. As demonstrated above, the Galilean transformations do not agree with the experimentally established fact of the constant velocity of light. Hence, it is necessary to derive new transformations for

coordinates in going over from one inertial frame of reference to another.

Let us consider two sets of coordinates, or frames of reference: xyz fixed to the surface of the earth, and $x'y'z'$ to a railway car (see Fig. 2.1). Here again there is no motion along the y - and z -axes and, therefore, $y' = y$ and $z' = z$ as in the Galilean transformations.

Free space is homogeneous and isotropic. This means that the transformations of coordinates are linear functions (Fig. 12.7). In

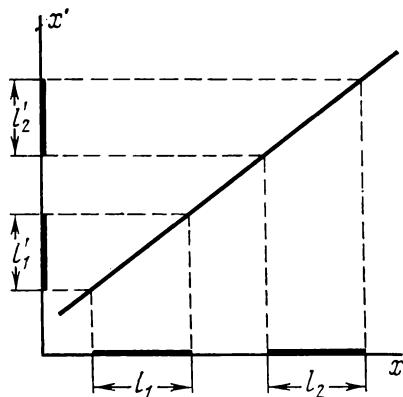


Fig. 12.7

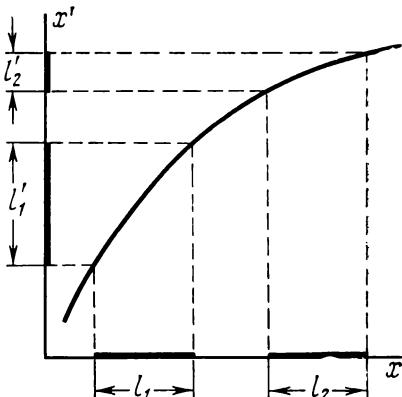


Fig. 12.8

fact, as is evident from the figure, the length of a line or a distance does not depend upon what region of space it is located in: if $l_1 = l_2$, then $l'_1 = l'_2$.

If the transformation had been nonlinear (Fig. 12.8), then it would follow from $l_1 = l_2$ that $l'_1 \neq l'_2$, i.e. the length of a rod would depend upon the region of space it is located in. This would disagree with the homogeneity of space.

Naturally, a similar argument can be made for time.

Therefore, we will seek for relativistic transformations in the form of the linear functions:

$$x' = Ax + Bt \quad \text{and} \quad t' = Mx + Nt \quad (12.1)$$

where A , B , M and N are constants whose values are to be determined.

2. Motion along the axis of abscissas in the car can be written as

$$\Delta x' = x'_2 - x'_1 = A(x_2 - x_1) + B(t_2 - t_1) = A\Delta x + B\Delta t \quad (12.2)$$

Similarly, for a time interval between two events we write

$$\Delta t' = M\Delta x + N\Delta t \quad (12.3)$$

Dividing equation (12.2) by equation (12.3) and taking into account that the velocity of a particle with respect to the car is

$$u' = \Delta x'/\Delta t \quad (12.4)$$

and that the velocity of the same particle with respect to the earth is

$$u = \Delta x/\Delta t \quad (12.5)$$

we obtain the expression for the law of addition of velocities (along the axis of abscissas):

$$u' = \frac{Au + B}{Mu + N} \quad (12.6)$$

3. Next we use equation (12.6) to determine the constants introduced into equation (12.1). We do this by considering the following special cases:

(a) Assume that the particle is at rest with respect to the car. Then $u' = 0$ and $u = v$. Substituting into equation (12.6) we obtain $0 = \frac{Av + B}{Mv + N}$, from which

$$B = -Av \quad (12.7)$$

(b) Assume that the particle is at rest with respect to the earth. Then $u' = -v$ and $u = 0$. Substituting into equation (12.6) and making use of equation (12.7), we obtain $-v = -Av/N$, from which

$$N = A \quad (12.8)$$

(c) Assume that a light wave propagates in the car instead of the moving particle. On the basis of the second postulate of the theory of relativity—the principle of invariance of the velocity of light in all inertial reference frames—we obtain

$$u' = u = c \quad (12.9)$$

Substituting into equation (12.6) and making use of the results of equations (12.7) and (12.8), we obtain $c = \frac{Ac - Av}{Mc + A}$ and then

$$M = -Av/c^2 \quad (12.10)$$

4. The values obtained for the constants B , M and N are substituted into equation (12.6). The new expression is the *relativistic law of addition of velocities* (along the axis of abscissas):

$$u' = \frac{u - v}{1 - uv/c^2} \quad \text{or} \quad u = \frac{u' + v}{1 + u'v/c^2} \quad (12.11)$$

12.6. LIMITING NATURE OF THE VELOCITY OF LIGHT

1. The limiting nature of the velocity of light follows as a consequence from the relativistic law of addition of velocities: *no body, in any reference frame, can move with a velocity exceeding the velocity of light in a vacuum.*

Let us suppose that a body moves along a railway car (the primed reference frame) with a velocity u' , not exceeding the velocity of light in a vacuum: $0 < u' < c$. The car is moving with respect to the earth (the unprimed reference frame) with a velocity v also less than the velocity of light: $0 < v < c$. Then we can readily show that the velocity of the body with respect to the earth (for *any* velocities u' and v , less than c) will also be less than the velocity of light.

As proof we shall find the difference between the velocity of light in a vacuum and the velocity of the body in the unprimed frame of reference:

$$c - u = c - \frac{u' + v}{1 + u'v/c^2} = \frac{(c - v)(c - u')}{c(1 + u'v/c^2)} \quad (12.12)$$

Since we assumed from the very beginning that the velocity of the body in the primed reference frame and the velocity of this frame itself are less than c , $c - u > 0$. Hence

$$u < c \quad (12.13)$$

2. This result is unconditionally true for particles of matter. In fact, there always exists a reference frame in which the particle is moving at a velocity less than that of light in a vacuum (in which, for example, it is at rest, i.e. $u' = 0$). But then it will be moving in any other reference frame at a velocity slower than that of light in a vacuum.

Note that in this derivation we did not resort to any assumptions as to the structure and properties of the particles of matter. Consequently, the result is entirely general, and valid for any particles of matter and any bodies.

Thus, the velocity of light in a vacuum is the highest possible velocity in nature.

12.7. LORENTZ TRANSFORMATIONS

1. Coordinate transformations that agree with the two basic postulates of the theory of relativity were first derived by Hendrik A. Lorentz (1853-1928).

To derive the required transformations we shall substitute the values of the constants B , M and N into equation (12.1) and obtain

$$\left. \begin{aligned} x' &= A(x - vt) \\ t' &= A(t - vx/c^2) \end{aligned} \right\} \quad (12.14)$$

According to the principle of relativity, the two reference frames are equally valid: we can assume that the reference frame fixed on the car is at rest; then the frame fixed on the surface of the earth will be moving with respect to the car with the velocity $v' = -v$ (reciprocity of the Lorentz transformations).

The transformation of coordinates will then be of the form:

$$\left. \begin{aligned} x &= A(x' + vt') \\ t &= A(t' + vx'/c^2) \end{aligned} \right\} \quad (12.15)$$

Substituting equations (12.15) into equations (12.14), we obtain

$$x' = A^2(x' + vt' - vt' - v^2x'/c^2)$$

and, after cancelling like terms,

$$A = \frac{1}{\sqrt{1-v^2/c^2}} \quad (12.16)$$

In their final form, the Lorentz transformations are written as

$$\left. \begin{aligned} x' &= \frac{x-vt}{\sqrt{1-v^2/c^2}}, & y' = y & \text{ and } z' = z \\ t' &= \frac{t-vx/c^2}{\sqrt{1-v^2/c^2}} \end{aligned} \right\} \quad (12.17)$$

or

$$\left. \begin{aligned} x &= \frac{x'+vt}{\sqrt{1-v^2/c^2}}, & y = y' & \text{ and } z = z' \\ t &= \frac{t'+vx'/c^2}{\sqrt{1-v^2/c^2}} \end{aligned} \right\} \quad (12.18)$$

2. Space and time are dealt with in classical mechanics as concepts independent of each other. Lorentz transformations demonstrate the intimate relationship between space and time coordinates. Not only are the space coordinates dependent on time (which was true of the Galilean transformations as well), but time depends upon the space coordinates, and also on the velocity of the reference frame.

3. The Lorentz transformations and the relativistic law of addition of velocities agree with the principle of inertia.

As a matter of fact, if a body is moving at uniform velocity in a straight line with respect to one inertial reference frame (for instance, the railway car), it will be moving at uniform velocity in a straight line with respect to any other inertial frame. From $u' = \text{const}$ and $v = \text{const}$ it follows, according to equation (12.11), that $u = \text{const}$.

12.8. LENGTH OR DISTANCE

1. Assume that a rod is arranged parallel to the axis of abscissas and is at rest with respect to a reference frame xyz (for instance, the earth's surface). We shall denote the rod's length by $l_0 = x_2 - x_1$ (see Fig. 2.2) in the reference frame with respect to which it is at

rest. In a reference frame $x'y'z'$ moving with respect to the frame xyz at a velocity v (for instance, in a railway car), the length of the rod is $l = x'_2 - x'_1$, provided that $t'_2 = t'_1$. According to the Lorentz transformations (12.18):

$$x_2 = \frac{x'_2 + vt'_2}{\sqrt{1-v^2/c^2}}$$

$$x_1 = \frac{x'_1 + vt'_1}{\sqrt{1-v^2/c^2}}$$

from which

$$x_2 - x_1 = \frac{x'_2 - x'_1 + v(t'_2 - t'_1)}{\sqrt{1-v^2/c^2}} \quad (12.19)$$

Putting $t'_2 = t'_1$ according to the specified condition, we find

$$l = l_0 \sqrt{1-v^2/c^2} \quad (12.20)$$

2. We shall call the length of a rod in the reference frame with respect to which it is at rest its *proper length* l_0 . According to equation (12.20) we find that the length of the rod in any other system with respect to which the rod is moving is shorter than the proper length.

Note that in all reference frames moving at different velocities with respect to the reference frame in which the rod is at rest, the length of the rod varies. From this it is obvious that the relativistic change in length is not due to deformations of the rod (or other similar processes).

12.9. TIME INTERVAL BETWEEN TWO EVENTS

1. We shall call any phenomena that occur with bodies *events*.

Suppose that at point A , where the body we are considering is located, an event occurs at the instant of time t_1 and another event at instant t_2 . Then the time interval between the two events is $\tau_0 = t_2 - t_1$. This time interval is measured in the reference frame xyz linked to the body being investigated (*proper time*). Let us find the time interval between the same two events in a reference frame $x'y'z'$ moving with velocity v (along the axis of abscissas) with respect to frame xyz .

Making use of equation (12.17) and the condition that the events occur at the same point A , i.e. $x_2 = x_1$, we obtain

$$\tau = t'_2 - t'_1 = \frac{t_2 - t_1}{\sqrt{1-v^2/c^2}} = \frac{\tau_0}{\sqrt{1-v^2/c^2}}$$

2. Hence, the required time interval is

$$\tau = \frac{\tau_0}{\sqrt{1-v^2/c^2}} \quad (12.21)$$

This result is fundamental. We found that the *time interval between two events has its minimum value in the reference frame linked to point A, where the events occurred.*

In any other reference frame this time interval is longer.

Time τ_0 is called *proper time*.

In other words, in comparison to the reference frame in which the events occur and which we have taken to be at rest, *time is slowed down in a moving reference frame*.

3. The slowing down of time (time dilatation) in a moving reference frame enables us to explain one interesting phenomenon.

The action of cosmic rays at the boundary of the atmosphere produces particles known as muons (see Sec. 83.3). These particles are then detected at the surface of the earth. The atmosphere is about 300 km thick. Even if the muons are moving with velocities near to that of light, the time required for them to pass through the atmosphere is about 10^{-3} s. But experiments show that muons are unstable and decay spontaneously. It was found that the mean life of a muon is only 10^{-6} s, i.e. one thousandth of the time required for it to pass through the atmosphere to the earth's surface! How can we reconcile these apparently conflicting facts?

There was found to be no contradiction. In the frame of reference linked to the muon its life is 10^{-6} s, as we expected. But the reference frame linked to the earth moves with respect to the muon with a velocity near to that of light. Time, therefore, is slowed down in this system, and the muon lives about 10^{-3} s, which is about a thousand times longer than in the frame linked to the muon.

12.10. THE TIME INTERVAL BETWEEN CAUSE AND EFFECT

1. Many events found in nature are related together by what we call a cause-and-effect, or causality, relationship. Before a bullet can hit a target, for instance, it is necessary to fire the gun. Here the shot is the event which is the cause and the hitting of the target by the bullet is the effect.

We should note two typical properties of this kind of relationship between events:

(a) first the event which is the cause occurs, and then the one which is the effect;

(b) if in some way we eliminate the event which is the cause, the other event, the effect, will not occur either.

We shall show that even though the course of time is different in different reference frames according to the theory of relativity, the sequence in time between cause and effect remains.

2. Suppose that in a reference frame linked to the earth a pistol is fired at the instant of time t_1 at a point with the coordinate x_1 , and that the bullet hits the target with the coordinate x_2 at instant t_2 .

(Fig. 12.9). The velocity of the bullet in this reference frame is

$$u = \frac{x_2 - x_1}{t_2 - t_1}$$

Let us now find the time interval between the same two events in a reference frame moving along the axis of abscissas with a velocity v .

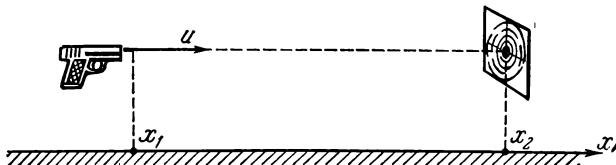


Fig. 12.9

From equation (12.17) we obtain

$$t'_2 - t'_1 = \frac{t_2 - t_1 - \frac{v}{c^2} (x_2 - x_1)}{\sqrt{1 - v^2/c^2}} = \frac{t_2 - t_1}{\sqrt{1 - v^2/c^2}} \left(1 - \frac{uv}{c^2} \right) \quad (12.22)$$

The velocity of the bullet and the velocity v of the new reference frame are less than that of light in a vacuum. Therefore, $1 - uv/c^2$ is a positive number. The sign of the time interval $t'_2 - t'_1$ coincides with the sign of $t_2 - t_1$.

But $t_2 > t_1$, because in the reference frame linked to the earth the bullet hits the target *after the shot was fired*. Therefore, in any other reference frame, $t'_2 > t'_1$.

Thus, though the time interval between two events may vary in different reference frames, the *event which is the cause occurs in all reference frames before the event which is the effect*.

12.11. THE RELATION BETWEEN RELATIVISTIC AND NEWTONIAN MECHANICS

1. Newtonian mechanics and, in particular, the Galilean transformations are based on the assumption that the course of time is identical in all reference frames. We have demonstrated that this assumption is essentially incorrect. Naturally, the question arises: how could this theory be successfully applied in practice for several centuries and give *correct* answers? And what is more, we continue to calculate the motion of celestial bodies, spaceships, automobiles, ships, etc. on the basis of the laws of Newtonian mechanics, making use of the Galilean transformations, and obtain excellent results. Isn't there a contradiction here?

As it turns out, there is no contradiction whatsoever. The point is that the bodies mentioned above move at velocities that are very much lower than that of light in a vacuum. In such cases the relativistic equations automatically reduce to the Newtonian ones with an accuracy sufficient for all practical purposes.

2. This can be demonstrated by supposing a body to be moving with a velocity $v = 10 \text{ km/s}$ with respect to the earth. This is the velocity of a space rocket. In ordinary engineering practice we deal with bodies travelling at much lower velocities. Let us now link a new reference frame to this body. Exact relations between the coordinates and times in the two frames are expressed by the Lorentz transformations. We can readily see, however, that if we use the Galilean transformations we obtain practically the same results.

In fact, for our case, the ratio

$$\frac{v^2}{c^2} = \left(\frac{10^4}{3 \times 10^8} \right)^2 \cong 10^{-9}$$

Hence, to distinguish the value $\sqrt{1 - v^2/c^2}$ from unity we must have a measuring instrument with an accuracy up to nine significant digits! We employ considerably less accurate instruments in practice. The same can be said of the quantity vx/c^2 ; it differs only slightly from zero.

But if we put $\sqrt{1 - v^2/c^2} \cong 1$ and $vx/c^2 \cong 0$ into the Lorentz transformations, they automatically become the Galilean transformations:

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}} \cong x - vt \quad \text{and} \quad t' = \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}} \cong t$$

Thus, to analyze phenomena occurring at a velocity considerably less than that of light in a vacuum, the Galilean transformations (and, for this reason, all formulas of Newtonian mechanics) can be used to advantage. The application of the Lorentz transformations under such conditions yields practically the same result, but the calculations will be much more complicated.

We have arrived at a very important conclusion: the *theory of relativity includes Newtonian mechanics as a limiting case of the mechanics of phenomena whose velocity is considerably less than that of light in a vacuum*.

3. This is an example that demonstrates the development of science. Any scientific theory describes a certain range of phenomena with a definite degree of accuracy that depends upon the level of development of science and measuring techniques. In the further advance of science we include a more and more extensive range of phenomena. At the same time the accuracy of our measurements steadily increases.

At a certain stage we may find that the old theory can no longer explain newly discovered phenomena. The conclusions of the old theory conflict with the new facts. Then a new theory is developed, frequently on the basis of entirely new principles. But the new theory does not discard the old as completely wrong. It includes the results of the old theory as a limiting case for the range of phenomena that was previously described satisfactorily. This proposition in science is called the *correspondence principle*.

It is often extremely difficult to disprove the basic principles of an old theory because one must refute concepts that appear to be "obvious", "self-evident" and, consequently, true. Sooner or later, however, a more advanced theory always wins. And subsequent generations of scientists often find it difficult to understand what brought about the earlier argument.

It was the same with the theory of relativity. When first proposed, it led to violent debates. Many scientists, unable to give up habitual conceptions, could not grasp its essence. Reactionaries of all kinds attacked the theory bitterly, attempting to show that it contradicted "common sense". But subsequent developments of science fully confirmed the truth of both the initial propositions and the conclusions drawn from them.

Chapter 13

MASS, MOMENTUM AND FORCE IN THE THEORY OF RELATIVITY

13.1. RELATIVISTIC MASS

1. It was assumed in Newtonian mechanics that the mass of a body is the same in various inertial frames of reference. Experiments with bodies moving at velocities much less than that of light seemed to confirm this statement. At these velocities no dependence of mass on velocity could be detected. This is why, in all the previous calculations, we assumed that the mass of a body is a constant value.

On the basis of this proposition, let us analyze the motion of a body subject to the action of a constant force. To simplify calculations we shall assume that at $t_1 = 0$, the initial velocity $u_1 = 0$, i.e. the body has no initial velocity. Then the fundamental law of dynamics becomes

$$F = mu/t \quad \text{or} \quad u = Ft/m \tag{13.1}$$

Under the assumption that the force and mass are constant, $a = F/m = \text{const}$. Hence, the velocity of a body is directly propor-

tional to the time the force acts. Thus

$$u = at \quad (13.2)$$

This means that the velocity of a body increases without limit if the constant force is applied long enough.

2. But this conclusion disagrees with the theory of relativity, one of whose fundamental conclusions states that *no body can move*

with a velocity exceeding the velocity of light in a vacuum (see Sec. 12.6)

It will be necessary, then, to revise the fundamental law of dynamics so that its conclusions agree with those of the theory of relativity. It turned out that this could be done by assuming that the *mass of a body differs in different frames of reference*. This is expressed by the relationship (Fig. 13.1):

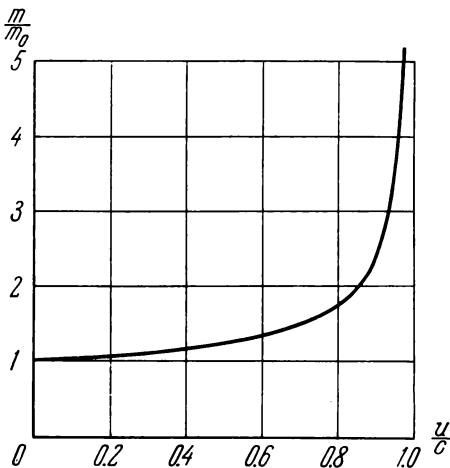


Fig. 13.1

body is at rest. It is called the *proper mass*, or *rest mass*. The quantity m is the mass of the same body in the frame of reference with respect to which it is moving with a velocity u . This is called the *relativistic mass* (see Sec. 21.8).

3. Momentum is defined in the theory of relativity in the same way as in Newtonian mechanics:

$$\mathbf{p} = mu \quad (13.4)$$

But, in contrast to Newtonian mechanics, the mass in equation (13.4) is the relativistic mass rather than the rest mass.

13.2. THE FUNDAMENTAL LAW OF DYNAMICS IN THE THEORY OF RELATIVITY

1. The fundamental law of dynamics [equations (7.2) and (7.4)] retains its definition: the *force is equal to the change in momentum in unit time*, i.e. the rate of change in momentum. Thus

$$\mathbf{F} = \frac{\mathbf{p}_2 - \mathbf{p}_1}{t_2 - t_1} = \frac{m_2 \mathbf{u}_2 - m_1 \mathbf{u}_1}{t_2 - t_1} \quad (13.5)$$

Now we shall demonstrate that the consequences of the fundamental law of dynamics, when it is written as above, do not contradict the theory of relativity.

2. Making the same assumptions ($F = \text{const}$, $t_1 = 0$ and $u_1 = 0$), we can write

$$F = \frac{mu}{t} = \frac{m_0 u}{t \sqrt{1 - u^2 c^2}} \quad (13.6)$$

Putting $F/m_0 = b = \text{const}$ and solving for the velocity u , we obtain

$$u = \frac{bt}{\sqrt{1 + b^2 t^2 / c^2}} \quad (13.7)$$

Let us examine this result.

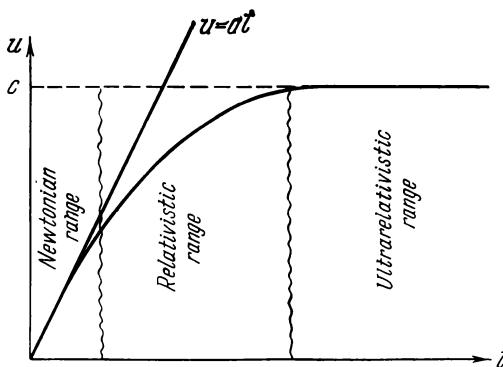


Fig. 13.2

3. For sufficiently short time intervals the product bt is considerably less than the velocity of light in a vacuum. Hence, $b^2 t^2 / c^2 \ll 1$ and, approximately,

$$u \cong bt$$

We have thus obtained the classical approximation as the limiting case for low velocities.

If the constant force acts long enough, the value of bt can become considerably greater than c . Then $b^2 t^2 / c^2 \gg 1$ and, approximately,

$$u \cong \frac{bt}{\sqrt{b^2 t^2 / c^2}} = c$$

Thus, after prolonged application of a constant force the *velocity of a body approaches that of light* instead of increasing without limit, as would follow from the laws of Newtonian mechanics (see also Fig. 13.2).

13.3. THE RELATION BETWEEN NEWTONIAN AND RELATIVISTIC DYNAMICS

1. It was shown in Sec. 12.11 that the theory of relativity does not annul Newtonian kinematics but incorporates it as a special case for relatively slow-moving bodies. The preceding section leads to the conclusion that the same is true for dynamics. Specifically, we have proved that it is quite permissible to apply the laws of Newtonian mechanics for relatively slow motion, but only the theory of relativity yields correct results in dealing with high velocities.

Let us attempt to introduce a criterion enabling us to decide which motions are slow and which have a velocity high enough to require relativistic treatment.

Let us assume that the apparatus we are using enables us to measure quantities with an accuracy to n significant digits. Then, if the relative error (i.e. the difference between two values divided by the quantity being measured) is less than 10^{-n} we shall not be able to detect it. Next we shall calculate at what velocity it will still be impossible to detect any change in the mass of a body. The relative error in measuring the mass is

$$\frac{\Delta m}{m} = \frac{m - m_0}{m} = 1 - \sqrt{1 - u^2/c^2} \quad (13.8)$$

This error should be less than 10^{-n} :

$$1 - \sqrt{1 - u^2/c^2} < 10^{-n} \text{ or } 1 - 10^{-n} < \sqrt{1 - u^2/c^2}$$

Since unity exceeds 10^{-n} , we can square both sides of the inequality:

$$1 - 2 \times 10^{-n} + 10^{-2n} < 1 - u^2/c^2 \text{ or } u < c \sqrt{2 \times 10^{-n} - 10^{-2n}}$$

Taking into account that $10^{-2n} \ll 10^{-n}$, we can write

$$u < c \sqrt{2 \times 10^{-n}}$$

Assume, for instance, that measurements are made with an accuracy to six significant digits ($n = 6$). Then $u < c \sqrt{2 \times 10^{-6}} = 423 \text{ km/s}$. Thus, at velocities under *four hundred kilometres per second* the rest mass differs from the relativistic mass by less than 10^{-6} , i.e. less than one ten-thousandth of one per cent!

Under realizable conditions the velocity of large bodies is substantially less than this limit: even the velocity of space rockets is only about 10 km/s, which is about one fortieth of the calculated value. And measurements are rarely made in engineering with such accuracy. It is quite clear, then, that the application of the laws of Newtonian mechanics for calculating the motion of bodies will yield ideal results, insofar as their accuracy is concerned.

In the world of atomic particles, however, velocities are frequently encountered that are near to that of light in a vacuum. Correct results are obtained only by making use of the laws of the theory of relativity. It was found possible, in particular, in analyzing the motion of fast-moving elementary particles to check experimentally the formula for relativistic mass, i.e. equation (13.3). The results of the experiment agreed excellently with theory.

2. These considerations enable us to introduce the following classification of motion (Fig. 13.2).

Newtonian range. The velocities of bodies are so low that measuring apparatus cannot detect the relativistic effects of time dilatation, contraction of length, increase of mass, etc. It is entirely permissible to apply the laws of Newtonian mechanics in this velocity range.

Relativistic range. The velocities are sufficiently high, and the relativistic effects become fully measurable. Obviously, only the theory of relativity yields correct results in this range.

Ultrarelativistic range. The velocity of bodies almost reaches that of light in a vacuum. More precisely, the difference between the velocity of a body and the velocity of light is less than the sensitivity of the measuring instrument. Here also, of course, correct results require the application of the theory of relativity.

In our example, where the measuring accuracy does not exceed 10^{-6} , the classical range consists of velocities less than 400 km/s; the ultrarelativistic range, of velocities that differ from the velocity of light by less than 300 m/s. Velocities between these two ranges belong to the relativistic range.

3. Consider the change in the momentum of a body subject to a constant force. Under the conditions that at $t_1 = 0$ the initial momentum $p_1 = 0$, it follows from equation (13.5) that

$$p = Ft \quad (13.9)$$

Thus, when a constant force is applied to a body, the momentum of the body increases proportionally to the time.

In the Newtonian range, momentum increases only because of the increase in velocity; the change in mass can be neglected. In the ultrarelativistic range, on the other hand, the velocity of the body practically does not change, and the momentum increases only as a result of the increase in mass. In the intermediate (relativistic) range the increase in momentum is accounted for by the increase in both factors, velocity and mass. Here the velocity increases slower than in the Newtonian range, precisely because of the increase in mass of the accelerated body.

Chapter 14

EQUATIONS OF MOTION AND THE UNCERTAINTY RELATION

14.1. INITIAL CONDITIONS AND MEASURING APPARATUS

1. The successful application of the method discussed in Chapter 8 for solving the basic problem of dynamics in Newtonian mechanics inevitably led physicists to believe that such methods were universal and could be employed to solve *any* problems. Nobody thought of asking in all seriousness whether one could actually specify the

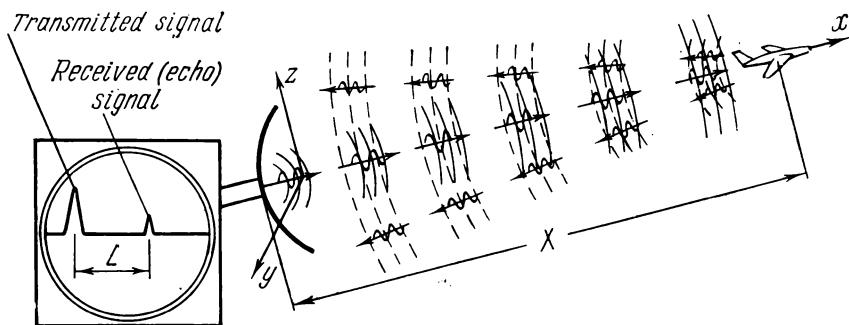


Fig. 14.1

initial conditions as precisely as required. The solutions were so patent that it seemed obvious that the initial conditions could be assigned with limitless degree of accuracy.

Of course, coordinates and momenta are measured with a definite degree of error. In the 19th century, however, it was assumed that by improving the measuring apparatus the error could be made infinitesimal. Further developments in science proved the unsoundness of this point of view.

Using an example, let us try to analyze, in general terms, the operation of instruments that can be used to measure the coordinates and velocities of bodies.

2. A radar set (derived from RAdio Detection And Ranging) is employed to determine the coordinates of an airplane (Fig. 14.1). The antenna of the set transmits radio waves in short pulses. These waves, travelling at the velocity of light, reach the airplane and, reflected by it, return to the antenna. Two marks are observed on the radar screen, from the transmitted signal and from the received (echo) signal. Knowing the velocity with which the ray moves across the

screen, u , and the velocity of radio waves ($c = 3 \times 10^8$ m/s) and measuring the distance L between the marks on the screen, we can use the proportion $\frac{L}{u} = \frac{2X}{c}$ to find the distance to the airplane, $X = \frac{c}{2u} L$. Distance L on the screen is read off directly in kilometres (because the value of $c/2u$ is constant and known). This enables us to determine the coordinate of the airplane at once from the marks, or pips, on the screen of the cathode-ray tube.

Note that it is impossible to measure the instantaneous value of the velocity with an instrument that serves to determine the instantaneous values of the coordinates. True, by observing the motion of the pip on

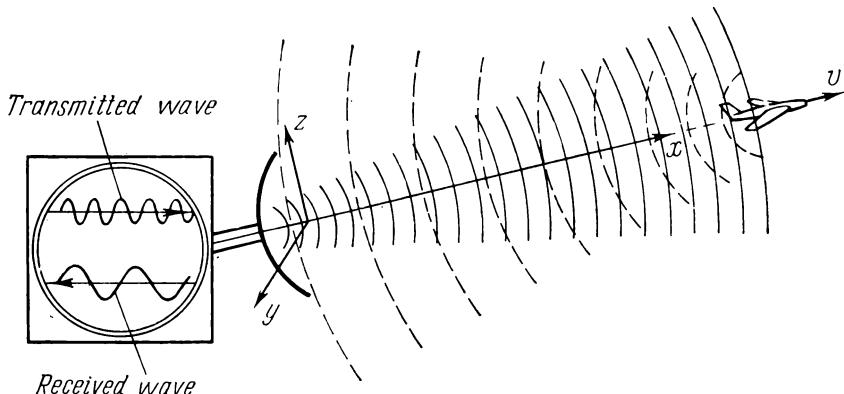


Fig. 14.2

the radar screen, we can determine the average velocity of the airplane over a certain time interval, but we cannot find the velocity of the airplane with this instrument at the instant we are measuring its coordinate.

3. Instruments have been developed for measuring the instantaneous velocity of a particle. For example, the radar set can be rebuilt to compare the frequency of oscillations in the transmitted wave with that of the wave reflected from the airplane.

If the airplane is flying with a velocity v along the axis of abscissas, the course of time, or its tempo, in a reference frame linked to the airplane changes with respect to the tempo in a frame linked to the radar set (see Sec. 12.9). The wave frequency will naturally change as the course of time changes. This will be evident on the screen of the rebuilt radar set. When the airplane is receding from the radar set, the waves are drawn out along the screen (Fig. 14.2); when it is approaching the set, the waves are crowded together (Fig. 14.3).

Applying the Lorentz transformations, we can derive the relationship between the frequency ν of the transmitted wave and the frequency ν' of the wave reflected from the airplane. Frequency is de-

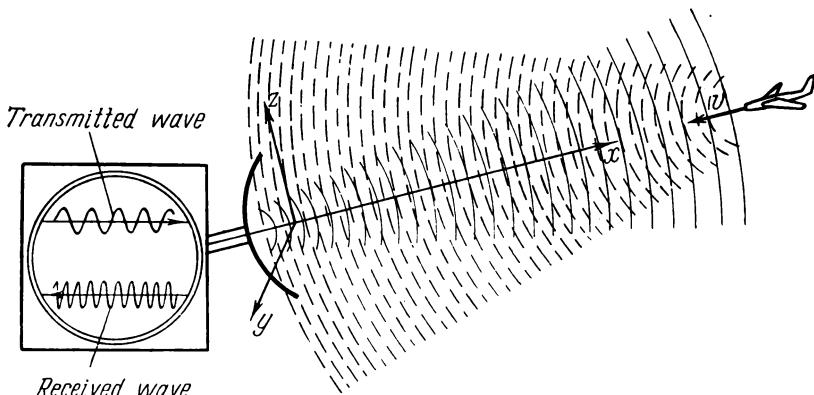


Fig. 14.3

fined as the number of oscillations per unit time. Hence, the frequency of oscillations of a wave in different reference frames is inversely proportional to the time interval between two events, e.g. between the reflection of the wave from the airplane and its reception

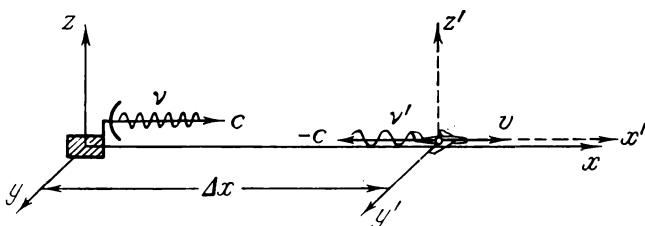


Fig. 14.4

by the radar set (Fig. 14.4). We denote the frequency by ν and the time interval by Δt in the reference frame of the radar set, and by ν' and $\Delta t'$ in the reference frame of the airplane. Then

$$\frac{\nu'}{\nu} = \frac{\Delta t}{\Delta t'} \quad (14.1)$$

But, according to equation (12.17),

$$\Delta t' = \frac{\Delta t - v \Delta x/c^2}{\sqrt{1 - v^2/c^2}}$$

where Δx is the distance travelled by the reflected wave. Since the velocity of the reflected wave is $\Delta x/\Delta t = -c$ (the minus sign because the reflected wave travels in the direction opposite to that of the axis of abscissas), we obtain

$$\Delta t' = \Delta t \frac{1+v/c}{\sqrt{1-v^2/c^2}} \quad (14.2)$$

Combining equations (14.1) and (14.2), we obtain

$$v' = v \frac{\sqrt{1-v^2/c^2}}{1+v/c} \quad (14.3)$$

Thus, the frequency of a wave reflected from a departing airplane is less than that of the wave sent out by the radar set. If the airplane is approaching the radar set, its velocity has the opposite sign and the frequency of the reflected wave increases:

$$v'' = v \frac{\sqrt{1-v^2/c^2}}{1-v/c} \quad (14.4)$$

This phenomenon is called the longitudinal *Doppler effect* in optics, after the Austrian physicist Christian Johann Doppler (1803-1853). With its aid we can readily measure the velocity of an airplane by comparing the frequencies of the transmitted and reflected waves.

The Doppler effect can also be applied to measure the velocity of a body which emits light, for instance, the velocity of stars (see Sec. 59.8).

4. But an instrument that enables us to measure the velocity of a particle (and, consequently, its momentum) does not allow us to measure its coordinate. As a matter of fact, a sine wave of definite frequency is required to measure velocities by means of the Doppler effect, while to measure the coordinates of an airplane the radar set must emit chains of short pulses. These are never of the sine type.

Therefore, to measure the coordinates and velocity (or momentum) of a moving body simultaneously it is necessary to use two instruments, one for measuring the coordinate and the other for measuring the velocity.

Niels Bohr (1885-1962), one of the founders of modern quantum physics, demonstrated that all measuring instruments can be divided into two classes in such a way that the instruments of one class yield data which are *complementary* to those provided by instruments of the other class. For example, one class of instruments are for measuring coordinates, and the other for measuring momenta or velocities. Full data on the motion of a body can be obtained only with the joint application of instruments of both kinds.

14.2. THE UNCERTAINTY RELATION

1. Instruments used to measure coordinates and velocities interact with the body whose parameters are being measured. Thus, when a radio wave overtakes and is reflected from an airplane, it exerts a definite force on the airplane, changing its coordinate and velocity.

In the case of an airplane, of course, these changes are so minute that they can be neglected even in the most precise calculations. However, as science penetrated into the subatomic world, it became clear that a measuring instrument can introduce appreciable disturbances into the motion of microparticles, and that the action of the instrument on the object cannot be neglected. This is associated with the special features of microparticles, which we will take up in the chapters devoted to quantum mechanics.

2. Werner Heisenberg (b. 1901) analyzed the results of all possible experiments that could be conducted for measuring the coordinates, or position, of a particle and its momentum (or velocity). He came to the conclusion that, owing to the special features of microparticles, simultaneous measurement of their coordinates and the corresponding momenta introduces uncertainties (variations) in their values, denoted by Δx and Δp_x . These uncertainties are related by the equation

$$\Delta x \times \Delta p_x \cong \hbar \quad (14.5)$$

where Planck's constant $h = 2\pi\hbar$; $\hbar = 1.05459 \times 10^{-34}$ kg-m²s. This is one of the most important of the universal physical constants, and is of especially great significance in describing phenomena of the subatomic world. This quantity is frequently encountered in quantum mechanics.

The coordinates and corresponding momenta along the other axes are related by similar equations:

$$\Delta y \times \Delta p_y \cong \hbar \text{ and } \Delta z \times \Delta p_z \cong \hbar$$

14.3. THE UNCERTAINTY RELATION AND CLASSICAL MECHANICS

1. If we compare the method of solving the basic problem of dynamics with Heisenberg's uncertainty principle, we see that they contradict each other. In fact, to set up the motion equations, the initial conditions must be specified with maximum accuracy. But the uncertainty principle shows that the initial conditions can be determined only approximately, and that the degree of approximation is not unlimited. How then can the law of motion, or motion equations, be found in classical mechanics with an arbitrary pre-assigned accuracy.

We shall try to answer this question by analyzing three problems. The first considers the motion of a macroscopic body (a rocket); the second, the motion of a microparticle (an electron) along a macroscopic path; the third, the motion of a microparticle along a micropath.

2. Problem 1. A rocket with a mass of 10^3 kg is in circular orbit about the earth. The orbit radius is 6500 km; the velocity of the rocket is 8 km/s. With what accuracy can the radius and velocity be specified?

To keep the rocket in circular orbit the velocity vector must be perpendicular to the radius at each point of the path. This means that the component of the velocity along the radius must be zero. But if this component is exactly zero, its uncertainty Δv_r is also equal to zero. Then, according to relation (14.5), the uncertainty of the value of the radius becomes infinitely large. Here, however, we can compromise. Assume that the orbit radius is specified with the uncertainty Δr and the radial component of the velocity is zero but with the uncertainty Δv_r , that can be determined from the Heisenberg relation (Fig. 14.5):

$$\Delta r \times \Delta p_r \approx \hbar$$

If the uncertainty of r and v is much less than the values themselves, we can consider the problem to be solved with sufficient accuracy.

Suppose in our problem that $\Delta r = 1 \text{ \AA} = 10^{-10} \text{ m}$, an uncertainty approximately equal to the diameter of the atom. Obviously, the error of any real measuring instrument is incomparably greater. Then the uncertainty of the radial component of the rocket's momentum is

$$\Delta p_r \approx \frac{\hbar}{\Delta r} = \frac{10^{-34}}{10^{-10}} = 10^{-24} \text{ kg-m/s}$$

and the uncertainty of the radial component of the velocity is

$$\Delta v_r = \frac{\Delta p_r}{m} \approx \frac{\hbar}{m \Delta r} = \frac{10^{-34}}{10^3 \times 10^{-10}} = 10^{-27} \text{ m/s}$$

But the orbital velocity of the rocket, $8 \times 10^3 \text{ m/s}$, is incommensurably larger than the calculated uncertainty!

As we see, the uncertainty is of no significance whatsoever in the motion of a macroscopic body. We can entirely neglect the uncertainty of the coordinates and velocities, assuming that the initial con-

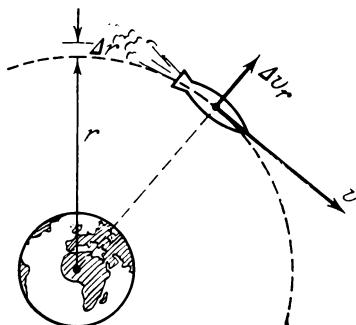


Fig. 14.5

ditions can be specified with any accuracy we require, and solving the basic problem of dynamics by classical methods.

3. Problem 2. An electron is travelling in a betatron (a device for accelerating electrons) in a circular path of radius 2.5 m with a velocity 99 per cent that of light, i.e. $v = 0.99 c = 2.97 \times 10^8$ m/s. With what accuracy can the orbital radius and velocity of the electron be specified?

In contrast to the preceding problem, it will be necessary to take into account the dependence of mass on velocity. The rest mass of an electron is $m_0 = 9.1 \times 10^{-31}$ kg. At this velocity its mass is

$$m = \frac{m_0}{\sqrt{1-v^2/c^2}} = \frac{m_0}{\sqrt{1-0.99^2}} = \frac{m_0}{\sqrt{1.99 \times 0.01}}$$

$$= \frac{m_0}{1.411 \times 0.1} = 7.1 m_0$$

Since the mass of an electron is much less than that of a rocket, the radius of the electron's orbit cannot be specified with as high a degree of accuracy as in the preceding problem. However, if we assume that the uncertainty of the radius is $\Delta r = 0.05$ mm, the orbit will be defined very accurately, the relative error being

$$\frac{\Delta r}{r} = \frac{0.05}{2.5 \times 10^3} = 2 \times 10^{-5} = 0.002\%$$

The uncertainty of the radial component of the velocity

$$\Delta v_r \cong \frac{\hbar}{m \Delta r} = \frac{10^{-34}}{7 \times 9 \times 10^{-31} \times 5 \times 10^{-5}} \cong 0.3 \text{ m/s}$$

At a velocity of the electron almost equal to that of light a radial component of 30 cm/s can certainly be neglected and set at zero. Therefore, here as well, when a microparticle travels along a macroscopic path, the uncertainty principle is of no actual significance. The initial conditions can again be specified with any required degree of accuracy and the problem can be solved by classical methods (taking the relativistic effects into consideration).

Further on we shall find that the complex nature of elementary particles (electrons, protons, neutrons, etc.) does not allow them, strictly speaking, to be dealt with as material points (or particles, in the sense we have defined the term in mechanics), and their paths cannot be determined so simply. However, in cases similar to the one just discussed, such an approach to the solution of the problem concerning their motion is quite permissible.

4. Problem 3. Let us consider the motion of a microparticle on a micro-orbit, for example, the motion of an electron in an atom. The radius of the atom is approximately $r \cong 0.5 \text{ \AA} = 5 \times 10^{-11}$ m; the orbital velocity of an electron is approximately $v \cong 10^6$ m/s. At this velocity relativistic effects can be neglected, and the mass

of the electron can be taken equal to its rest mass. We assume the uncertainty of the radius to be one per cent of the radius, i.e. $\Delta r = 0.01 r = 5 \times 10^{-13}$ m. Then the uncertainty of the radial component of the velocity is

$$\Delta v_r \cong \frac{\hbar}{m\Delta r} = \frac{10^{-34}}{9 \times 10^{-31} \times 5 \times 10^{-13}} = 2.2 \times 10^8 \text{ m/s}$$

which is almost the velocity of light.

What do we find? The uncertainty of the radial component of the velocity is 220 times greater than the orbital velocity. There is obviously no point in considering the motion of an electron on its orbit because its velocity is absolutely indeterminate.

On the other hand, if we assume a reasonable uncertainty of the radial component of the velocity, for instance, $\Delta v_r/v = 1$ per cent, we can say nothing of the path. In fact, if $\Delta v_r = 0.01 v = 10^4$ m/s, the uncertainty of the radius is

$$\Delta r \cong \frac{\hbar}{m\Delta v_r} = \frac{10^{-34}}{10^4 \times 9 \times 10^{-31}} = 1.1 \times 10^{-8} \text{ m} = 110 \text{ \AA}$$

which is 220 times as much as the radius itself! Evidently, it is quite meaningless to speak of any definite orbit under such conditions.

5. We see, therefore, that the uncertainty principle does not influence the applicability of the methods of Newtonian mechanics or the theory of relativity to *macroscopic* objects. Using the forces and the initial conditions, we can find the law of motion of a body construed as a particle with any pre-assigned degree of accuracy. The uncertainty principle does not actually limit the application of these laws to *microparticles* that travel in *macroscopic* regions of space. It makes sense, for example, to speak of the path of an electron in a radio tube, cathode-ray tube, accelerator, etc.

In microregions, the atom, atomic nucleus, crystal lattices of solids etc., it is an entirely different matter. Here, owing to the uncertainty principle, methods of classical mechanics turn out to be absolutely inapplicable. Such classical concepts as the law of motion of a particle, its path, instantaneous velocity and others become meaningless. There is no point in speaking of the path of an electron in an atom or of the velocity of a proton in an atomic nucleus. Such problems require methods of *quantum mechanics* for their solution.

PART TWO ▼▼ CONSERVATION LAWS

Chapter 15

LAW OF CONSERVATION OF LINEAR MOMENTUM

15.1. CLOSED SYSTEM OF BODIES

1. A precise solution of the character of motion of a certain body requires that all the forces be taken into account that act on the body due to the interaction with it of all the bodies surrounding it. But there is actually an infinitely great number of such forces. Moreover, such forces are variable because the surrounding bodies have very complex motion due to their interaction with one another. In consequence, no feasible solution can be found to a problem when stated this way, due to the insurmountable mathematical difficulties.

The problem can be substantially simplified when we consider that not all the surrounding bodies act on the given body with the same forces. If, for instance, a sputnik, or artificial satellite, is orbiting the earth at a radius of about $r = 8000$ km, the sun acts on it with a force that is much less than the force of attraction of the earth. Thus

$$\frac{F_e}{F_{\text{sun}}} = \frac{GmM_eR_e^2}{r^2GmM_{\text{sun}}} = \frac{M_eR_e^2}{M_{\text{sun}}r^2}$$

where $R_e = 1$ AU (astronomical unit) = 1.5×10^8 km

$$M_e = 6 \times 10^{24} \text{ kg}$$

$$M_{\text{sun}} = 2 \times 10^{30} \text{ kg.}$$

Substituting the numerical values we find that the attraction of the earth is 1000 times greater than the force exerted on the sputnik by the sun. We shall leave it to the student to show that the force exerted by the moon is only 1/200 as much as that of the earth, and that the action of the other planets is negligibly small.

2. This calculation suffices to show that, as a first approximation, we can disregard the action of all bodies on the sputnik except the earth. Consequently, we can deal with a *system consisting of two*

bodies, the sputnik and the earth, and assume that their interaction in the main determines the motion of the sputnik. All other bodies can be regarded as *external* with respect to this system and their action can be taken into consideration in the form of corrections to the main force.

3. It is customary to call forces with which component parts of a given system interact *internal forces*. Forces with which the whole system or its separate component bodies interact with surrounding bodies are said to be *external*.

A system of bodies is said to be *closed* (or isolated) if, in comparison to the internal forces, the external forces can be neglected.

4. Thus, in the example given above, the system of bodies consisting of the earth and the sputnik can, as a first approximation, be regarded as a closed one.

To an even greater degree of accuracy, the solar system is a closed one. In fact, the forces of interaction between the sun and the planets considerably exceed the forces exerted on the planets by even the nearest stars. The star nearest to the solar system is located at the vast distance of $R = 4.5$ light years* $= 4.2 \times 10^{13}$ km, while the distance from the earth to the sun is only $R_e = 1.5 \times 10^8$ km. Assuming that the mass of the star is approximately equal to that of the sun we obtain

$$\frac{F_{\text{sun}}}{F_{\text{star}}} = \frac{GmMR^2}{R_e^2 GmM} = \frac{R^2}{R_e^2} = \left(\frac{4.3 \times 10^{13}}{1.5 \times 10^8} \right)^2 \cong 8 \times 10^{10}$$

Hence the internal force (i.e. the force of interaction between the sun and the earth) is 80 thousand million times greater than the external force (i.e. the force exerted on the earth by the nearest star). Naturally, we have every reason to regard the system (sun + earth) as a closed one.

5. The concept of a closed system is a very useful abstraction because in such systems all phenomena are described by means of the simplest and most general laws. Therefore, wherever possible, it proves expedient to abstract oneself from the action of the external forces and deal with a given system of bodies as being closed. Then, if necessary, the solution obtained as a first approximation should be corrected by taking into account the disturbances introduced by the action of the external forces.

15.2. LAW OF CONSERVATION OF LINEAR MOMENTUM

1. Valid for closed systems, the law of conservation of linear momentum can be stated as follows: the *total linear momentum of a closed system remains constant* for any processes that occur within the system.

* A light year is the distance light travels in a vacuum in one year.

This does not imply that this law requires a constant momentum of each body included in the system. The contrary is true: owing to the action of the internal forces, the momenta of the bodies in the system are continually varying. *Only the vector sum of the momenta of all the component parts of the system remains constant.*

2. We shall derive the law of conservation of linear momentum from Newton's second and third laws. For the sake of simplicity we shall limit ourselves to the case when the system comprises two bodies. A discussion of the more general case, when the system consists of many bodies, would not yield anything new in principle, but would considerably complicate the derivation.

Assume that at the instant t_1 the first body has the mass m'_1 and the velocity \mathbf{v}'_1 , and the second body the mass m'_2 and velocity \mathbf{v}'_2 , while at instant t_2 the masses and velocities are, respectively, m_1 , \mathbf{v}_1 , m_2 and \mathbf{v}_2 . For the first body Newton's second law can be written as

$$\mathbf{F}_{12} = \frac{m_1 \mathbf{v}_1 - m'_1 \mathbf{v}'_1}{t_2 - t_1}$$

and for the second body as

$$\mathbf{F}_{21} = \frac{m_2 \mathbf{v}_2 - m'_2 \mathbf{v}'_2}{t_2 - t_1}$$

According to Newton's third law, $\mathbf{F}_{21} = -\mathbf{F}_{12}$ or

$$\frac{m_2 \mathbf{v}_2 - m'_2 \mathbf{v}'_2}{t_2 - t_1} = -\frac{m_1 \mathbf{v}_1 - m'_1 \mathbf{v}'_1}{t_2 - t_1}$$

Next we cancel the denominators, transfer all unprimed terms to the left-hand and all primed ones to the right-hand side of the equation, and obtain

$$m_2 \mathbf{v}_2 + m_1 \mathbf{v}_1 = m'_2 \mathbf{v}'_2 + m'_1 \mathbf{v}'_1 \quad (15.1)$$

or, more concisely,

$$m_2 \mathbf{v}_2 + m_1 \mathbf{v}_1 = \text{const} \quad (15.2)$$

for any instant of time.

If the closed system consists of several bodies, the law of conservation of linear momentum is written as:

$$m_1 \mathbf{v}_1 + m_2 \mathbf{v}_2 + \dots + m_n \mathbf{v}_n = \text{const} \quad (15.3)$$

3. In deriving the law of conservation of momentum we used only Newton's laws, and then in a form which is valid both for relativistic and Newtonian mechanics. Consequently, the law of conservation of linear momentum is applicable in both Newtonian and relativistic mechanics, but in the latter the dependence of mass on velocity must be taken into account.

15.3. RECOIL PHENOMENA

1. A common phenomenon is the splitting of a body into two parts under the action of internal forces. If the internal forces are very much greater than the external ones, the system can be assumed to be closed, and the law of conservation of linear momentum can be applied.

To simplify the discussion we shall limit ourselves to the case in which the body was at rest with respect to the reference frame before it was split into parts. Then we can affirm that the two new bodies produced by splitting will move in opposite directions at velocities inversely proportional to their masses. Let the masses of the bodies equal m_1 and m_2 , and their corresponding velocities, v_1 and v_2 . According to the law of conservation of linear momentum

$$\mathbf{p} = \mathbf{p}_1 + \mathbf{p}_2$$

Since, before splitting, the body was at rest with respect to the reference frame: $\mathbf{p} = 0$. Then

$$0 = m_1 v_1 + m_2 v_2 \quad (15.4)$$

or

$$v_2 = -\frac{m_1}{m_2} v_1 \quad (15.5)$$

which proves that vectors v_1 and v_2 are opposite in direction.

2. Recoil phenomena are observed, for instance, in firing a shell from a field gun. If the gun is not secured it will recoil in the direction opposite to the flight of the shell. In a machine gun, submachine gun or any automatic gun, the recoil of the bolt is utilized to extract the fired cartridge case and to cock the gun.

The motion of all transportation facilities is associated with recoil phenomena. In the rotation of the driving wheels of an automobile, a friction force is developed between the tires and the earth. This is an internal force (in the system consisting of the automobile and the earth); as a result of it the automobile moves in one direction and the earth in the opposite direction. The earth moves, of course, at a negligibly low velocity, because the mass of the earth is incommensurably greater than that of the automobile.

A ship travels in exactly the same way: its propeller pushes the water back beyond the stern, and, consequently, the ship moves forward.

3. Many phenomena in nuclear physics are accompanied by recoil too. A uranium nucleus, for example, splits into two fragments of approximately equal mass if it absorbs a neutron. Since, before the reaction, the nucleus and neutron move at very low (thermal) velocities, it can be assumed with some approximation that before fission the

system was at rest. But then the fragments produced in fission must move in opposite directions. This can be clearly seen in the photograph of Fig. 15.1, obtained in a Wilson cloud chamber (see Sec. 36.9).

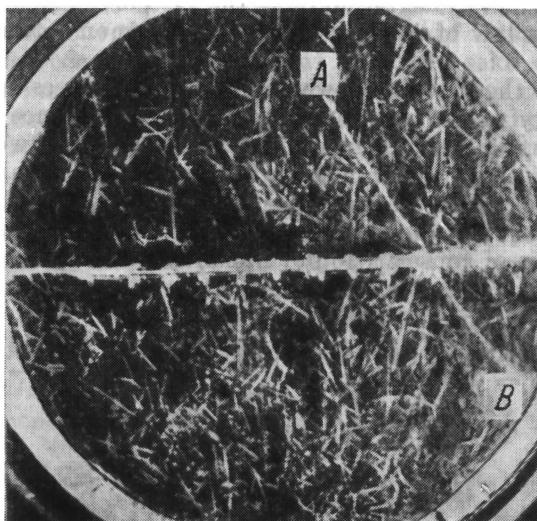


Fig. 15.1

in the middle of which is a plate with a layer of uranium oxide. Tracks *A* and *B* were made by two fission fragments recoiling in opposite directions.

Further on (Sec. 17.4) we shall discuss still another phenomenon associated with recoil in nuclear reactions.

15.4. MEASUREMENT OF MASS

Making use of the law of conservation of linear momentum we can compare the masses of two bodies without weighing them. We tie two bodies of different masses together with a string and place a compressed spring between them. If we cut the string, the two bodies begin to move along a certain axis in opposite directions at the velocities v_1 and v_2 . Since, according to equation (15.5), the velocities are inversely proportional to the masses, we can determine the mass of one of the bodies without weighing it if we know their velocities and the mass of the other body.

This method of measuring masses has practically no application in mechanics because it is quite difficult to measure the velocities of the bodies since, due to friction, these velocities continuously vary.

In nuclear physics, however, where the velocity of a particle can be measured by its track, this method proves useful, for instance, to compare the masses of fragments produced in nuclear fission.

15.5. JET PROPULSION (PROPULSION BY REACTION)

1. The motion of a rocket is explained by the same law as recoil phenomena, i.e. the law of conservation of linear momentum. As the fuel burns, gases are ejected at a high velocity through the exhaust. As a result, the rocket moves in the opposite direction in such a way that the sum of the momenta of the rocket and gases remains constant.

Assume that at a certain instant t_0 of time the mass of the rocket (together with the fuel and oxidizer) equals m and its velocity is v with respect to the earth. At the instant t_1 , when a certain amount of the fuel has been burned, the mass of the rocket will be m_1 and its velocity with respect to the earth, v_1 . The velocity of the gases relative to the rocket—*exhaust velocity*—is equal to u and is opposite to the velocity of the rocket. Then the velocity of the gases with respect to the earth at instant t_0 is equal to the difference in the velocity of the rocket and the exhaust velocity: $v_{\text{gas}} = v_1 - u$.

The momentum of the rocket at the instant t_0 equals $p_0 = mv$. At instant t_1 the total momentum of the rocket and gases is

$$p_1 = m_1 v_1 + m_{\text{gas}} v_{\text{gas}} = m_1 v_1 + (m - m_1) (v_1 - u)$$

Since the rocket and the gas constitute a closed system, the law of conservation of linear momentum is applicable. Thus

$$mv = m_1 v_1 + (m - m_1) (v_1 - u) \quad (15.6)$$

Removing the brackets and rearranging the terms we obtain

$$m(v - v_1) = u(m_1 - m)$$

But $\Delta v = v_1 - v$ equals the increment of velocity of the rocket, and $\Delta m = m_1 - m$ equals the mass of the fuel burned during the time $\Delta t = t_1 - t_0$. Therefore

$$-m\Delta v = u\Delta m \quad (15.7)$$

2. Gases ejected from the exhaust nozzle of the rocket act on the rocket with a certain force called *jet*, or *reactive, thrust*. It can be determined by making use of the fundamental law of dynamics, equation (7.2). For this purpose we divide both sides of equation (15.7) by Δt . Since $m\Delta v/\Delta t = F$ is the *propulsive force* and $\mu = \Delta m/\Delta t$ is the *fuel consumption per second*, we obtain

$$F = -\mu u \quad (15.8)$$

Hence, the jet, or reactive, thrust is proportional to the fuel consumption per second and to the exhaust velocity of the gases; it is in the direction opposite to gas exhaust.

15.6. ROCKET FUEL CALCULATIONS

1. To get an idea of the amount of rocket fuel required to put a rocket into orbit we shall carry out approximate calculations. Assume that a rocket with a mass of about 10 metric tons is to be lifted with an acceleration of $a = 5g = 49 \text{ m/s}^2$. The propulsive force is $F = ma = 49 \times 10^4 \text{ N}$. The gas exhaust velocity for modern rockets is $u = 4 \text{ km/s}$. Then the fuel consumption is

$$\mu = \frac{F}{u} = \frac{49 \times 10^4}{4 \times 10^3} = 122.5 \text{ kg/s}$$

A rocket moving with an acceleration of about 50 m/s^2 will reach the circular orbital velocity (8 km/s) in the time

$$t = \frac{v}{a} = \frac{8000}{50} = 160 \text{ s}$$

Therefore, the total amount of fuel required is

$$M_{\text{fuel}} = \mu t = 122.5 \text{ kg/s} \times 160 \text{ s} \approx 19600 \text{ kg} \approx 20 \text{ metric tons}$$

2. The preceding calculation is no more than a rough approximation and is unsuitable even for a crude estimate of the amount of fuel

required to put a rocket into orbit. The point is that a rocket is a body of variable mass since its mass continually decreases as the fuel burns out. Therefore, to put a 10-ton rocket into orbit, its mass at takeoff, even according to our approximate calculations, would have to be about 30 metric tons. To impart the required acceleration to this mass, we would need a propulsive force three times greater. This would lead to an increase in fuel consumption and, inevitably, a much greater amount would be required.

3. The required amount of fuel can be calculated by the Tsiolkovsky formula which we give here without proof:

$$\frac{v}{u} = 2.3 \log \frac{M_0}{M} \quad (15.9)$$

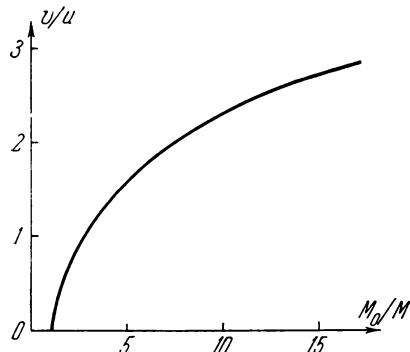


Fig. 15.2

where v = final velocity of the rocket (the initial velocity $v_0 = 0$)

u = gas exhaust velocity

M = mass of the rocket in orbit

$M_0 = M_{\text{fuel}} + M$ = initial mass of the rocket with fuel.

The graph of this function is illustrated in Fig. 15.2.

Putting $v = 8 \text{ km/s}$, $u = 4 \text{ km/s}$ and $M = 10 \text{ metric tons}$, we obtain

$$\log \frac{M_0}{M} = \frac{8}{4 \times 2.3} = 0.870$$

from which $M_0/M = 7.42$, $M_0 = 74.2 \text{ metric tons}$ and $M_{\text{fuel}} = 64.2 \text{ metric tons}$ —almost 3.5 times as much as obtained in the crude estimate. A much greater amount of fuel is required—about 160 metric tons—to reach the escape velocity (11.2 km/s).

15.7. CENTRE OF MASS

Assume that two particles with masses m_1 and m_2 are located along the axis of abscissas at points with the coordinates x_1 and x_2 . The distance between the particles is $l = x_2 - x_1$ (Fig. 15.3). We shall

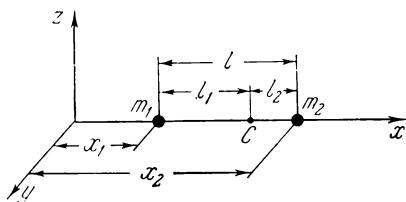


Fig. 15.3

call point C , which divides the distance between the particles into lengths inversely proportional to the masses of the particles, the *centre of mass* (or *centre of inertia*) of the given system of particles. Thus, by definition,

$$\frac{l_1}{l_2} = \frac{m_2}{m_1} \quad (15.10)$$

Since $l_1 = x_c - x_1$ and $l_2 = x_2 - x_c$, where x_c is the coordinate of the centre of mass, then

$$m_1(x_c - x_1) = m_2(x_2 - x_c)$$

from which

$$x_c = \frac{m_1 x_1 + m_2 x_2}{m_1 + m_2} \quad (15.11)$$

This expression can be readily generalized for the case of several particles located arbitrarily. The abscissa of the centre of mass is

$$x_c = \frac{m_1 x_1 + m_2 x_2 + \dots + m_n x_n}{m_1 + m_2 + \dots + m_n} \quad (15.12)$$

Similar expressions are obtained for the coordinates y_c and z_c , along the y - and z -axes, of the centre of mass of a system of particles.

15.8. MOTION OF THE CENTRE OF MASS

1. We shall solve the problem of the motion of the centre of mass for a case when the bodies concerned are moving at velocities that are much less than that of light, and the masses of the bodies remain constant.

Writing equation (15.12) for two different instants of time and subtracting one from the other we obtain

$$\Delta x_c = \frac{m_1 \Delta x_1 + m_2 \Delta x_2 + \dots + m_n \Delta x_n}{m_1 + m_2 + \dots + m_n} \quad (15.13)$$

After dividing both sides of the equation by $\Delta t = t_2 - t_1$ and putting $\Delta x_c / \Delta t = v_c^{(x)}$ (component of the velocity vector along the axis of abscissas), we obtain

$$v_c^{(x)} = \frac{m_1 v_1^{(x)} + m_2 v_2^{(x)} + \dots + m_n v_n^{(x)}}{m_1 + m_2 + \dots + m_n} \quad (15.14)$$

Similar expressions are obtained for the velocity components along the y - and z -axes. The vector equation for the velocity of the centre of mass has the form

$$\mathbf{v}_c = \frac{\mathbf{m}_1 \mathbf{v}_1 + \mathbf{m}_2 \mathbf{v}_2 + \dots + \mathbf{m}_n \mathbf{v}_n}{\mathbf{m}_1 + \mathbf{m}_2 + \dots + \mathbf{m}_n} = \frac{\mathbf{P}}{M} \quad (15.15)$$

where M = total mass of the system

\mathbf{P} = its total linear momentum.

2. If we are dealing with a closed system of particles, its total momentum remains constant. But then the velocity of its centre of mass is also constant.

In other words, the *centre of mass of a closed system has inertial motion*, i.e. it moves at uniform velocity in a straight line regardless of the motion of the separate bodies making up the system.

Note should be taken of the meaning of this statement. Internal forces may act in a closed system. As a result, bodies of the system may move with acceleration and their velocities (and momenta) may continuously vary. This, however, has no effect on the motion of the centre of mass. Hence, the *velocity of the centre of mass is not changed by internal forces*.

3. The student can easily find many examples that confirm this statement. If you are in an automobile, you can press against the

walls with any arbitrary force; this has no effect on the motion of the automobile: if it was at rest with respect to the earth, it will remain at rest; if it is in motion, it will continue at the same velocity.

It is equally clear that Baron Munchausen was exaggerating (to put it mildly) when he told how he had pulled himself by his hair out of a bog (Fig. 15.4). Since his centre of mass was sinking, no efforts that he himself could apply to his hair could change the rate at which he sank.

4. If a man walks along a rowboat on a lake, the boat will move in the opposite direction. The reason is that the friction between the



Fig. 15.4

boat and the water is negligibly small at very low velocities (see Sec. 11.8) and, therefore, the man and the boat constitute a closed system. Motion of the man in one direction causes motion of the boat in the opposite direction so that the centre of mass C_s of the system remains at the same place (Fig. 15.5).

This principle is valid for rotational motion as well as translational motion. It is, for instance, inexact to speak of the moon revolving about the earth. Actually, the system (earth + moon) rotates about the common centre of mass. Let us find its location. The distance between the centres of the earth and moon equal 384 000 km, and the mass of the earth is 81 times that of the moon. From equation (15.12) it follows that the distance from the centre of mass to the earth's centre is

$$x_c = \frac{(1 \times 384\,000) + (81 \times 0)}{1 + 81} \cong 4700 \text{ km}$$

But the radius of the earth is 6370 km. Therefore, the centre of mass of the system (earth + moon) is at an appreciable distance from the

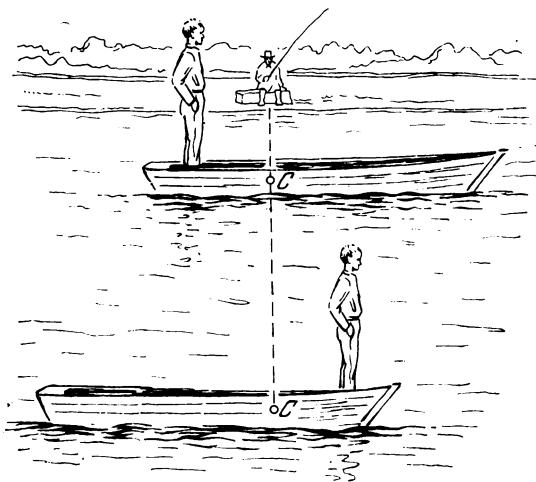


Fig. 15.5

centre of the earth. However, if we carry out the same calculation for the system (sun + earth) and analyze the result, we shall find that the centre of mass of this system coincides almost exactly with the centre of the sun.

Chapter 16

TOTAL AND KINETIC ENERGY

16.1. TOTAL ENERGY OF A BODY

1. The concept of energy is one of the basic concepts of physics. From his elementary physics course the student knows the *law of conservation of energy*, one of the most important laws of nature, and many applications of this law used to explain mechanical, thermal and electrical phenomena. The idea of energy permeates many engineering problems because the most vital of them involve the production, transmission and utilization of various kinds of energy. This and subsequent chapters will discuss the concept of energy in detail and show how it can be used to solve many physics problems.

2. The most general concept of energy can be found in the ideas of relativity theory. The *total energy* of a body isolated from external influence can be defined as the *product of its relativistic mass by the*

square of the velocity of light. Thus

$$\mathcal{E} = mc^2 \quad (16.1)$$

where \mathcal{E} = total energy

$$m = \frac{m_0}{\sqrt{1-u^2/c^2}} = \text{relativistic mass}$$

$$c = 3 \times 10^8 \text{ m/s} = \text{velocity of light.}$$

It follows from this definition that the *total energy of a body differs in different reference frames*. It depends upon the velocity of the body with respect to the reference frame, because its relativistic mass depends upon its velocity (see Sec. 13.1), and the body has different velocities in different reference frames.

A body possesses minimum energy in the reference frame with respect to which it is at rest. This value of the energy of a body is called the *rest energy*:

$$\mathcal{E}_0 = m_0 c^2 \quad (16.2)$$

16.2. KINETIC ENERGY

1. The *kinetic energy of a body is the difference between its total and rest energies.* Thus

$$K = \mathcal{E} - \mathcal{E}_0 = (m - m_0)c^2 \quad (16.3)$$

where K is the kinetic energy. It depends upon the velocity of the body with respect to the reference frame.

We shall now change this expression for the kinetic energy to make it more convenient for further discussion. From equation (13.3) it follows that $m_0/m = \sqrt{1-u^2/c^2}$. Then

$$K = mc^2 \left(1 - \frac{m_0}{m}\right) = mc^2 \left(1 - \sqrt{1-u^2/c^2}\right)$$

After multiplying and dividing this equation by $1 + \sqrt{1-u^2/c^2}$ we obtain

$$K = \frac{mu^2}{1 + \sqrt{1-u^2/c^2}} = \frac{p^2}{m(1 + \sqrt{1-u^2/c^2})} \quad (16.4)$$

In Newtonian mechanics the velocity of a body with respect to a reference frame is very much less than the velocity of light. Hence we can neglect the fraction u^2/c^2 in comparison to unity, assuming that $\sqrt{1-u^2/c^2} \cong 1$. Then we obtain the very simple formula for finding the kinetic energy used in Newtonian mechanics:

$$K = \frac{mu^2}{2} = \frac{p^2}{2m} \quad (16.5)$$

Though this formula is only approximate, its accuracy is quite sufficient for all practical purposes in engineering. Using the results of Sec. 13.3 we find that even at velocities of hundreds of kilometres per second, the approximate expression (16.5) differs from the exact equation (16.4) by less than one ten-thousandth of one per cent.

2. At velocities much lower than the velocity of light, the kinetic energy of a body is much less than the rest energy. As a matter of fact

$$\frac{K}{\mathcal{E}_0} = \frac{u^2}{2c^2} \ll 1$$

Even at a velocity of $u \approx 300$ km/s, which is ten times as much as the orbital velocity of the earth, the kinetic energy comes to

$$\frac{K}{\mathcal{E}_0} = \frac{9 \times 10^{10}}{2 \times 9 \times 10^{16}} = 5 \times 10^{-7} = 5 \times 10^{-5} \%$$

i.e. five hundred-thousandths of one per cent of the rest energy of a body!

At ultrarelativistic velocities, on the contrary, which differ only slightly from the velocity of light, practically all the energy of a body is in its kinetic energy. Here the rest energy is substantially less than the kinetic energy. In up-to-date synchrophasotrons, for example, protons are accelerated to velocities of the order of $u = 0.999923 c$. Then

$$\begin{aligned} \frac{K}{\mathcal{E}_0} &= \frac{(m - m_0)c^2}{m_0 c^2} = \frac{1}{\sqrt{1 - u^2/c^2}} - 1 = \frac{10^3}{\sqrt{153.9941}} - 1 \\ &= 80.6 - 1 \approx 80 \end{aligned}$$

Thus, at these velocities, the kinetic energy of the proton is 80 times its rest energy.

Neglecting the rest energy in comparison with the kinetic energy, equation (16.3) can be written in the following approximate form for ultrarelativistic velocities:

$$K \approx \mathcal{E} = mc^2 \tag{16.6}$$

16.3. ENERGY AND LINEAR MOMENTUM

1. Let us derive the relation between the total energy of a body and its linear momentum; we shall need it further on. This we do by squaring the expression for the relativistic mass, equation (13.3), and carrying out certain simple transformations. Thus

$$m^2 c^2 - m^2 u^2 = m_0^2 c^2$$

or, after multiplying by c^2 ,

$$m^2 c^4 - m^2 u^2 c^2 = m_0^2 c^4 \tag{16.7}$$

Since the linear momentum of a body is $p = mu$, the total energy is $\mathcal{E} = mc^2$ and the rest energy is $\mathcal{E}_0 = m_0c^2$, we can write

$$\mathcal{E}^2 = \mathcal{E}_0^2 + p^2c^2 \quad (16.8)$$

In the ultrarelativistic range

$$K \cong \mathcal{E} \cong pc \quad (16.9)$$

2. An analysis of equation (16.8) indicates that there are two ways of changing the total energy of a body.

In the first place, the momentum of the body can be changed without changing the rest energy. This can be done by changing the velocity of the body with respect to the chosen reference frame. In the second place, the rest energy can be changed. In this case the rest mass of the body will naturally be changed as well. It is not clear, on the face of it, what processes are required to change the rest mass of bodies. Actually, such processes are feasible, as shall be demonstrated in Chapters 20 and 21.

16.4. KINETIC ENERGY AND WORK

1. Let us consider a case in which the force F acting on a body makes an angle α with the direction of motion. In the general case we can assume that the force is variable and that the body moves along a curvilinear path. Next we resolve the force into two components (Fig. 16.1): one tangential $F_t = F\cos\alpha$ and one normal $F_n = F\sin\alpha$, and see how they influence the kinetic energy of the body.

The normal component of the force only changes the direction of the velocity vector, but does not change its magnitude (see Sec. 4.8). Hence, the normal component of the force has no influence on the kinetic energy of the body. As matter of fact, the expression for the kinetic energy, equation (16.4), contains the square of the velocity. Therefore, the kinetic energy depends only on the magnitude of the velocity (or momentum), but not on the direction of these vectors.

The tangential component of the force, on the other hand, affects precisely the magnitude of the velocity (or momentum) but not its direction. Consequently, the change in the kinetic energy of a body is due to the action of the tangential component of the applied force. This component, according to equation (13.5), can be written as

$$F_t = \frac{p_2 - p_1}{t_2 - t_1} = \frac{\Delta p}{\Delta t}$$

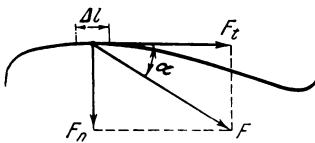


Fig. 16.1

In fact, the tangential component of the force characterizes the change in the magnitude of the momentum per unit time. Therefore, equation (13.5) should refer to the difference of the magnitudes of the vectors and not to that of the vectors themselves.

2. Assume, due to the force, that the body moves a short distance Δl during a short interval Δt of time. At constant rest energy, the change in the kinetic energy of the body equals the change in its total energy. Thus

$$\Delta K = K_2 - K_1 = (\mathcal{E}_2 - \mathcal{E}_0) - (\mathcal{E}_1 - \mathcal{E}_0) = \mathcal{E}_2 - \mathcal{E}_1 = \Delta \mathcal{E}$$

The change in the total energy can be calculated by means of equation (16.8):

$$\mathcal{E}_2^2 = \mathcal{E}_0^2 + p_2^2 c^2 \quad \text{and} \quad \mathcal{E}_1^2 = \mathcal{E}_0^2 + p_1^2 c^2$$

Subtracting the second equation from the first we obtain

$$(\mathcal{E}_2 - \mathcal{E}_1)(\mathcal{E}_2 + \mathcal{E}_1) = (p_2 - p_1)(p_2 + p_1)c^2$$

Putting

$$\mathcal{E}_1 = \mathcal{E} - \Delta \mathcal{E}/2; \quad \mathcal{E}_2 = \mathcal{E} + \Delta \mathcal{E}/2$$

$$p_1 = p - \Delta p/2; \quad p_2 = p + \Delta p/2$$

then

$$\Delta \mathcal{E} \times 2\mathcal{E} = \Delta p \times 2pc^2 \quad \text{or} \quad \Delta \mathcal{E} \times mc^2 = \Delta p \times muc^2$$

Thus

$$\Delta \mathcal{E} = \Delta K = u \Delta p \tag{16.10}$$

After multiplying and dividing equation (16.10) by Δt , we recall that

$$u \Delta t = \Delta l \quad \text{and} \quad \frac{\Delta p}{\Delta t} = F_t = F \cos \alpha$$

At constant rest energy the change in the total energy of a body is equal to the change in its kinetic energy. Then we obtain

$$\Delta K = F_t \Delta l = F \Delta l \cos \alpha \tag{16.11}$$

3. We shall call the physical quantity $F_t \Delta l = F \Delta l \cos \alpha$ an *element of work*. Thus

$$\Delta W = F \Delta l \cos \alpha = F_t \Delta l \tag{16.12}$$

From equations (16.11) and (16.12) it follows that

$$\Delta W = \Delta K \tag{16.13}$$

i.e. an element of work is equal to an infinitesimal change in the kinetic energy of a body.

4. Let us consider the case of the work done by an arbitrary force over a finite length of path (Fig. 16.2). First we shall divide the

path into the small portions $\Delta l_1, \Delta l_2, \dots, \Delta l_n$. According to equation (16.13), for each portion we have

$$\Delta W_1 = \Delta K_1 = K_1 - K_0; \quad \Delta W_2 = \Delta K_2 = K_2 - K_1; \quad \dots; \\ \Delta W_n = \Delta K_n = K - K_n$$

Adding together these equations and denoting the total work done over a finite length of the path by $W = \Delta W_1 + \Delta W_2 + \dots + \Delta W_n$, we obtain

$$W = K - K_0 \quad (16.14)$$

because all the intermediate terms cancel out.

Therefore, the work done by a force over a finite length of path is equal to the change in the kinetic energy of the body to which the force is applied, i.e. to the difference in the values of the kinetic energy of

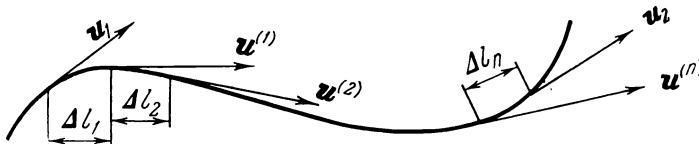


Fig. 16.2

the body at the final and initial points of the path. Equation (16.14) was derived on the basis of the formulas of relativity theory and is valid, therefore, for the motion of bodies at arbitrary velocities (but under the condition that the rest energy does not change).

5. Taking the importance of the result, it will be advisable for the student, as homework, to derive equation (16.14) for the case of the work done by a constant force and with the approximation $u \ll c$ (Newtonian mechanics). He can make use, for example, of equations (4.9), (4.10) and (7.2) to obtain

$$W = \frac{mu^2}{2} - \frac{mu_0^2}{2} \quad (16.15)$$

16.5. POWER

1. The *average power* during the time interval Δt is defined as the work divided by the time interval, or

$$P_{av} = \frac{\Delta W}{\Delta t} \quad (16.16)$$

The *instantaneous power* is the limit that the average power approaches in an infinitesimal time interval:

$$P = \lim_{\Delta t \rightarrow 0} \frac{\Delta W}{\Delta t} \quad (16.17)$$

2. The instantaneous power can be expressed in terms of the force and the instantaneous velocity. For this purpose we substitute equation (16.12) into equation (16.17) and obtain

$$P = \lim_{\Delta t \rightarrow 0} \left(F_t \frac{\Delta l}{\Delta t} \right) = F_t \lim_{\Delta t \rightarrow 0} \frac{\Delta l}{\Delta t} = F_t u$$

where u is the magnitude of the instantaneous velocity. Hence,

$$P = F_t u = F u \cos \alpha \quad (16.18)$$

Note that the work and power equal zero if the force is perpendicular to the path ($\alpha = \pi/2$).

16.6. UNITS OF ENERGY, WORK AND POWER

1. Since work is a measure of the change in energy, work and energy are measured by the same units.

The unit of work in the International System of Units (SI) is the *joule* (J). A joule is the work done by a force of 1 N over a distance of 1 m. Thus

$$1 \text{ J} = 1 \text{ N} \times 1 \text{ m}$$

Work has the dimensions

$$[W] = [F] [l] = ML^2T^{-2}$$

The same dimensions can be obtained in another way:

$$[W] = [\mathcal{E}] = [m] [c^2] = ML^2T^{-2}$$

An idea of the size of the joule can be obtained from the following: if a body with a mass of 2 kg moves with a velocity of 1 m/s, its kinetic energy is equal to 1 joule.

2. In the cgs system, the unit of work is the *erg*:

$$1 \text{ erg} = 1 \text{ dyne} \times 1 \text{ cm}$$

Evidently

$$1 \text{ J} = 10^7 \text{ ergs}$$

In engineering, energy is frequently measured in *kilogram-metres* (kgf-m):

$$1 \text{ kgf-m} = 1 \text{ kgf} \times 1 \text{ m} = 9.81 \text{ J}$$

In thermal calculations, energy is sometimes measured in *calories* (cal). By definition

$$1 \text{ cal} = 4.1868 \text{ J} \approx 4.2 \text{ J} = 0.427 \text{ kgf-m}$$

Units of energy applied in atomic physics are the *electron volt* (eV) and its derived units: *kiloelectron volt* ($1 \text{ keV} = 10^3 \text{ eV}$), *million*

electron volt ($1 \text{ MeV} = 10^6 \text{ eV}$) and the *gigaelectron volt* ($1 \text{ GeV} = 10^9 \text{ eV}$).

$$1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$$

The purpose of calling units of energy by these names will become clear further on (see Sec. 18.7).

3. The SI unit for power is the *watt* (W):

$$1 \text{ W} = 1 \text{ J/s}$$

Other units that find application are:

$$1 \text{ erg/s} = 10^{-7} \text{ W}$$

$$1 \text{ kgf-m/s} = 9.81 \text{ W}$$

$$1 \text{ hp (metric)} = 75 \text{ kgf-m/s} = 735.499 \text{ W} \cong 736 \text{ W}$$

$$1 \text{ kcal/h} = \frac{10^3 \times 4.2 \text{ J}}{3600 \text{ s}} = 1.163 \text{ W}$$

It will be left as an exercise for the student to derive the relation between the units: kilowatt-hour and kilocalorie, metric horsepower and kilocalorie per hour, and the hectowatt-hour and kilogram-metre.

16.7. MOMENTUM AND ENERGY OF A LOCALIZED PARTICLE

1. A particle is said to be *localized* in a certain region of space with the typical dimension a if its coordinates comply with the relationships

$$\begin{aligned} x_0 &\leqslant x \leqslant x_0 + a; \quad y_0 \leqslant y \leqslant y_0 \\ &+ a \text{ and } z_0 \leqslant z \leqslant z_0 + a \end{aligned} \quad (16.19)$$

This means that the particle is within the limits of the cube shown in Fig. 16.3.

Naturally, the uncertainty of the coordinates of the particle coincides with the typical dimension of the region in which the particle is localized. Thus

$$\Delta x \cong \Delta y \cong \Delta z \cong a \quad (16.20)$$

Applying the uncertainty relation, equation (14.5), we shall find the uncertainty of the projections of the linear momenta:

$$\Delta p_x \cong \Delta p_y \cong \Delta p_z \cong \hbar/a \quad (16.21)$$

2. Let us try to estimate the average value of the momentum of this particle. It follows from equation (16.21) that the difference

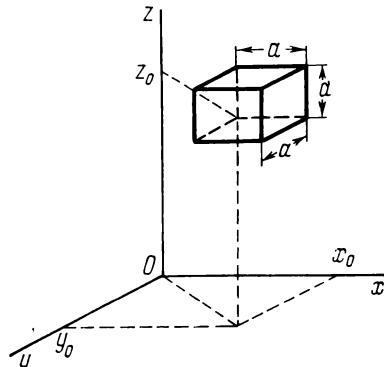


Fig. 16.3

between two arbitrary values of the projection of the momentum on any coordinate axis does not exceed \hbar/a . Thus

$$0 \leq |p_{x2} - p_{x1}| \leq \hbar/a$$

But the sum of the magnitudes of two quantities is not less than the magnitude of their difference:

$$|p_{x1}| + |p_{x2}| \geq |p_{x2} - p_{x1}| \cong \hbar/a$$

The average value of the projection of the momentum on the axis of abscissas is

$$\bar{p}_x \cong 1/2(|p_{x1}| + |p_{x2}|) \geq \hbar/2a$$

Similarly for the other projections:

$$\bar{p}_x \cong \bar{p}_y \cong \bar{p}_z \geq \hbar/2a \quad (16.22)$$

From this we can find the average value of the linear momentum of the particle. Thus

$$\bar{p} = \sqrt{\bar{p}_x^2 + \bar{p}_y^2 + \bar{p}_z^2} \geq \sqrt{3\hbar^2/4a^2} \cong \hbar/a \quad (16.23)$$

where a is the dimension of the region of localization*.

3. Let us derive the expression for the average kinetic energy of a localized particle. According to equations (16.5) and (16.23) we can write for the nonrelativistic case:

$$\bar{K} = \frac{\bar{p}^2}{2m} \geq \frac{\hbar^2}{2ma^2} \quad (16.24)$$

According to equation (16.9), we obtain for an ultrarelativistic particle

$$\bar{K}_{rel} \cong \bar{p}c \geq \hbar c/a \quad (16.25)$$

If the number of particles in unit volume (i.e. their concentration) equals n , then, evidently, $na^3 = 1$, from which it follows that $a = n^{-1/3}$. Substituting into equations (16.24) and (16.25), we express the kinetic energy in terms of the particle concentration:

$$\bar{K} \geq \frac{\hbar^2 n^{2/3}}{2m} \quad (16.26)$$

$$\bar{K}_{rel} \geq \hbar c n^{1/3} \quad (16.27)$$

4. Thus, a particle localized in a certain region of space cannot be at rest. Its average momentum is $\bar{p} \cong \hbar/a$ and its average kinetic energy is $\bar{K} \cong \hbar^2/2ma^2$ (or $\bar{K}_{rel} \cong \hbar c/a$).

On the face of it, this result seems paradoxical; it contradicts our everyday experience in which we observe any amount of bodies at

* As we can see, the average value of the momentum of a localized particle is not less than the uncertainty of the momentum, i.e. $\bar{p} \geq \Delta p$. The same is true for the energy: $\bar{E} \geq \Delta E$.

rest. Two problems will help us to gain an understanding of this seeming paradox.

Problem 1. Assume that a one-kopeck coin ($m = 1 \text{ g} = 10^{-3} \text{ kg}$) is lying on the table. Let us estimate its average momentum and velocity. The region of localization here is, at the least, no less than the size of the coin: $a \cong 1 \text{ cm} = 10^{-2} \text{ m}$. The average value of the momentum is

$$\bar{p} \cong \frac{\hbar}{m} \cong \frac{10^{-34}}{10^{-3}} \cong 10^{-32} \text{ kg}\cdot\text{m/s}$$

and the average velocity is

$$\bar{u} = \frac{\bar{p}}{m} \cong \frac{10^{-32}}{10^{-3}} = 10^{-29} \text{ m/s}$$

Such a velocity can be neglected, of course, and the coin can be assumed to be at rest. We see again, as in Sec. 14.3 that in the domain of Newtonian mechanics we can ignore consequences of the uncertainty relation.

Problem 2. Let us estimate the momentum and energy of an electron in an atom ($a \cong 1 \text{ \AA} = 10^{-10} \text{ m}$). Here

$$\bar{p} \cong \frac{\hbar}{a} \cong \frac{10^{-34}}{10^{-10}} = 10^{-24} \text{ kg}\cdot\text{m/s}$$

$$\bar{u} = \frac{\bar{p}}{m} \cong \frac{\hbar}{ma} \cong \frac{10^{-34}}{9 \times 10^{-31} \times 10^{-10}} = 10^6 \text{ m/s}$$

As we see, the electron moves in the atom with a considerable, though not relativistic velocity. The average value of the kinetic energy is

$$\bar{K} \cong \frac{\hbar^2}{2ma^2} \cong \frac{10^{-68}}{2 \times 9 \times 10^{-31} \times 10^{-20}} \cong 5 \times 10^{-19} \text{ J} \cong 3 \text{ eV}$$

When we have gained more detailed knowledge of the atom's structure (see Chapters 71 and 72) we shall see that we have obtained a correct result as to the order of magnitude.

Therefore, we cannot make use of Newtonian conceptions in the subatomic domain. Here, only quantum ideas, including the conclusions of the uncertainty relation, yield correct results.

Chapter 17

ELEMENTARY COLLISION THEORY

17.1. WHAT IS A COLLISION?

1. In everyday life a collision is understood to be a phenomenon similar to that observed when *billiard balls* strike each other, i.e. when there is direct contact between the colliding bodies. This phe-

nomenon is understood in a wider sense in physics. By a *collision* or *impact* we mean *any short interaction of particles*. Hence, we can speak of the collision of molecules though they interact at a distance through their electric fields. We can likewise speak of the collision of a neutron or alpha particle with a nucleus though here there is also no direct contact of the particles, and the interaction is due to nuclear or electric forces.

2. A feature of the theory of collisions is that we do not analyze the interaction mechanism in detail. The reason for this is that any analysis of the forces exerted in the collision turns out to be extremely difficult and, in many cases, quite impossible. Such is the case, for instance, with nuclear forces whose variation with distance has not yet been found.

This leads to a situation in which we cannot describe the law of motion of particles in detail, we cannot calculate their path, instantaneous velocities, etc. But frequently this is of no special interest to us, the more so because it is simply impossible due to the uncertainty principle to precisely specify the path and instantaneous velocities (or momenta) of microparticles.

3. The problem of analyzing a collision is usually stated as follows: the linear momenta and kinetic energies of the particles before the collision are given; how can these quantities be found after the collision? It turns out that no detailed analysis of the interaction is required to solve this problem.

In the following, the theory of collision is only dealt with in the approximation of Newtonian mechanics, when the velocities of the particles are much less than that of light in a vacuum, and the masses of the particles are sufficiently large. Moreover, for the sake of simplicity, we shall consider the collision of only two particles. Notwithstanding the seeming limitedness of this formulation of the problem, it will be found useful for the analysis of many real physical phenomena. We shall consider certain problems for whose solution it proves sufficient to apply the law of conservation of linear momentum.

17.2. COMPLETELY INELASTIC COLLISION

1. A collision is said to be *completely inelastic* if the bodies move at the same common velocity after collision, sticking together to form a single new body. Examples are the firing of shells into a railway car filled with sand, when the shells remain embedded in the sand, the collision of a meteorite with the earth, etc.

Suppose that a particle of mass m_1 moves with a velocity \mathbf{v}_1 and a particle of mass m_2 with a velocity of \mathbf{v}_2 . The particles collide at a certain point and stick together. What is the velocity of the newly formed body (Fig. 17.1)? Since we are dealing with a nonrelativistic

case, the masses of the bodies do not vary with their velocity, and the mass of the new body is $M = m_1 + m_2$. On the basis of the law of conservation of linear momentum we can write

$$m_1 v_1 + m_2 v_2 = (m_1 + m_2) v \quad (17.1)$$

from which

$$v = \frac{m_1 v_1 + m_2 v_2}{m_1 + m_2} \quad (17.2)$$

2. It should be noted that velocities and momenta are vectors, and, therefore, equation (17.1) cannot be dealt with as an arithmetical addition. Addition is to be performed here according to the rule for adding vectors (Fig. 17.2). In particular, if the bodies are

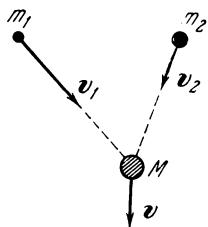


Fig. 17.1

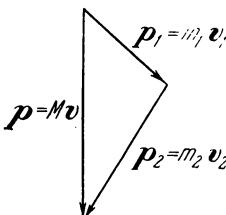


Fig. 17.2

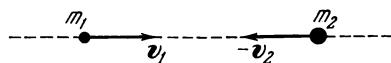


Fig. 17.3

moving toward each other along a straight line, in a head-on collision, one vector should have a plus sign and the other vector a minus sign (Fig. 17.3).

3. In an inelastic collision, the sum of the kinetic energies of the particles before the collision is not equal to the kinetic energy of the body formed after the collision.

This is obvious for the case of the inelastic collision of two bodies whose momenta are equal and opposite in direction. Here, the kinetic energy of the body after the collision equals zero, while the sum of the kinetic energies of the colliding bodies is naturally not equal to zero. The same is true for the case of a body which, with respect to the given reference frame, splits into two fragments.

4. Let us consider, as an example, the problem of the so-called alpha radioactive decay. Here the nucleus of some radioactive element emits an alpha particle (helium nucleus), and the newly formed nucleus is subject to recoil.

Assume that we know the kinetic energy K of the alpha particle and wish to find the kinetic energy of the new nucleus, which we denote by the letter R . This problem is readily solved by the conservation of linear momentum principle.

Let the mass of the alpha particle equal m and the mass of the nucleus before decay equal M . Then the mass of the new nucleus equ-

als $M - m$. Assuming that, before decay, the nucleus was at rest with respect to the laboratory reference frame, we obtain on the basis of the conservation of linear momentum that

$$0 = \mathbf{p}_\alpha + \mathbf{p}_R \quad (17.3)$$

Consequently, the momenta of the alpha particle and the new nucleus are equal in magnitude and opposite in sign.

Let the alpha particle move with a nonrelativistic velocity; the velocity of the new nucleus will be even less. We can find their kinetic energies according to equation (16.5). Thus

$$K = \frac{p_\alpha^2}{2m} \quad \text{and} \quad R = \frac{p_R^2}{2(M-m)} \quad (17.4)$$

The ratio of the kinetic energies is

$$\frac{K}{R} = \frac{M-m}{m} \quad (17.5)$$

The change in the energy of the initial nucleus is

$$\Delta E = K + R = K \frac{M}{M-m} \quad (17.6)$$

Let us consider a concrete numerical example. Investigations show that in the decay of a radium nucleus an alpha particle is emitted with the kinetic energy $K = 5.681$ MeV, and that the mass of a radium nucleus is 56 times greater than the mass of an alpha particle ($M = 56m$). Then the kinetic energy of the nucleus formed as a result of radioactive decay is

$$R = \frac{Km}{M-m} = \frac{5.681 \times 1}{55} = 0.103 \text{ MeV}$$

and the total energy of the reaction is

$$\Delta E = K + R = 5.784 \text{ MeV}$$

17.3. ELASTIC COLLISION

1. A collision is said to be *elastic* if, in addition to the sum of the linear momenta, the sum of the kinetic energies of the colliding bodies is also conserved. This by no means implies that the energy of each of the bodies is not changed. On the contrary, the velocities of the bodies are changed by the collision and, at the same time, their momenta and kinetic energies. But the *sum of the momenta* and the *sum of the kinetic energies* remains constant in an elastic collision.

The following system of equations is obtained in an elastic collision of two bodies:

$$\left. \begin{aligned} p_1 + p_2 &= p'_1 + p'_2 \\ K_1 + K_2 &= K'_1 + K'_2 \end{aligned} \right\} \quad (17.7)$$

where the unprimed letters refer to the momenta and kinetic energies before the collision; the primed letters, to these quantities after the collision.

2. Let us consider an elastic *central collision* of two balls, i.e. a collision in which the velocity vectors are along a straight line joining the centres of the two balls. To simplify the discussion we

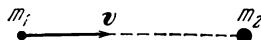


Fig. 17.4

shall deal with this phenomenon in a reference frame with respect to which the second ball m_2 is at rest (Fig. 17.4) and ball m_1 moves with a velocity of v . Then equations (17.7) will take the form

$$\left. \begin{aligned} m_1 v &= m_1 v_1 + m_2 v_2 \\ \frac{m_1 v^2}{2} &= \frac{m_1 v_1^2}{2} + \frac{m_2 v_2^2}{} \end{aligned} \right\} \quad (17.8)$$

or the form

$$m_1 (v - v_1) = m_2 v_2$$

$$m_1 (v^2 - v_1^2) = m_2 v_2^2$$

Dividing the second equation by the first we obtain

$$v + v_1 = v_2 \quad (17.9)$$

Substituting equation (17.9) into the first equation of system (17.8) we obtain

$$m_1 v = m_1 v_1 + m_2 v + m_2 v_1$$

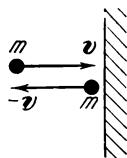
from which

$$v_1 = \frac{m_1 - m_2}{m_1 + m_2} v \quad \text{and} \quad v_2 = \frac{2m_1}{m_1 + m_2} v \quad (17.10)$$

3. Now let us dwell on several interesting special cases.

(a) If $m_2 = m_1$, then $v_1 = 0$ and $v_2 = v$. Therefore, in an elastic central collision of a moving ball with a ball of equal mass which is at rest, the first ball stops and the second ball starts moving at the same velocity the first ball had before the collision. If, in addition, the balls are indistinguishable (for instance, if they are molecules), then the first ball will seem to have "passed through" the second ball, continuing to move at the same velocity, while the stationary ball simply remains in the same position.

(b) If $m_1 > m_2$, then $0 < v_1 < v$ and $v_2 > v$. In the limiting case, when $m_1 \gg m_2$, the velocity of the massive ball is almost unchanged by the collision, and the light ball, at rest before the collision, will fly off with a velocity $v_2 \cong 2v$.



(c) If $m_1 < m_2$, then $-v \ll v_1 < 0$, i.e. the light ball, striking the heavy one, bounces back in the reverse direction. At this, the massive ball m_2 , acquires the velocity $0 \leq v_2 < v$.

In the limiting case, when $m_2 \gg m_1$, it can be assumed that the ratio m_1/m_2 approaches zero. Then from equation (17.10) it follows that $v_2 = 0$, and the velocity of the first ball will be

$$v_1 = \frac{m_1 - m_2}{m_1 + m_2} v = \frac{(m_1/m_2) - 1}{(m_1/m_2) + 1} v = -v$$

Hence, if a ball collides with a massive wall, it will bounce back with the same velocity (Fig. 17.5). The velocity of the wall remains practically constant.

17.4. NEUTRON MODERATION

1. It is necessary, for certain nuclear reactions, to slow down the neutrons, i.e. to reduce their kinetic energy from several MeV to hundredths of an electron volt. This slowing-down, or *moderation*, as it is called, is accomplished by elastic collisions of the neutrons with the nuclei of the moderator substance. Let us find the *moderating ratio r*, which is the ratio of the energy lost by the neutron in one collision to its initial kinetic energy.

Let the mass of a nucleus of the moderator be M , and its velocity after collision be $u = \frac{2m}{m+M} v$ [see equation (17.10)]. The energy lost by the neutron is equal to that acquired by the moderator nucleus. Thus

$$r = \frac{\Delta K}{K} = \frac{Mu^2 \times 2}{2 \times mv^2} = \frac{4mM}{(m+M)^2} \quad (17.11)$$

Carbon is frequently used as a moderator. The mass of its nucleus is 12 times that of a neutron. For $M = 12 m$, we obtain from equation (17.11)

$$r = \frac{4 \times 1 \times 12}{13^2} = 0.284 = 28.4\%$$

2. We shall carry out the following calculation to find the number of collisions required to completely slow down a neutron. If the initial kinetic energy of the neutron is K then after one collision

$$K_1 = K - \Delta K = K(1 - r)$$

and after the second collision

$$K_2 = K_1 - \Delta K_1 = K_1 (1 - r) = K (1 - r)^2$$

Evidently, after n collisions, the kinetic energy will be

$$K_n = K (1 - r)^n$$

Putting $K = 1.75$ MeV = 1.75×10^6 eV, $K_n = 0.025$ eV and $r = 0.284$, we obtain: $0.025 = 1.75 \times 10^6 \times 0.716^n$ or $7 \times 10^7 = 0.716^{-n}$, from which

$$n = \frac{-\log(7 \times 10^7)}{\log 0.716} = \frac{-7.845}{-0.145} = 54$$

Hence, if all the elastic collisions of the neutron with nuclei were central ones, in 54 collisions its kinetic energy would be reduced from 1.75 MeV to 0.025 eV, i.e. to one seventy millionth of its initial value! However, off-centre collisions occur as well as central ones, and in them the neutron loses less energy. Therefore, to obtain the corresponding moderation, the number of collisions will have to be increased somewhat.

3. A more efficient moderator is *heavy water* in which the atoms of hydrogen are replaced by atoms of heavy hydrogen—deuterium. Here $M = 2 m$ and $r = \frac{4mM}{(m+M)^2} = \frac{8}{9}$. Substituting into equation (17.11) we obtain $K_n = K/9^n$. Putting, as above, $K = 1.75$ MeV and $K_n = 0.025$ eV, we obtain $n = 7$.

17.5. PRESSURE OF A STREAM OF PARTICLES ON A WALL

1. Let us consider a stream of particles of mass m which all move with a velocity v perpendicular to a wall whose mass is substantially greater than that of all the particles (Fig. 17.6). We can see from the drawing that during the time Δt , all the particles contained in the volume $V = A\Delta l = Av\Delta t$ hit the wall. If the particle concentration (number of particles in unit volume) is equal to n , then, during the time Δt , N particles hit the wall, where

$$N = nV = nAv\Delta t \quad (17.12)$$

2. The force acting on the particle as it hits the wall is equal to the change in linear momentum in unit time. Thus

$$f = \Delta(mv)/\Delta t \quad (17.13)$$

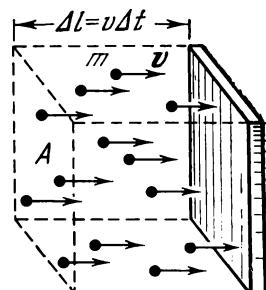


Fig. 17.6

According to Newton's third law, the particle exerts a force on the wall, which is equal in magnitude but opposite in direction. Since N

particles hit the wall during the time Δt , the total force on the wall is

$$F = -fN = -nAv \Delta(mv) \quad (17.14)$$

3. The change in the momentum of a particle depends upon the kind of collision it has with the wall.

In an elastic collision of a particle with a stationary wall, its velocity is reversed (see Sec. 17.3).

Thus the initial velocity $v_i = v$ and the final velocity $v_f = -v$. Then the change in momentum is

$$\Delta(mv)_{el} = mv_f - mv_i = -2mv \quad (17.15)$$

Substituting into equation (17.14) we obtain

$$F_{el} = 2Anmv^2 \quad (17.16)$$

By definition, the *pressure* p is the ratio of the force F to the area A on which this force acts (the direction of the force being perpendicular to the area). Thus

$$p = F/A$$

It follows that the pressure exerted by a stream of particles that strike a wall with elastic collisions can be expressed by

$$p_{el} = 2nmv^2 \quad (17.17)$$

4. In inelastic collisions, when the particles stick to the wall, the pressure is only one half as much. Here the initial velocity $v_i = v$ and the final velocity $v_f = 0$. Then the change in momentum is

$$\Delta(mv)_{inel} = mv_f - mv_i = -mv \quad (17.18)$$

Substituting into equation (17.14) we obtain

$$F_{inel} = Anmv^2$$

$$p_{inel} = nmv^2 \quad (17.19)$$

5. The formulas obtained for the forces exerted by a stream on a wall are very general. They are suitable, for example, for calculating the pressure of a gas (see Sec. 26.1), for calculating the pressure of light, etc.

Let us apply them for calculating the force exerted by a stream of liquid on a wall. The product $\rho = mn$ is the density of the liquid. In fact, if we multiply the concentration—the number of molecules in unit volume—by the mass of one molecule, we obtain the mass of the substance in unit volume, i.e. the density of the substance.

Thus

$$F_{el} = 2A\rho v^2 \quad (17.20)$$

$$F_{inel} = A\rho v^2 \quad (17.21)$$

In essence, we have already applied similar formulas in Chapter 11 to estimate the resistance to which a body moving in a liquid is subjected [see equation (11.10)]. There the equation was derived from an analysis of the dimensionality. In this chapter, the formula was derived on the basis of the law of conservation of linear momentum.

Chapter 18

CONSERVATIVE FORCES AND POTENTIAL ENERGY

18.1. WORK DONE BY A VARIABLE FORCE

1. As mentioned above (Sec. 16.4), an element of work done by the force \mathbf{F} over an infinitesimal displacement Δl is the product of the magnitude of the force by the magnitude of the displacement by the cosine of the angle between these vectors. Thus

$$\Delta W = F \Delta l \cos \alpha$$

In other words, an *element of work is equal to the product of the tangential component of the force by the magnitude of the displacement:*

$$\Delta W = F_t \Delta l$$

Depending on whether angle α is acute or obtuse, the element of work ΔW can be either positive (if $0 < \alpha < \pi/2$, then $\Delta W > 0$) or negative (if $\pi/2 < \alpha < \pi$, then $\Delta W < 0$).

2. To calculate the work done on a finite length of path, we divide the path into small displacements and calculate the element of work done over each displacement. Then the sum of these elements of work will be the work done over a finite length of path. Thus

$$W = \Delta W_1 + \Delta W_2 + \dots + \Delta W_n = F_t^{(1)} \Delta l_1 + F_t^{(2)} \Delta l_2 + \dots + F_t^{(n)} \Delta l_n \quad (18.1)$$

Strictly speaking, a certain error is always introduced here that depends upon how the path is divided into small portions. An exact answer is obtained only *in the limit*, i.e. when the path is divided into an infinitely great number of infinitesimal displacements. Thus

$$W = \lim_{\Delta l \rightarrow 0} (F_t^{(1)} \Delta l_1 + F_t^{(2)} \Delta l_2 + \dots + F_t^{(n)} \Delta l_n) \quad (18.2)$$

3. The work done by a constant force over a straight portion of the path can be written as

$$W = F_t l = Fl \cos \alpha \quad (18.3)$$

which is obtained by factoring out F_t in equation (18.1) and adding the elements of displacement Δl , which give the path l traversed by the body on which the constant force acts.

This work can be represented by the graph in Fig. 18.1. The graph of the tangential component of the force when $F_t = \text{const}$ is a straight line parallel to the axis of abscissas. Then the work done over the portion $l = l_2 - l_1$ of the path is *numerically* equal to the area

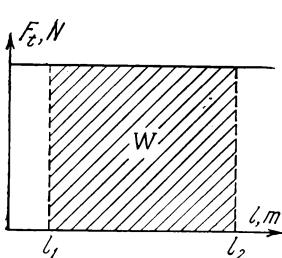


Fig. 18.1

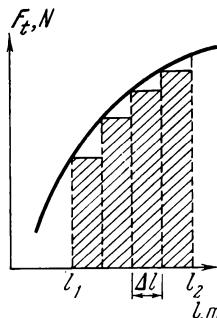


Fig. 18.2

of the hatched rectangle. Here the scale must be taken into account: unit area represents unit work (in our figure, 1 cm^2 represents 1 J).

4. In the general case, the tangential component of the work is a variable quantity (Fig. 18.2). To calculate the work graphically, we divide the path l into small portions. Then we multiply the lower value of the force for each portion (the left-hand ordinates in our case) by the corresponding displacement and add the products thus obtained together. This gives us the area of the internal stepped figure which represents a certain work W_{\min} . Since we used the minimum value of the force for each displacement, the amount of work W_{\min} is less than the required work W .

If we now repeat this calculation but multiply the higher values of the force (the right-hand ordinates in Fig. 18.3), the area of the outward-projecting stepped figure represents a certain amount of work W_{\max} , which is greater than the required work W . Thus

$$W_{\min} < W < W_{\max}$$

If we divide the path into more portions, let us say, twice as many, the area of the internal stepped figure is increased and that of the external figure is reduced (Fig. 18.4). It is shown in courses of higher mathematics that if the number of portions goes to infinity (and, correspondingly, $\Delta l \rightarrow 0$), the areas of the two stepped figures have a common limit, equal to the area under the curve, which exactly represents the work done by the variable force.

Therefore, the work done by a variable force is represented on a graph by the area under the curve of the force. This area is bounded underneath

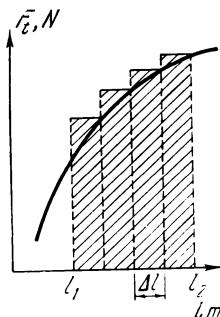


Fig. 18.3

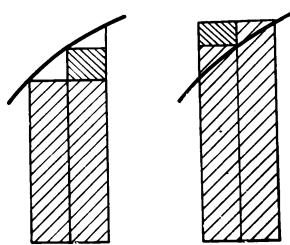


Fig. 18.4

by the axis of abscissas and at the sides by the ordinates of the end points of the curve.

18.2. WORK DONE BY AN ELASTIC FORCE

1. First we shall calculate the work done by the external force to stretch a spring. According to Newton's third law, the external force is equal in magnitude to the elastic force but is opposite in direction ($F_{ext} = -F_{el}$). Then, from the equation for the elastic force, we can write

$$F_{ext} = kx \quad (18.4)$$

where k is the coefficient, or modulus, of elasticity. A graph of this force is shown in Fig. 18.5. The work done by the external force over the portion of the path $l = x_2 - x_1$ is numerically equal to the area under the curve, i.e. the area of the hatched trapezoid:

$$W_{ext} = \frac{F_1 + F_2}{2} l = \frac{kx_1 + kx_2}{2} (x_2 - x_1) = \frac{kx_2^2}{2} - \frac{kx_1^2}{2} \quad (18.5)$$

2. The work done by the elastic force over the same portion of the path differs only in sign. Hence

$$W_{el} = \frac{kx_1^2}{2} - \frac{kx_2^2}{2} \quad (18.6)$$

3. When $x_1 < x_2$, i.e. in stretching the spring, the elastic force does negative work, which complies with the sign rule for forces: forces of attraction are considered negative and forces of repulsion, positive (see Sec. 10.2).

In fact, when the distance is increased between attracting bodies, the forces of attraction make an obtuse angle with the direction of

the displacement ($\pi/2 < \alpha \leq \pi$) and the cosine of an obtuse angle is a negative number. Here the force of attraction performs negative

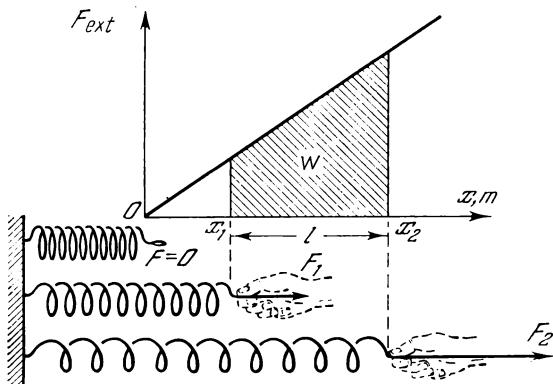


Fig. 18.5

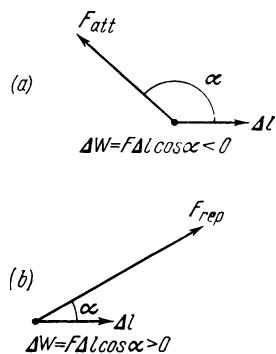


Fig. 18.6

work (Fig. 18.6a). But a force of repulsion makes an acute angle with the direction of the displacement ($0 \leq \alpha < \pi/2$); hence it does positive work (Fig. 18.6b).

18.3. WORK DONE BY A COULOMB FORCE

1. We encountered no particular difficulties in deriving the equation for the work done by an elastic force because this work is represented by the area of an ordinary trapezoid. It is somewhat more difficult to calculate the work done by a Coulomb (or a gravitational) force.

Difficulties arise from the fact that the force is inversely proportional to the square of the distance, and the graph of such a force is a curved line (Fig. 18.7). Since we cannot as yet find the area under such a curve, we shall use numerical methods to calculate the work done by such forces.

2. Let us find the work done by a Coulomb force in moving a charge along a radius from $r_1 = 2 \text{ m}$ to $r_2 = 3 \text{ m}$.

To simplify the calculations assume that $q_1 q_2 / 4\pi\epsilon_0 = 1$. Then the equation for the force will be

$$F = 1/r^2$$

$$(18.7)$$

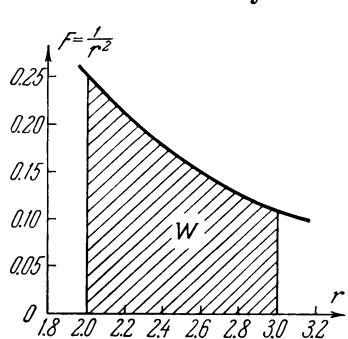


Fig. 18.7

Let us divide the path $l = r_2 - r_1$ into ten equal portions. We shall calculate each element of work as the product of the mean force by the displacement Δr . Thus

$$\Delta W = F_m \Delta r \quad (18.8)$$

the mean force being one half of the sum of the forces at the beginning and end of each displacement:

$$F_m = 1/2 (F_b + F_e)$$

The results of these calculations are given in Table 18.1.

Table 18.1

n	r	$F = \frac{1}{r^2}$	$F_m = \frac{F_n + F_{n+1}}{2}$	$\Delta W = F_m \Delta r$	$W_n = \Delta W_1 + \Delta W_2 + \dots + \Delta W_n$	$\frac{1}{r}$	$\frac{1}{r_1} - \frac{1}{r_n}$
1	2.000	0.2500		0.0238	0.0238	0.5000	0.0238
2	2.100	0.2268	0.2384	0.0217	0.0455	0.4762	0.0455
3	2.200	0.2066	0.2167	0.0198	0.0653	0.4545	0.0652
4	2.300	0.1890	0.1978	0.0181	0.0834	0.4348	0.0833
5	2.400	0.1736	0.1813	0.0167	0.1001	0.4167	0.1000
6	2.500	0.1600	0.1668	0.0154	0.1155	0.4000	0.1154
7	2.600	0.1479	0.1539	0.0142	0.1297	0.3846	0.1296
8	2.700	0.1372	0.1425	0.0132	0.1429	0.3704	0.1429
9	2.800	0.1276	0.1324	0.0123	0.1552	0.3571	0.1552
10	2.900	0.1189	0.1232	0.0115	0.1667	0.3448	0.1667
11	3.000	0.1111	0.1150	0.0108	0.1775	0.3333	0.1774
12	3.100	0.1041	0.1076			0.3226	

An analysis of this table indicates that to a good degree of accuracy (four significant digits) the work done by the force [equation (18.7)] can be expressed as

$$W = \frac{1}{r_1} - \frac{1}{r_2} \quad (18.9)$$

3. If our calculations were not for a special case when $q_1 q_2 / 4\pi\epsilon_0 = 1$, but for the general case of arbitrary charges, we would obtain for the work done by a Coulomb force the equation

$$W_{\text{Coul}} = \frac{q_1 q_2}{4\pi\epsilon_0 r_1} - \frac{q_1 q_2}{4\pi\epsilon_0 r_2} \quad (18.10)$$

We leave it as a problem for the student to show that, in accordance with the proposition derived at the end of the preceding section, the forces of repulsion do positive work when the distance between the charges is increased, and the forces of attraction do negative work.

4. Let us calculate the work done by a Coulomb force when a charge is moved along an arbitrary curvilinear path (Fig. 18.8).

We divide this curve into infinitesimal displacements Δl which, as a first approximation, we consider to be infinitesimal lengths. An element of work done over such a length equals

$$\Delta W = F \Delta l \cos \alpha = F \Delta r \quad (18.11)$$

Fig. 18.8

because, as is evident in the drawing, $\Delta l \cos \alpha = \Delta r$. This equation agrees with equation (18.8) in which we dealt with the displacement of a charge along a radius. The sum of the elements of work equals

$$W = F_1 \Delta l_1 \cos \alpha_1 + F_2 \Delta l_2 \cos \alpha_2 + \dots = F_1 \Delta r_1 + F_2 \Delta r_2 + \dots$$

If we use the mean values of the forces and carry out a numerical calculation we shall obtain equation (18.10) again.

Therefore, the *work done by a Coulomb force does not depend upon the shape of the path of the charge*; it is determined only by the radius vectors of the initial and final points of the path.

18.4. WORK DONE BY A GRAVITATIONAL FORCE

1. The calculation of the work done by a gravitational force is very much the same as that for the work of a Coulomb force. It is only necessary in equation (18.10) to replace the term $q_1 q_2 / 4\pi\epsilon_0$ by $Gm_1 m_2$ and to change the sign because a gravitational force is always a force of attraction. We shall obtain

$$W_{grav} = -\frac{Gm_1 m_2}{r_1} + \frac{Gm_1 m_2}{r_2} \quad (18.12)$$

2. Let a body of mass m be raised to a certain altitude h above the earth's surface (Fig. 18.9). Putting $m_1 = m$ (the mass of the body), $m_2 = M$ (mass of the earth), $r_1 = R$ (radius of the earth) and $r_2 = R + h$, we obtain the expression for the work done by the force of gravity:

$$W = -\frac{GmM}{R} + \frac{GmM}{R+h} = -\frac{GmMh}{R(R+h)} \quad (18.13)$$

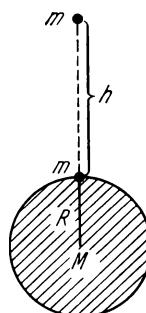


Fig. 18.9

The work done by the external force is opposite in sign. Hence,

$$W_{ext} = \frac{GmMh}{R(R+h)} \quad (18.14)$$

If the altitude h is much less than the earth's radius, it can be assumed with some approximation that $R + h \cong R$. Recalling equation (9.11), we can write

$$W_{ext} = GmMh/R^2 = mgh \quad (18.15)$$

As a matter of fact, we could have reached the same result in an entirely elementary way. At low altitudes above the earth the force of gravity $P = mg$ is a constant, and for a constant force $W_{ext} = Ph = mgh$.

Though the derivation given above is more complicated, it enables us to assess the error we introduce when we substitute approximate equation (18.15) for the exact equation (18.14). Thus, assuming that we are to calculate the work done to an accuracy of one per cent, we can write

$$\frac{W_{appr} - W_{exact}}{W_{exact}} = \frac{W_{appr}}{W_{exact}} - 1 \leqslant 0.01$$

Substituting the exact and approximate values of the work and cancelling like factors, we obtain

$$\frac{R+h}{R} - 1 = \frac{h}{R} \leqslant 0.01$$

Therefore, up to an altitude $h = 0.01 R = 0.01 \times 6371 \cong 64$ km we can, with an accuracy within 1 per cent, use approximate equation (18.15).

3. It is quite obvious, then, that the *work done by a gravitational force*, like that done by a Coulomb force, *does not depend on the shape of the path* of the particle to which the force is applied. The amount of work is determined only by the radius vectors of the initial and final points of the path followed. This follows directly from the identical equations obtained for calculating the forces done by gravitational and Coulomb forces.

18.5. CONSERVATIVE FORCES

1. It was shown in the preceding sections that forces exist—elastic, Coulomb and gravitational—for which the work done does not depend on the shape of the path of the particle that the force is applied to, but only on the coordinates of the beginning and end of the path. Such forces are said to be *conservative*. We shall now show that the *work done by a conservative force on a particle moving along any closed path equals zero for the round trip*.

For this proof let us consider a certain closed path (Fig. 18.10). Let us denote the work done on the particle along the branches

M_1KM_2 and M_1LM_2 by W_K and W_L respectively. It follows from the definition of a conservative force that if a body moves from point M_1 to M_2 , the work done by the force applied to it does not depend on whether it moves along curve M_1KM_2 or curve M_1LM_2 . Consequently, $W_K = W_L$. Assume that positive work is done when the body moves from point M_1 to point M_2 , then negative work is done as it moves from point M_2 to point M_1 : $W'_L = -W_L$. It is a fact that

if the direction of the force is not changed, but the direction of motion is reversed, then, according to the definition, the sign of the work is reversed.

The work done by a force on a particle during a complete cycle (in a closed path) is equal to the sum of the works done on the branches M_1KM_2 and M_2LM_1 . Thus

$$W_{cc} = W_K + W'_L = W_K - W_L = 0$$

We leave it to the student to carry out a calculation for a concrete force acting on a particle that moves along a given path. He should make use of the formulas for either an elastic, Coulomb or gravitational force, equation (18.6), (18.10) or (18.12).

2. We should note that there are also *nonconservative forces* (sometimes called *dissipative forces*).

The work done by such a force essentially depends upon the shape of the path of the particle to which the force is applied. An example is the force of friction. Thus if we push some object, for instance, along the floor, we do work on it in overcoming the friction force. But here, as we can see from experience, the amount of work done depends essentially on the path along which the body is pushed. If, in particular, the friction force is the same at all points of the floor surface, work done along the shortest path between two points is less than for any other path.

The work done by a nonconservative force during a complete cycle (in a closed path) is not equal to zero, i.e. the work done by the friction force (or any other forces of resistance) is always negative because the force is directed opposite the displacement. Naturally, the work done by this force is not equal to zero whatever the path of the body.

18.6. POTENTIAL ENERGY OF ELASTIC, COULOMB AND GRAVITATIONAL INTERACTIONS

1. Now we shall introduce the concept of the *potential energy* U of interacting bodies. This is the name given to energy that depends upon the mutual position of these bodies. It is also known as *mutual energy*, or *energy of configuration*.

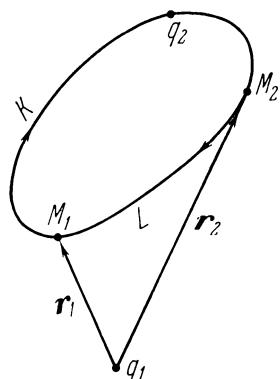


Fig. 18.10

In all the examples dealt with above [equations (18.6), (18.10) and (18.12)], the work done by the conservative forces was equal to the difference between two terms which were functions of the coordinates at the beginning and end of the path. Consequently, the work was equal to the change in the potential energy of the system of interacting bodies. Thus

$$W = U_1 - U_2 \quad (18.16)$$

2. Comparing equations (18.6) and (18.16) we obtain the following expression for the *potential energy of elastic forces*:

$$U_{el} = kx^2/2 \quad (18.17)$$

Evidently, the potential energy of the body equals zero at $x = 0$, i.e. when the spring is unstretched.

3. A similar comparison of equations (18.10) and (18.16) yields the expression for the *potential energy of Coulomb forces*:

$$U_{Coul} = \frac{q_1 q_2}{4\pi\epsilon_0 r} \quad (18.18)$$

We should note here that the potential energy tends to zero when the distance between the interacting charges increases without limit ($r \rightarrow \infty$). If charges q_1 and q_2 are like charges, there is a repulsive force between them and the potential energy is positive. If they are of opposite sign, the force between them is one of attraction and the potential energy is negative.

4. The *potential energy of gravitational forces* can be found by comparing equations (18.12) and (18.16):

$$U_{grav} = -\frac{Gm_1 m_2}{r} \quad (18.19)$$

Here the potential energy is negative because a gravitational force is always one of attraction. When the distance between the interacting bodies increases without limit ($r \rightarrow \infty$), the potential energy tends to zero, i.e. it grows at the expense of the work done by the external force, which increases the distance between the interacting bodies.

5. Of definite interest is the potential energy of the gravity force near the earth's surface. If in equation (18.16) we put the potential energy at the earth's surface $U_1 = 0$, denote the energy U_2 at the altitude h by U and substitute for the work done by the gravity force, the expression from equation (18.13), we can write

$$U = \frac{GmMh}{R(R+h)} \quad (18.20)$$

For low altitudes ($h \ll R$) we can use the approximate equation

$$U = mgh \quad (18.21)$$

The last two expressions are often called the *potential energy of a body in a gravitational field*. This terminology is not quite exact: actually, the potential energy is possessed by a system consisting of the body and the earth, and not by the body alone. However, if this is not forgotten in laying emphasis on the existence of a gravitational field (which is set up by the earth), there is no harm in such usage. One can speak in a similar way of the potential energy of a charge in an electric field. But it is quite impermissible to speak only of the potential energy of a body without specifying the field in which it is located.

6. We should note that we can find the potential energy only to an accuracy within some certain arbitrary constant addend. As a matter of fact if we add a certain constant to all the quantities of potential energy, the new value U' of the potential energy will be

$$U' = U + \text{const}$$

but the amount of work done will not be changed. Thus

$$W' = U'_1 - U'_2 = (U_1 + \text{const}) - (U_2 + \text{const}) = U_1 - U_2 = W$$

This result is frequently stated as follows: the *zero level of potential energy can be chosen arbitrarily*. If we transfer to some other zero level, the potential energy is changed at all points by the same addend; the work done by the force is not changed by this transformation.

It was precisely with this in mind that we dealt in detail with the zero level of potential energy of elastic, Coulomb and gravitational interactions.

18.7. POTENTIAL OF AN ELECTROSTATIC FIELD

1. In Chapter 10 we described an electric field by a *force characteristic* called the vector of field strength, or intensity, \mathbf{E} . But we can introduce an *energy characteristic* of the field as well. It is the scalar quantity φ , called the *electric potential*.

The electric potential of a point in a field is equal in magnitude to the ratio of the potential energy of a test charge, placed at the point of the field, to the value of the charge. Thus

$$\varphi = U/q \quad (18.22)$$

If the test charge is moved from a point in the field with the potential φ_1 to a point with the potential φ_2 , the field does the work

$$W = U_1 - U_2 = q(\varphi_1 - \varphi_2) \quad (18.23)$$

Thus, the *work done by the field in moving a test charge from one point of the field to another is equal to the product of the charge by the difference in potentials* (or simply *potential difference*).

2. Electric forces are conservative. The work they do is *path-independent*, i.e. it does not depend upon the shape of the path along which the charge is moved, but only on the electric potentials of the points at the beginning and end of the path. It follows that a potential electric field is incapable of moving electric charges along closed paths.

3. The SI unit of electric potential is the *volt* (V). A volt is equal to the potential of a point in a field where a test charge of one coulomb has a potential energy of one joule. Naturally, if the potential difference between two points of a field is equal to 1 V, the electric forces moving a test charge of 1 C between these points does work equal to 1 J. Thus

$$1 \text{ V} = \frac{1 \text{ J}}{1 \text{ C}}$$

Now we can give the grounds for the name of the energy unit, the *electron volt*, introduced in Sec. 16.6. An electron volt is the energy acquired by a particle carrying the elementary charge (i.e. the charge of the electron) when it moves between points of a field with a potential difference of 1 V. Since $e = 1.6 \times 10^{-19}$ C and $\varphi_1 - \varphi_2 = 1$ V, then

$$1 \text{ eV} = 1.6 \times 10^{-19} \text{ C} \times 1 \text{ V} = 1.6 \times 10^{-19} \text{ J}$$

We leave it as an exercise for the student to show that the unit of potential in the cgs electrostatic system of units is equal to 300 V.

18.8. ELECTRIC POTENTIAL OF THE FIELD SET UP BY A POINT CHARGE

1. If the point charge setting up a field is denoted by Q and the test charge by q , then the potential energy of the test charge located at a distance r from the source of the field can, in accordance with equation (18.18), be written as

$$U = \frac{qQ}{4\pi\epsilon_0 r} \quad (18.24)$$

From this it follows that the electric potential of the field set up by a point charge equals

$$\varphi = \frac{Q}{4\pi\epsilon_0 r} \quad (18.25)$$

This expression indicates that the potential of the field of a point charge becomes zero at an infinitely distant point. Such a choice of the zero level of the potential may be convenient but is by no means obligatory. In principle, a zero potential can be chosen at any point

of the field (see Sec. 18.6, item 6). This will only change the potentials of all the points of the field by the same constant addend. But the work done by the electric forces is not changed because it depends on the potential difference, whose magnitude is unchanged by this choice, and not on the absolute values of the potentials.

2. Similar to the potential of the field set up by an electric charge is the concept of the potential of a gravitational field, which we can introduce. The gravitational potential of the field set up by a particle of mass M at a distance r from the particle is equal, according to equation (18.19), to

$$\varphi_{grav} = -GM/r \quad (18.26)$$

This potential becomes zero at an infinitely distant point. If we assume that the gravitational potential equals zero at the earth's surface, then, according to equation (18.20), the potential of the gravitational field at the altitude h above the earth's surface equals

$$\varphi_{grav} = \frac{GMh}{R(R+h)} = gh \frac{R}{R+h} \quad (18.27)$$

At low altitudes ($h \ll R$), the approximate value is

$$\varphi_{grav} = gh \quad (18.28)$$

18.9. ENERGY OF AN ELECTRIC FIELD

1. As was shown in Sec. 18.6, the potential energy of two charges, q_1 and q_2 , equals

$$U = \frac{q_1 q_2}{4\pi\epsilon_0 r}$$

But the expression $q_1/4\pi\epsilon_0 r = \varphi_1$ is the electric potential of the field set up by charge q_1 at the point where charge q_2 is located. Therefore, the equation for the potential energy of two interacting charges can be written as

$$U = q_1\varphi_2 = q_2\varphi_1 \quad (18.29)$$

The half-sum of two equal quantities is equal to each of them; hence

$$U = 1/2 (q_1\varphi_2 + q_2\varphi_1) \quad (18.30)$$

The potential energy of a set of many charges is calculated in a similar way: it is the half-sum of the product of each charge by the potentials of the fields set up by all the other charges. Thus, q_1 is multiplied by the sum of all the potentials except φ_1 , q_2 by the sum of all the potentials except φ_2 , etc. The reason for this is that the field set up by a certain fixed charge acts on any other charges except this one.

Chapter 19

LAW OF CONSERVATION OF ENERGY IN NEWTONIAN MECHANICS

19.1. MECHANICAL ENERGY AND ITS CONSERVATION

1. Let us consider a system of bodies subject to only conservative forces. *The physical quantity E , equal to the sum of the kinetic and potential energies, is called the total mechanical energy of the system.* Thus

$$E = K + U \quad (19.1)$$

In Newtonian mechanics these quantities are of vital importance.

2. Assume that the conservative system in question is a closed one. This means that it is subject to only internal forces. Equation (16.14) infers that the work done by the internal forces equals the change in kinetic energy, and equation (18.16) that it is equal to the change in potential energy. Thus

$$W = K_2 - K_1 \text{ and } W = U_1 - U_2$$

Equating the right-hand sides of these two equations we write

$$K_2 - K_1 = U_1 - U_2$$

or

$$K_2 + U_2 = K_1 + U_1 \quad (19.2)$$

Consequently, the sum of the kinetic and potential energies, i.e. the total mechanical energy of the system is conserved.

$$E = K + U = \text{const} \quad (19.3)$$

We thus obtain the *law of conservation of energy* in Newtonian mechanics. It can be stated thus: *the mechanical energy of a closed conservative system is retained.*

19.2. MECHANICAL ENERGY AND FRICTION

The law of the conservation of energy in Newtonian mechanics, formulated above, is an idealization, to some degree, of the conditions of a real experiment. It is a fact that friction forces act in all real systems, and that they are nonconservative. Therefore, the law of the conservation of mechanical energy is only approximately true, and is the more accurate, the less the friction forces in comparison to elastic, electric or gravitational forces.

Experience shows that systems of bodies exist where the friction forces can be neglected as a first approximation. In such cases, calculations based on the law of conservation of mechanical energy agree well with data obtained in practice.

Even if the friction forces cannot be neglected, it is possible to introduce a correction, rewriting the law of conservation of mechanical energy as follows:

$$E_1 = E_2 + W_{fr} \quad (19.4)$$

where E_1 and E_2 = mechanical energy at the beginning and end of the process

W_{fr} = work done by the forces of friction.

This, incidentally, frequently proves to be an efficient way to find the friction force.

Some examples of the application of the law of conservation of mechanical energy will be considered in the following.

19.3. SPACE VELOCITIES

1. Let us attempt to calculate the velocity that must be imparted to a body to make it fly off on a tangent and be put into orbit about the earth. We denote the velocity of the body in orbit by v_{orb} , the velocity imparted to it at the earth's surface by v and the radius of the orbit by r (Fig. 19.1).

A body in circular orbit is subject to the gravitational force of the earth which imparts to it the normal, or centripetal, acceleration $a_n = v_{orb}^2/r$. According to Newton's second law

$$G \frac{mM}{r^2} = \frac{mv_{orb}^2}{r} \quad (19.5)$$

Fig. 19.1

But we are seeking for the required velocity at the earth and not the velocity in orbit. We can find it by making use of the law of conservation of energy in Newtonian mechanics. The sums of the potential and kinetic energies at the earth's surface and in orbit are equal to each other. Thus

$$K_{sur} + U_{sur} = K_{orb} + U_{orb} \quad (19.6)$$

or

$$\frac{mv^2}{2} - \frac{GmM}{R} = \frac{mv_{orb}^2}{2} - \frac{GmM}{r} \quad (19.7)$$

Combining with equation (19.5), we obtain

$$\frac{mv^2}{2} - \frac{GmM}{R} = \frac{GmM}{2r} - \frac{GmM}{r}$$

and, finally,

$$v = \sqrt{GM \left(\frac{2}{R} - \frac{1}{r} \right)} \quad (19.8)$$

In these calculations we did not take the resistance of the air into account. Actually, under real conditions in the atmosphere, the initial velocity imparted to the body would have to be increased. Owing to the fact that such calculations of the true required velocity are quite complex, they cannot be given here.

2. Let us calculate the numerical value of the *circular orbital velocity* v_{co} , i.e. the velocity it is necessary to impart to a body to put it into circular orbit near the earth's surface. Here $r \approx R$ and, therefore

$$v_{co} = \sqrt{GM \left(\frac{2}{R} - \frac{1}{R} \right)} = \sqrt{\frac{GM}{R}} \quad (19.9)$$

Since at low altitudes $g = GM/R^2$,

$$v_{co} = \sqrt{gR} = \sqrt{9.81 \text{ m/s}^2 \times 6.37 \times 10^6 \text{ m}} = 7.91 \times 10^3 \text{ m/s} \approx 8 \text{ km/s}$$

Thus, the circular orbital velocity is 8 km/s.

3. The *escape velocity* v_{es} is the minimum velocity that it is necessary to impart to a body for it to leave the earth's gravitational field and become an artificial planet of the sun. Here we can assume that the body is launched in an infinitely distant orbit about the earth. Putting $r \rightarrow \infty$ in equation (19.8) we can write

$$\begin{aligned} v_{es} &= \sqrt{2GM/R} \\ &= v_{co}\sqrt{2} = 7.91 \times 1.414 \\ &= 11.2 \text{ km/s} \end{aligned}$$

The escape velocity is almost one and a half times the circular orbital velocity.

19.4. LOOPING THE LOOP

Assume that a bicyclist coasts, without rotating the pedals, down the vertical portion of a loop-the-loop track (Fig. 19.2). Find the minimal height H at which he must start his descent so that he doesn't fall away from the track at the top point of the loop, i.e. at the most dangerous portion of his path.

At the top point, the bicyclist is subject to two forces: the force of gravity $P = mg$ and the reaction N of the track. Their resultant

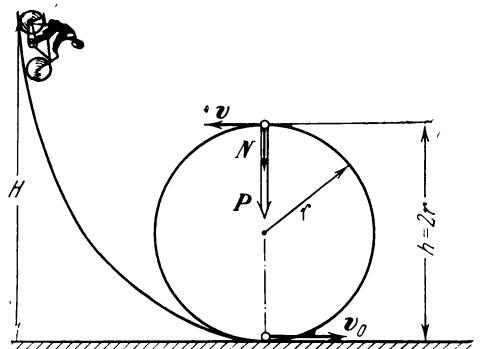


Fig. 19.2

imparts the normal acceleration $a_n = v^2/r$ to the bicyclist. According to Newton's second law

$$mg + N = mv^2/r \quad (19.10)$$

To find the velocity at the top point of the loop, we shall make use of the law of conservation of energy: the potential energy at the beginning of the descent (at the height H) should equal the sum of the kinetic and potential energies at the top point of the loop (at the height $h = 2r$). Thus

$$mgH = mg \times 2r + \frac{mv^2}{2} \text{ or } H = 2r + \frac{v^2}{2g}$$

Substituting the value of v^2 from equation (19.10), we obtain

$$H = \frac{5r}{2} + \frac{Nr}{2mg}$$

The minimal height H (neglecting friction) is obtained when the bicyclist just barely touches the track at the top point of the loop. Then the reaction of the track $N = 0$ and $H = 2.5r$.

19.5. POTENTIAL ENERGY CURVES

1. Often a particle can move only along a certain given curve, for instance along the axis of abscissas. In such cases, its potential energy depends only on a single variable, i.e. the potential energy can be expressed as the function $U = f(x)$.

The graph showing the dependence of the potential energy on the distance is called the *potential energy curve*. It has been found that an analysis of the shape of this curve yields much information on the kind of motion of the particle.

As an example, let us consider the motion of a particle subject to an elastic force (Fig. 19.3). At $x = x_0$ the spring is unstretched and the force acting on the particle equals zero. When the particle is displaced from its equilibrium position, it is subject to the force $F = -k(x - x_0)$. Note

that when $x > x_0$, the force is negative (attraction due to tension), and at $x < x_0$, it is positive (repulsion due to compression).

The potential energy of the particle [see equation (18.17)] is

$$U = \frac{k(x - x_0)^2}{2} \quad (19.11)$$

Fig. 19.3

It is represented in the graph (Fig. 19.3) as a parabola with its vertex at $x = x_0$. The mechanical energy of the particle, $E = K + U$, is constant; it is shown in the graph as a straight line parallel to the axis of abscissas.

Primarily, it is evident from the graph that the kinetic energy at any point can be found as the intercept from the straight line E to the parabola because $K = E - U$. The kinetic energy of the particle reaches its maximum value at $x = x_0$; here $U = 0$ and $K_{\max} = E$. At points $x = x_1$ and $x = x_2$, the kinetic energy equals zero because $U_{\max} = E$.

It is also evident from the graph that the particle cannot be displaced to the right of point x_2 or to the left of point x_1 . Indeed, the kinetic energy cannot be a negative quantity and, consequently, the potential energy cannot become greater than the total mechanical energy*. In such cases, the particle is said to be in a *potential energy well* with the coordinates $x_1 \leq x \leq x_2$.

2. An analysis of the inclination of the potential energy curve enables us to determine the sign of the force directly, and thereby, the nature of its action (attraction or repulsion). In fact, an element of work $\Delta W = F\Delta x$, but, on the other hand, $\Delta W = U_1 - U_2 = -\Delta U$. Therefore, if the force is a function of only one coordinate, for example the abscissa x , then $F\Delta x = -\Delta U$, or

$$F = -\frac{\Delta U}{\Delta x} \quad (19.12)$$

But in the graph $\Delta U/\Delta x = \tan \alpha$, where α is the angle between a tangent to the potential energy curve and the axis of abscissas (Fig. 19.4). Strictly, the exact value of the force is obtained only in the limit, when the displacement Δx tends to zero. Thus

$$F = -\lim_{\Delta x \rightarrow 0} \frac{\Delta U}{\Delta x}$$

However, we are interested here only in the sign of the force, and that can be obtained without the passage to the limit.

If the potential energy increases, the potential energy curve makes an acute angle with the axis of abscissas. The tangent of an acute angle is positive, and the force is of the opposite sign, i.e. negative. Hence, it is a force of attraction.

If the potential energy decreases, the potential energy curve makes an obtuse angle with the axis of abscissas, and its tangent is negative. Here, the force is positive, i.e. a force of repulsion.

Finally, at maxima or minima of the energy curve, the force evidently equals zero because its sign reverses at such points. In the

* Quantum mechanics introduces essential corrections into this conclusion, arising from the special nature of microparticles and the uncertainty principle (see Secs. 70.2 through 70.6).

graph, the tangent to the potential energy curve at a maximum or minimum is parallel to the axis of abscissas (Fig. 19.5).

It will be left as an exercise for the student to show that in the case illustrated in Fig. 19.6, a particle with the total energy $E < U_0$ will either be in a potential energy well $x_1 \leq x \leq x_2$ or will move

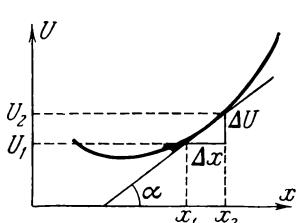


Fig. 19.4

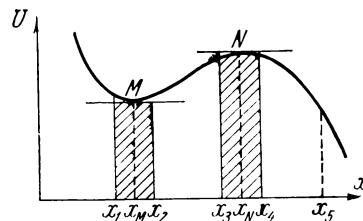


Fig. 19.5

away to an infinitely distant point, and that it (the particle) will not be able to surmount the potential barrier $x_2 \leq x \leq x_3$ from either the right or the left.

3. Attention should be drawn here to an extremely interesting property of the potential energy of forces of attraction. If the distance decreases between the particles being attracted to each other,

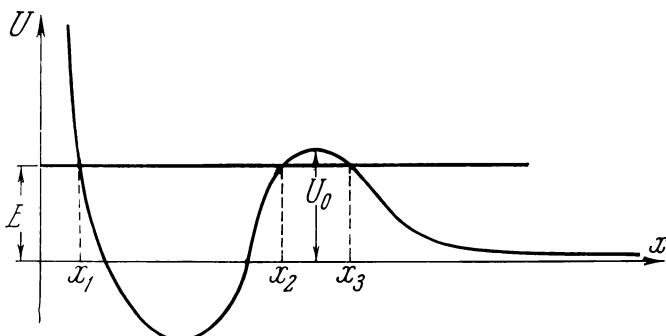


Fig. 19.6

the potential energy also decreases and the force of attraction increases. Inversely, an increase in the distance between particles being attracted to each other is accompanied by a growth in the potential energy and a reduction in the force of attraction.

This property is entirely general and is possessed by any forces of attraction. True, its proof in the general form is somewhat complicated, and we shall not undertake it here. It will be left to the student to test the validity of this statement, using Coulomb and gravitational forces as examples.

This can be expediently done by analyzing the potential energy curve shown in Fig. 19.7 for the gravitational interaction of two particles and comparing the energies and forces at points with the coordinates r_1 and r_2 . In plotting the curve we assumed that the potential energy equals zero at an infinitely great distance between the interacting particles [see equation (18.19)]. However, this is of no significance in principle; the result of our reasoning is entirely independent of the place we choose for the zero level of the potential energy.

19.6. POTENTIAL ENERGY AND EQUILIBRIUM

1. An analysis of its potential energy curves gives one an idea of the equilibrium conditions of a body (or, more precisely, a particle).

A body is in equilibrium if the resultant of all forces applied to it equals zero. Evidently, equilibrium corresponds to points with either minimum or maximum potential energy (points M and N in Fig. 19.5) because it is precisely at these points that the force becomes zero. But the points of maximum and minimum potential energy are not equivalent.

This can be demonstrated by assuming that the particle is at point x_M where the potential energy has a minimum. Over the portion $x_1 \leq x < x_M$ the potential energy decreases. This means that the particle is subject to a positive force of repulsion, returning it to point M . Over the portion $x_M < x \leq x_2$ the potential energy increases. Therefore, the particle is subject to a negative force of attraction, again returning it to point M .

Thus, if a particle located at a point of minimum potential energy is displaced from this point, it will be returned by the force to this point. We conclude that a *minimum value of the potential energy is a condition for stable equilibrium*.

By similar reasoning with respect to point N , where the potential energy of the particle has a maximum, the student can readily see that forces acting on the particle in the near vicinity of this point do not tend to return it to the point. Hence, a *maximum value of the potential energy corresponds to a state of unstable equilibrium*.

2. In analyzing the conditions of equilibrium we considered only the *immediate neighbourhood* of the point of the field we were inter-

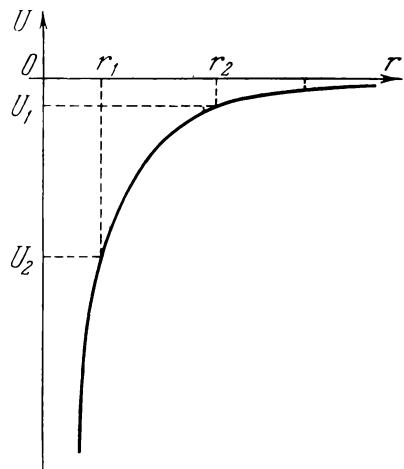


Fig. 19.7

rested in. This can be defined as a region in which there are no additional maxima or minima of energy. If this factor is disregarded, we can reach incorrect conclusions.

For example, in analyzing the forces acting on a particle displaced to the right of the position of stable equilibrium ($x_i > x_M$), we contended that they would be forces of attraction. But this is true only as long as the particle remains to the left of an energy maximum. If we displace the particle farther to the right, for instance, to point $x_5 > x_N$ (see Fig. 19.5), we shall obtain a force of repulsion, and the particle will not return to its previous position.

Chapter 20

INTERNAL ENERGY

20.1. INTERNAL ENERGY OF A SYSTEM OF PARTICLES

1. It was shown in Chapter 16 that the total energy of a body can be expressed as the sum of its kinetic and rest energies. The kinetic energy of the body depends upon its velocity with respect to the chosen reference frame, i.e. relative to external bodies.

The rest energy of a body depends upon processes occurring inside it. Hence the rest energy is also called the *internal energy*.

2. Let a body of rest mass M_0 consist of N particles whose rest masses are $m_0^{(1)}, m_0^{(2)}, \dots, m_0^{(N)}$. Further, let these particles move with arbitrary velocities v_1, v_2, \dots, v_N , but in such a way that their centre of mass remains stationary in the given reference frame, i.e. $m_1 v_1 + m_2 v_2 + \dots + m_N v_N = 0$.

Finally, let us assume that these particles are at considerable distances from one another, so that, as a first approximation, their interaction can be neglected.

It would seem, on the face of it, that the rest energy of a body is simply equal to the sum of the rest energies of its component particles. But this is not so. Indeed, even though the body is at rest in the given reference frame, the particles making up the body move in quite definite ways. Therefore, the *internal energy* \mathcal{E}_0 (i.e. its rest energy) of a body equals the sum of the total energies of the particles of which the body consists. Thus

$$\mathcal{E}_0 = \mathcal{E}^{(1)} + \mathcal{E}^{(2)} + \dots + \mathcal{E}^{(N)} \quad (20.1)$$

Since, according to equation (16.3), $\mathcal{E}^{(i)} = \mathcal{E}_0^{(i)} + K^{(i)}$, where i is the number of the particle, we can write

$$\begin{aligned} \mathcal{E}_0 &= \mathcal{E}_0^{(1)} + \mathcal{E}_0^{(2)} + \dots + \mathcal{E}_0^{(N)} + K^{(1)} + K^{(2)} + \dots + K^{(N)} \\ &= \mathcal{E}_0^{(1)} + \mathcal{E}_0^{(2)} + \dots + \mathcal{E}_0^{(N)} + K_{int} \end{aligned} \quad (20.2)$$

where $K_{int} = K^{(1)} + K^{(2)} + \dots + K^{(N)}$ is the total kinetic energy of the internal motion of the particles making up the body.

3. In the preceding, we dealt with a system of noninteracting particles. Actually, the particles interact quite strongly with one another in all bodies (with the exception of dilute gases), usually by forces of attraction. The potential energy of forces of attraction is a *negative* value because we took the potential energy to be equal to zero in the absence of forces of interaction (see Sec. 18.6).

Let us denote by U_{int} the total potential energy of interaction of the particles. Actually, U_{int} represents the energy of the fields by means of which interaction is accomplished between the particles inside a body. Then, generalizing equation (20.2) for the case of interacting particles, we can write

$$\mathcal{E}_0 = \mathcal{E}_0^{(1)} + \mathcal{E}_0^{(2)} + \dots + \mathcal{E}_0^{(N)} + K_{int} + U_{int} \quad (20.3)$$

Thus, the *internal energy of a body equals the sum of the rest energies of the particles constituting the body plus the total kinetic energies of these particles and potential energies of the fields that accomplish interaction between the particles.*

4. Dividing both sides of equation (20.3) by c^2 , and recalling that $\mathcal{E}_0/c^2 = M_0$ is the rest mass of the body and that $\mathcal{E}_0^{(i)}/c^2 = m_0^{(i)}$ is the rest mass of the i -th particle, we obtain

$$M_0 = m_0^{(1)} + m_0^{(2)} + \dots + m_0^{(N)} + \frac{1}{c^2} (K_{int} + U_{int}) \quad (20.4)$$

Thus we find that, in the general case, the *rest mass of a body does not equal the sum of the rest masses of the particles making up the body.*

Specifically, in the presence of large forces of attraction, the magnitude of the potential energies of the fields may turn out to be substantially greater than the total kinetic energy of internal motion. And, since the potential energy of forces of attraction is a negative value, we may find that $K_{int} + U_{int} < 0$. Hence, in this case, the rest mass of the body will be less than the sum of the rest masses of the particles constituting the body.

We shall encounter this phenomenon again in the second volume in studying nuclear forces and methods of utilizing nuclear energy.

20.2. CHANGES IN INTERNAL ENERGY WHEN A BODY IS DEFORMED

1. A body is deformed when a force is applied to it, i.e. its size and shape are changed. Deformation of a body involves changes in the distances between its molecules and, thereby, changes in the energy of molecular interaction. Consequently, *deformation of a body is accompanied by a change in its internal energy.*

But if the internal energy of the body changes, its mass also changes. Why cannot we detect this change in mass? The following example may serve as an answer to this question.

A steel bar of length $l = 1 \text{ m}$ and cross section $A = 1 \text{ cm}^2$ is elongated $\Delta l = 1 \text{ mm}$ by a force of $F = 2 \times 10^4 \text{ N}$. The work done by the elastic force is

$$W = \frac{kx^2}{2} = \frac{F\Delta l}{2} = 10 \text{ J}$$

It is equal to the change in the internal energy of the body: $W = \Delta \mathcal{E}_0$. The change in its mass is

$$\Delta m = \frac{\Delta \mathcal{E}_0}{c^2} = \frac{10}{9 \times 10^{16}} = 1.1 \times 10^{-16} \text{ kg}$$

Before being deformed the mass of the bar was

$$m = \rho l A = 7.8 \times 10^3 \times 1 \times 10^{-4} = 0.78 \text{ kg}$$

As is evident from these calculations, the change in mass of the bar is extremely small and appears only in the sixteenth place after the decimal point! No instruments exist that can detect such a minute quantity. Therefore, the change in mass of a body when it is deformed is neglected in practice.

2. Next, let us calculate the change in the energy of a single molecule when a body is deformed. One kilogram of steel comprises 10^{25} molecules. Our bar with a mass of 0.78 kg is made up of $N = 0.78 \times 10^{25}$ molecules. Hence, the change in the energy of a single molecule is

$$\varepsilon = \frac{\Delta \mathcal{E}_0}{N} = \frac{10}{0.78 \times 10^{25}} = 1.3 \times 10^{-24} \text{ J}$$

This energy can be expressed more conveniently in electron volts. As mentioned previously, $1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$. Therefore

$$\varepsilon = \frac{1.3 \times 10^{-24}}{1.6 \times 10^{-19}} \cong 10^{-5} \text{ eV}$$

Thus, the change in the energy of a molecule when a body is deformed is of the order of several millionths of an electron volt. In the following we shall compare this value with the changes in the energy of molecules in other physical processes.

20.3. CHANGES IN INTERNAL ENERGY OF A BODY IN THERMAL PROCESSES

1. Processes involving changes in the temperature of a body, as well as with changes in its state of aggregation, such as melting or freezing, evaporating or condensing, are said to be thermal. These processes will be dealt with in detail later in the text. Here we shall restrict ourselves to a general description of these processes as regards the changes in internal energy.

Changes in the temperature of a body are associated with changes in the kinetic energy of motion of its molecules. It should be emphasized, however, that, at the same time, the potential energy of their interaction also changes (excepting for dilute gases). Indeed, any change in temperature is accompanied by a change in the distances between the equilibrium positions of the lattice points in the crystal lattices of the body. We perceive this as thermal expansion of the body. This, naturally, changes the interaction energy of the molecules. Changes in the state of aggregation are the result of changes in the molecular structure of bodies, which cause changes both in the interaction energy of the molecules and in the nature of their motion.

Thus, *thermal processes are associated with changes in the internal energy of bodies.*

2. Let us estimate the change that occurs in the energy of molecules in thermal processes. It is known that in raising the temperature of 1 kg of water by 1 deg C, $1 \text{ kcal} = 4.18 \times 10^3 \text{ J}$ of energy is added to the water. But one kilogram of water comprises 3.3×10^{25} molecules. Hence

$$\varepsilon = \frac{\Delta \mathcal{E}_0}{N} = \frac{4.18 \times 10^3}{3.3 \times 10^{25} \times 1.6 \times 10^{-19}} \cong 10^{-3} \text{ eV}$$

This is approximately a hundred times as much as the energy involved in elastic deformation. The energy required to convert 1 kg of water into steam is $539 \text{ kcal} = 2.25 \times 10^6 \text{ J}$. The energy of each molecule increases by about

$$\varepsilon = \frac{\Delta \mathcal{E}_0}{N} = \frac{2.25 \times 10^6}{3.3 \times 10^{25} \times 1.6 \times 10^{-19}} \cong 0.5 \text{ eV}$$

3. Notwithstanding the fact that in thermal processes the internal energy increases by a thousand times more than in elastic deformation, it is still impossible to measure the change in the mass of the body. Indeed, even though the internal energy of 1 kg of water is increased by $2.25 \times 10^6 \text{ J}$ when it is evaporated, its mass is increased by only

$$\Delta m = \frac{\Delta \mathcal{E}_0}{c^2} = \frac{2.25 \times 10^6}{9 \times 10^{16}} = 2.5 \times 10^{-10} \text{ kg}$$

which is far beyond the limits of sensitivity of measuring apparatus.

20.4. CHANGES IN INTERNAL ENERGY IN CHEMICAL REACTIONS

1. Chemical reactions are ones involving the rearrangement of molecules, their decomposition into simpler components or, inversely, the formation of more complex molecules of simpler ones or of separate atoms (decomposition and fusion reactions). Here, an essential change occurs in the forces of interaction between the atoms, this being accompanied by a corresponding change in the interaction energy. Changes occur, moreover, in the nature of both the motion of the

molecules and in their interaction, because the molecules of the newly formed substance interact differently from those of the initial substances.

Thus, *chemical reactions are associated with changes in the internal energy of bodies.*

Note, also, that chemical reactions are always accompanied by thermal processes—a change in temperature and, frequently, in the state of aggregation of the system. Thus, in the combustion of solid carbon, gaseous products are formed: carbon monoxide (CO) and carbon dioxide (CO_2).

2. Let us estimate the change in the energy of molecules in chemical reactions. About $7000 \text{ kcal} = 3 \times 10^7 \text{ J}$ of energy are evolved in the combustion of 1 kg of coal. Since 1 kg of coal consists of 5×10^{25} atoms of carbon, the change in the energy of a single molecule is

$$\varepsilon = \frac{\Delta \mathcal{E}_0}{N} = \frac{3 \times 10^7}{5 \times 10^{25} \times 1.6 \times 10^{-19}} \cong 4 \text{ eV}$$

Therefore, the change in energy per molecule in chemical reactions is several electron volts, which is hundreds and thousands of times more than in thermal processes and millions of times more than in elastic deformation.

It should be noted that this energy corresponds to the average value of the kinetic energy of an electron in an atom (Sec. 16.7). This was to be expected since chemical reactions consist in the rearrangement of the electron shells of atoms and molecules.

3. But here again the change in the mass of the interacting bodies is so small that it lies beyond the limits of accuracy of the most precise measuring instruments: it appears only in the tenth significant digit, and we are incapable of measuring mass with such accuracy. Indeed, in the combustion of 1 kg of coal, the change in mass is

$$\Delta m = \frac{\Delta \mathcal{E}_0}{c^2} = \frac{3 \times 10^7}{9 \times 10^{16}} = 3.3 \times 10^{-10} \text{ kg}$$

20.5. CHANGES IN INTERNAL ENERGY IN NUCLEAR REACTIONS

1. Under certain conditions, the nuclei of atoms undergo transformations called *nuclear reactions*. Regardless of the mechanism of the processes that occur in such reactions (and they may differ greatly), they are all associated with considerable changes in the energy of the interacting particles. Consequently, *nuclear reactions are accompanied by a change in the internal energy of the body of which the nuclei subject to transformation are a part.* Here the rest masses of the reacting nuclei are changed. This enables the energy yield of the reaction (see Sec. 80.4) to be estimated from the known masses of the initial particles and the products of the reaction.

It was found that if 1 kg of heavy hydrogen (deuterium) is converted into helium, only 993.75 g of helium are obtained as the result of the reaction, i.e. the rest mass is reduced by $\Delta m_0 = 6.25$ g. And such a reduction in mass (by 0.6 per cent) can readily be measured by sensitive instruments.

The change in the internal energy in this reaction constitutes

$$\Delta \mathcal{E}_0 = \Delta m_0 c^2 = 6.25 \times 10^{-3} \times 9 \times 10^{16} = 5.62 \times 10^{15} \text{ J}$$

For comparison, we shall calculate the amount of coal that must be burned to obtain the same energy:

$$m_{\text{coal}} = \frac{5.62 \times 10^{14}}{3 \times 10^7} \cong 2 \times 10^7 \text{ kg}$$

(here 3×10^7 J is the energy evolved in the combustion of 1 kg of coal). Since a railway coal car accommodates about 50 tons of coal, we find that the energy evolved in the conversion of 1 kg of deuterium into helium is equal to that evolved in burning 400 cars of coal.

2. Let us estimate the change in energy of a single nucleus in such a reaction. One kilogram of helium consists of 1.5×10^{26} atoms. Therefore

$$\varepsilon = \frac{\Delta \mathcal{E}_0}{N} = \frac{5.62 \times 10^{14}}{1.5 \times 10^{26} \times 1.6 \times 10^{-19}} = 2.34 \times 10^7 \text{ eV} = 23.4 \text{ MeV}$$

This energy is a million times more than that obtained in chemical reactions and a thousand million times more than in thermal processes.

It should be noted that these figures also characterize the order of magnitude of nuclear forces as compared to the interaction forces between atoms and between molecules.

CHAPTER 21

THE LAW OF CONSERVATION OF ENERGY

21.1. WORK AS A MEASURE OF THE CHANGE IN TOTAL AND INTERNAL ENERGY

1. We found in Sec. 20.2 that in the deformation of a body the distances are changed between the particles of which it consists. And this leads to a change in the internal energy of the body.

Besides, the temperature of the body is changed in inelastic deformation, as is readily evident from a simple experiment. If we rapidly bend a piece of wire back and forth several times, the temperature

at the bend will increase sharply. This can be felt by touching the bend with a finger.

But work is done in deforming a body. This is precisely the work that is the measure of the change in the internal energy.

2. The internal energy of a body is also changed in its inelastic collision with another body. At this, as shown in Sec. 17.2, the kinetic energy of the colliding bodies decreases. But it is not to be implied that the kinetic energy simply vanishes without leaving a trace. We shall find, without difficulty, that it is converted into internal energy of the body.

Let us conduct the following experiment. We place a piece of wire on an anvil and strike it several heavy blows with a hammer. Now, if we touch the flattened part of the wire we shall find that its temperature has been substantially raised.

It is evident that in an inelastic collision of bodies, the kinetic energy is converted into internal energy. But a measure of the change in kinetic energy is work done by a force. Quite naturally, the work done by this force can also serve as a measure of the change in internal energy.

3. Finally, a change in internal energy is due to the action of friction forces since, as we know from experience, friction is always accompanied by a change in temperature of the rubbing bodies. Naturally, the work done by the forces of friction can serve as a measure of the change in internal energy.

4. The examples given above enable us to establish a more general relationship between the work done by a force and the change in energy than that derived in Chapter 16.

Let the total energy of a body in a certain state be equal to $\mathcal{E}^{(1)} = \mathcal{E}_0^{(1)} + K^{(1)}$. If a certain external force acts on this body its state is changed, and its total energy will change correspondingly. It will become equal to $\mathcal{E}^{(2)} = \mathcal{E}_0^{(2)} + K^{(2)}$. The change in energy is $\Delta\mathcal{E} = \mathcal{E}^{(2)} - \mathcal{E}^{(1)} = (\mathcal{E}_0^{(2)} - \mathcal{E}_0^{(1)}) + (K^{(2)} - K^{(1)})$ or, more briefly,

$$\Delta\mathcal{E} = \Delta\mathcal{E}_0 + \Delta K \quad (21.1)$$

But we already know that the work done by an external force can change both the kinetic and internal energies of a body, i.e. its total energy. Therefore, we can write

$$\Delta\mathcal{E} = W_{ext} \quad (21.2)$$

i.e. the work done by an external force is a measure of the change in the internal energy of a body.

21.2. HEAT EXCHANGE

1. There is still another method of changing the energy of a body; it is not associated with the work done by forces. If a body is placed in the flame of a Bunsen burner its temperature is changed and, con-

sequently, so is its internal energy. If we now submerge the heated body in cold water, we find that its temperature is changed again, and again this is associated with a change in its energy. But this change in energy did not require any work to be done by a force acting on the body.

Actually, the change in the internal energy of the body occurred as a result of molecular collisions. The molecules of a heated body move at higher velocities; those of a cold body, at lower ones. In contact between these bodies, elastic collisions occur between their molecules. A rapidly moving particle, colliding with another particle having a lower velocity, transmits a part of its kinetic energy to the slower particle. It follows, naturally, that when two bodies with different temperatures come into contact, the total kinetic energy of the molecules of the hotter body is reduced, due to molecular collisions, and the total kinetic energy of the molecules of the cooler body is increased. As a result of this phenomenon, energy is transmitted from a hot body to a cool one without any work being done.

The process of transmitting energy from one body to another without doing work is called heat exchange.

2. The process of heat exchange considered above between bodies in direct contact is called *heat conduction*, or *heat exchange by conduction*.

Heat exchange can also be accomplished by convection. A layer of liquid or gas, in contact with a body having a high temperature, is heated and rises, permitting a colder layer to take its place. Heat exchange accomplished by a current of liquid or gas is called *convective heat exchange*.

3. The third general method of heat exchange is by means of visible and invisible rays. This can be demonstrated by the following experiment. Let us clamp a heated body, for instance an electric iron, on a support. Now let us put a hand under the iron, at a distance of 15 to 20 cm from its flat surface. We feel how the iron transmits energy to our hand. This process cannot be convection, because the current of heated air rises upward and does not reach our hand. It cannot be heat conduction either, because all gases, and specifically air, are poor conductors of heat.

It was found that any heated body emits infrared rays and, at temperatures above 800 °C, visible light as well. The process of transmitting energy by electromagnetic waves (infrared and visible radiation) is called *heat exchange through radiation*, or *radiative heat exchange*.

4. In the experiment described above we were actually dealing with radiation. Indeed, if we place a sheet of tin between the iron and our hand, we shall no longer feel that our hand is being heated. The reason is that the metal partly reflects and partly absorbs the radiation which, thereby, does not reach the hand.

Finally, it should be noted that while heat conduction and heat convection require the presence of some substance between the two bodies, heat radiation can take place in a vacuum as well. Thus, the energy radiated by the sun reaches the earth after passing through outer space with its high vacuum.

21.3. QUANTITY OF HEAT

1. A measure of the change in the energy of a body in mechanical processes is the work done by external forces on the body [see equation (21.2)]. We now introduce a physical quantity that characterizes the change in energy of a body in heat exchange. This is called the *quantity of heat* and is denoted by Q .

The quantity of heat is the energy transmitted by heat exchange.

It follows from this definition that for the case when the energy of the body is changed only by heat exchange

$$Q = \Delta \mathcal{E} \quad (21.3)$$

i.e. the *quantity of heat is a measure of the change in the energy of a body in heat exchange.*

2. The SI unit for the quantity of heat, as that for work and energy, is the *joule* (see Sec. 16.6). Thus

$$[Q] = [\mathcal{E}] = [W] = \text{J}$$

A nonsystem unit of the quantity of heat employed in practice is the *calorie*. By definition

$$1 \text{ Cal} = 4.1868 \text{ J}$$

3. Let us recall the formulas for calculating the amount of heat that are studied in elementary school:

(a) When the temperature of a body is changed

$$Q = mc\Delta t \quad (21.4)$$

where m = mass of the body

Δt = change in temperature

c = specific heat of the substance constituting the body.

(b) In melting a crystal (or in freezing a melt):

$$Q = m\lambda \quad (21.5)$$

where λ is the heat of fusion.

(c) In evaporating a liquid (or in condensing a vapour):

$$Q = mL \quad (21.6)$$

where L is the heat of vapourization.

(d) In a chemical reaction:

$$Q = mq \quad (21.7)$$

where q is the heat of reaction. Thus, the amount of heat evolved in burning 1 kg of fuel is called the calorific power (heat of combustion).

4. It should be noted that in everyday usage the term "heating" has two meanings: to transmit energy in the form of heat (as in heat exchange) and to raise the temperature. As a matter of fact, these two processes are not equivalent. The temperature of a body can be changed without transmitting heat to it. In an inelastic collision, inelastic deformation, in friction, etc. the temperature of a body is raised by doing work on it and not by heat exchange. On the contrary, in such processes as melting or boiling, the temperature of the body is not changed though energy is transmitted to it in the form of heat. Hence, for a precise description of a phenomenon, the ambiguous term "heating" should be avoided, indicating the exact essence of the process: a change in temperature or the transmission of energy by heat exchange.

21.4. THE FIRST LAW OF THERMODYNAMICS

The science that studies the conditions for the conversion of energy from one form to another, and characterizes these conversions quantitatively is called *thermodynamics*.

1. It was shown above [equations (21.2) and (21.3)] that the energy of a body can be changed either due to the action of external forces or by heat exchange with the environment. Also possible is the case when the energy of a body is changed by both of these causes. Then the change in the energy of the system is equal to the sum of the work done by the external forces plus the amount of heat obtained from the environment. Thus

$$\Delta E = W_{ext} + Q \quad (21.8)$$

2. The work done by a system in overcoming the external forces differs from the work done by the external forces only by its sign. Thus $W = -W_{ext}$. Then equation (21.8) can be rewritten as

$$\Delta E = -W + Q$$

or

$$Q = \Delta E + W \quad (21.9)$$

i.e. the *quantity of heat* (energy obtained by the system by heat exchange) is expended in changing the energy of the system and in work done to overcome external forces.

3. This expression is called the *first law of thermodynamics*. It is exceptionally general in meaning and can be applied to describe a very wide range of phenomena. Further on, we shall use this law to describe many thermal and other processes.

In the subsequent sections of this chapter we shall find the relationship between the first law of thermodynamics and one of the most important laws of nature—the law of the conservation of energy.

21.5. AN ADIABATICALLY ISOLATED SYSTEM

1. A system is said to be *closed* if the external forces can be neglected in comparison to the internal forces (see Sec. 15.1). From this definition it follows that in a closed system the work done by the external forces equals zero.

An *adiabatically isolated system* is one that has no heat exchange with its environment. According to this definition, the quantity of heat received by an adiabatically isolated system equals zero.

2. The concept of an adiabatically isolated system is an idealization of certain real systems. It is impossible, of course, to completely exclude heat exchange from any system, but it is feasible to make the processes occurring in the given system approach adiabatic processes.

Heat exchange usually takes place very slowly, considerably more slowly than the transmission of energy by other processes. For this reason any rapid process, such as an explosion or flash, can be regarded as adiabatic to a high degree of accuracy.

A system can also be made to approximate an adiabatic one by impeding heat exchange with its environment, i.e. by providing reliable heat insulation. A good example of an adiabatic jacket is a Dewar flask, which is the main part of the common thermos bottle (Fig. 21.1). This is a thin-walled glass vessel with double walls between which a good vacuum is provided. The vacuum minimizes heat conduction and convection. The inside wall of the vessel is silvered. The obtained mirror surface reflects rays and thereby impedes radiative heat exchange.

A system enclosed in a jacket of the Dewar type can be assumed with a high degree of accuracy to be adiabatically isolated.

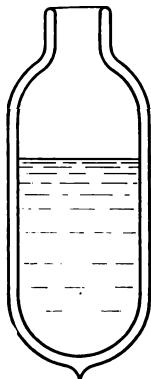


Fig. 21.1

21.6. THE LAW OF CONSERVATION OF ENERGY

1. Let us consider a certain system of bodies that does not interact mechanically with surrounding bodies (being a closed system) and has no heat exchange with external bodies (being an adiabatically isolated system). From the fact that it is closed, it follows that the system does no work to overcome external forces, i.e. $W = 0$. From the condition of adiabatic isolation, it follows that the system neither receives nor gives up energy by heat exchange, i.e. $Q = 0$.

Then from the first law of thermodynamics, equation (21.9), we find that in this case the total energy of the system does not change: $\Delta E = 0$ and, therefore, $E = \text{const}$.

Hence, the *total energy of a closed and adiabatically isolated system is a constant*. This is the most general expression for the law of conservation of energy.

2. The law of conservation of energy in Newtonian mechanics (see Sec. 19.1) is a special case of the law obtained above. Indeed, in Chapter 19 we dealt only with conservative systems, and in such systems there is no heat exchange. Hence, only one condition—the closed nature of the system—was required for the conservation of its energy.

3. The law of conservation of energy does not require at all that the energies of each of the bodies making up the system remain unchanged. Both mechanical interaction and heat exchange may occur between these bodies. Therefore, the energies of the separate bodies included in the given system will undoubtedly change. To conserve the energy of the system it must not interact with external bodies nor participate in heat exchange. Then the total energy of the system will remain constant, even though the energies of its components change.

21.7. THE LAW OF CONSERVATION OF MASS

1. If the total energy of a system is conserved, then its relativistic mass, proportional to this energy, is also conserved. Hence the *law of conservation of mass* can be derived from the law of conservation of energy:

The relativistic mass of a closed and adiabatically isolated system is a constant.

As we see, the law of conservation of mass is a corollary of the law of conservation of energy. That is why we often speak of the unified law of conservation of energy and mass.

The question of the rest mass (proper mass of the system) is considerably more complicated.

2. Newtonian mechanics deals with bodies moving with low velocities. Here the mass of a body practically coincides with its proper mass. The change in the energy of a body due to its motion, deformation, thermal or chemical processes is negligibly small in comparison to its rest mass (see Secs. 16.2, 20.2, 20.3 and 20.4) so that the latter can be assumed to be constant. Moreover, Newtonian mechanics does not deal with the conversion of certain forms of energy into other forms. For instance, it is impossible, within the limits of this theory, to transform particles of matter into other particles, into an electromagnetic or other fields. Therefore, mass is considered in Newtonian mechanics to be a constant value for a given body (independent of its energy), and the mass of a system, to be the sum

of the masses of the bodies included in the system. Naturally, the *law of conservation of mass* is valid in Newtonian mechanics independently of the law of conservation of energy. Experience excellently supports this law for the processes treated in Newtonian mechanics.

3. In the theory of relativity, the rest mass of a system is not equal to the sum of the rest masses of the bodies of which the system consists (see Sec. 20.1). Strictly, this alone is sufficient to contend that, in the general case, the rest mass is not a conserved quantity.

But, besides this, certain particles may be converted into others, and particle production and annihilation is possible at high kinetic energies. We restrict ourselves to a single example.

4. Let a fast proton, accelerated in a synchrotron to a velocity near to the speed of light, strike a stationary proton. It was found that four particles are produced, if the fast proton possesses sufficient kinetic energy, instead of the initial two: three protons and one antiproton.

The reaction of this event is written as follows:

$${}_1p^1 + {}_1p^1 \rightarrow {}_1p^1 + {}_1p^1 + {}_1p^1 + {}_{-1}\tilde{p}^1$$

where ${}_1p^1$ is a proton and ${}_{-1}\tilde{p}^1$ is the antiproton (a particle of a mass equal to that of the proton but oppositely charged). Let us denote the rest energy of the proton by \mathcal{E}_0 . Then the rest energy of the system before the collision equals $2\mathcal{E}_0$, and after the collision it becomes $4\mathcal{E}_0$. Thus the rest mass is not conserved here, and kinetic energy is converted into rest energy.

Making use of the laws of conservation of energy and momentum, let us determine the minimum kinetic energy of a proton which is capable, upon striking a stationary hydrogen target, of producing an antiproton. For the sake of simplicity we shall assume that all the particles move with the same momentum p_1 after the collision. According to the law of conservation of linear momentum

$$p = 4p_1 \quad (21.10)$$

where p is the momentum of the fast proton before it strikes the target. Using the relationship, equation (16.8), between energy and momentum, we can write

$$\mathcal{E}^2 - \mathcal{E}_0^2 = p^2 c^2 \quad \text{and} \quad \mathcal{E}_1^2 - \mathcal{E}_0^2 = p_1^2 c^2 \quad (21.11)$$

where \mathcal{E}_0 = rest energy of the proton

\mathcal{E} = energy of the fast proton before collision with the target

\mathcal{E}_1 = energy of the proton or antiproton after the collision.

Combining equations (21.10) and (21.11) to eliminate p , we obtain

$$\mathcal{E}^2 - 16\mathcal{E}_1^2 + 15\mathcal{E}_0^2 = 0 \quad (21.12)$$

It follows from the law of conservation of energy that the total energy of the two protons before the collision is equal to the sum of

the total energies of the four particles after the collision. Thus

$$\mathcal{E} + \mathcal{E}_0 = 4\mathcal{E}_1 \quad (21.13)$$

Substituting into equation (21.12) and performing simple mathematical operations, we obtain

$$\mathcal{E} = 7\mathcal{E}_0 \quad (21.14)$$

from which the kinetic energy of the fast proton before the collision was

$$K = \mathcal{E} - \mathcal{E}_0 = 6\mathcal{E}_0 \quad (21.15)$$

Of essential significance is the fact that of the kinetic energy $K = 6\mathcal{E}_0$, only $2\mathcal{E}_0$ is converted into rest energy; the remainder, $4\mathcal{E}_0$, is conserved in the form of kinetic energy of the products of the reaction: three protons and one antiproton.

5. It is left for the student to carry out numerical calculations, taking the rest mass of the proton to be $m_0 = 1.67 \times 10^{-27}$ kg. The kinetic energy of the fast proton will be found to be $K \approx \approx 6 \text{ GeV} = 6 \times 10^9 \text{ eV}$, and its velocity to be 99 per cent of the speed of light.

21.8. MORE ABOUT RELATIVISTIC MASS

1. In Sec. 13.1 we introduced equation (13.3) for the relativistic mass without any proof. Now we shall show that this formula can be obtained on the basis of three principles: the law of conservation of linear momentum (Sec. 15.2), the law of conservation of relativistic mass (Sec. 21.7) and the relativistic law for the addition of velocities (Sec. 12.5).

Let us imagine two entirely identical bodies of equal rest mass m_0 which can be established by weighing (see Sec. 6.3). Assume that the body at the right is at rest with respect to the frame of reference and the other body is moving to the right with the velocity u , parallel to the axis of abscissas. The momentum of the left-hand body equals mu . Following an inelastic collision with the right-hand body the momentum of the system is MV , where M is the total mass of the system of bodies and V is its velocity.

According to the law of conservation of relativistic mass

$$M = m + m_0 \quad (21.16)$$

while from the law of conservation of linear momentum

$$mu = (m + m_0)V \quad (21.17)$$

2. Let us now consider a reference frame attached to the left-hand body. Here, only the direction of the velocity vectors is changed, all the other quantities remain as previously. The law of conservation of linear momentum will then be

$$-mu = -V(m + m_0) \quad (21.18)$$

On the other hand, in changing from one inertial reference frame to another, the velocity is transformed according to formula (12.11) which, for the given case, is written as follows:

$$V' = -V = \frac{V-u}{1-Vu/c^2} \quad (21.19)$$

The roots of quadratic equation (21.19) are

$$V_1 = \frac{c^2}{u} (1 + \sqrt{1-u^2/c^2})$$

and

$$V_2 = \frac{c^2}{u} (1 - \sqrt{1-u^2/c^2})$$

The first root has no physical meaning because a body cannot move with a velocity higher than the speed of light (see Sec. 12.6). Thus, the velocity of the system following an inelastic collision is

$$V = \frac{c^2}{u} (1 - \sqrt{1-u^2/c^2}) = \frac{u}{1 + \sqrt{1-u^2/c^2}} \quad (21.20)$$

After substituting into equation (21.17) we obtain the expression (13.3) for the relativistic mass:

$$m = \frac{m_0}{\sqrt{1-u^2/c^2}}$$

3. If, instead of the relativistic formula, equation (21.19), we had used the classical law of the addition of velocities, equation (2.6), we would have obtained: $V' = -V = V - u$, from which $V = u/2$. After substituting into equation (21.17) we would have: $mu = 1/2(m + m_0)u$, from which $m = m_0$.

Thus, the invariance of mass in Newtonian mechanics follows from the laws of conservation of mass and momentum, together with the Galilean transformations. The same conservation laws, in conjunction with the Lorentz transformations, lead to the relativistic law of the dependence of the mass of a body on its velocity with respect to the frame of reference.

Chapter 22

THE LAW OF CONSERVATION OF ANGULAR MOMENTUM

22.1. FEATURES OF ROTATIONAL MOTION

Rotating bodies are frequently encountered in practice; they include all possible kinds of flywheels, shafts, generator and motor rotors, screws, drills, milling cutters, etc. A feature of rotational motion is that all the points of the body move in concentric circles whose centres lie on the axis of rotation. All these points move at *different* linear velocities but have the same angular velocity $\omega = \lim_{\Delta t \rightarrow 0} \frac{\Delta\alpha}{\Delta t}$ (see Sec. 4.7).

In studying the laws of motion of a particle we introduced a number of dynamic quantities: linear momentum, force, kinetic energy, etc. Actually, we used the same quantities for describing the laws of translational motion of rigid bodies. If a rigid body has translational motion, all its points move along identical paths (see Sec. 1.1) and, consequently, at the same velocity. For this reason the equation for the linear momentum or kinetic energy is of exactly the same form as for a particle. Indeed, the linear momentum of a body equals the sum of the linear momenta of the separate particles of which the body consists. Thus

$$\begin{aligned} \mathbf{p} &= \mathbf{p}_1 + \mathbf{p}_2 + \dots + \mathbf{p}_r = m_1 \mathbf{v} + m_2 \mathbf{v} + \dots + m_n \mathbf{v} \\ &= (m_1 + m_2 + \dots + m_n) \mathbf{v} = m \mathbf{v} \end{aligned}$$

where m_1, m_2, \dots, m_n = masses of the separate particles
 $m = m_1 + m_2 + \dots + m_n$ = mass of the body.

Similarly, the kinetic energy of a body in translational motion is of the same form as that of a particle.

But this reasoning is invalid if a rigid body is rotating, because the velocities of its particles differ. Here all the dynamic characteristics should be expressed in terms of the angular velocity. This is precisely why we must introduce, for such phenomena, a number of new physical quantities: moment of force, moment of inertia and angular momentum.

22.2. KINETIC ENERGY AND MOMENT OF INERTIA

1. Let us derive the equation for the kinetic energy of a rigid body rotating about a fixed axis. We shall solve this problem with the approximation of Newtonian mechanics, i.e. under the condition that all the particles of the body move at velocities considerably less than the speed of light in a vacuum. For the sake of simplicity, we shall

first deal with a system of two particles and then generalize the result to suit any rigid body.

2. Let two particles of masses m_1 and m_2 be located at the distance l from each other (Fig. 22.1). We assume the system to be rigid, i.e. the distance between the particles remains constant. The system rotates about an axis with the angular velocity ω . Then, according to

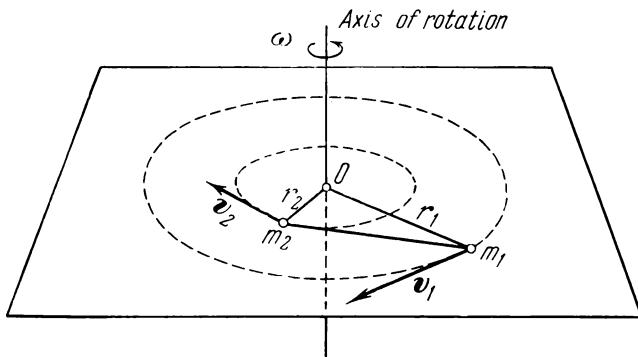


Fig. 22.1

equation (4.12), the velocity of the first point is $v_1 = r_1\omega$ and that of the second is $v_2 = r_2\omega$, where r_1 and r_2 are the distances from the particles to the axis of rotation. The kinetic energies of the particles, according to equation (16.5), are

$$K_1 = \frac{m_1 v_1^2}{2} = \frac{1}{2} m_1 r_1^2 \omega^2 \quad \text{and} \quad K_2 = \frac{m_2 v_2^2}{2} = \frac{1}{2} m_2 r_2^2 \omega^2 \quad (22.1)$$

But the kinetic energy of a system is equal to the sum of the kinetic energies of the particles of which it consists. Thus

$$K = K_1 + K_2 = \frac{1}{2} \omega^2 (m_1 r_1^2 + m_2 r_2^2) \quad (22.2)$$

3. The physical quantity

$$I = m_1 r_1^2 + m_2 r_2^2 \quad (22.3)$$

is called the *rotational inertia*, or *moment of inertia*, of a system of two particles. It characterizes the distribution of the mass of these particles with respect to the axis of rotation.

The SI unit of the moment of inertia is the *kilogram-metre* ($\text{kg}\cdot\text{m}^2$). Substituting into equation (22.2) we obtain

$$K = I \omega^2 / 2 \quad (22.4)$$

Thus, the *kinetic energy of a system of particles equals one half of the product of the moment of inertia of the system by the square of the angular velocity of rotation*.

4. If the system consists of n , and not two, particles, the equation of its kinetic energy remains the same, but that for the moment of inertia becomes

$$I = m_1 r_1^2 + m_2 r_2^2 + \dots + m_n r_n^2 \quad (22.5)$$

Any rigid body can be divided into a sufficiently large number of portions with the masses $\Delta m_1, \Delta m_2, \dots, \Delta m_n$ so that each portion can be taken as a particle. Then the moment of inertia of the whole body is

$$I = \Delta m_1 r_1^2 + \Delta m_2 r_2^2 + \dots + \Delta m_n r_n^2 \quad (22.6)$$

Generally speaking, the result depends upon the way the body is divided into portions and the position of the axis of rotation. An exact result is obtained only when we pass over to the limit with an infinitely large number of divisions. The calculation of the moments of inertia of rigid bodies is a quite difficult mathematical problem, and we shall not go into it here.

22.3. DEPENDENCE OF THE MOMENT OF INERTIA ON THE LOCATION OF THE AXIS OF ROTATION

The moment of inertia of a body depends, not only on its mass, but on the location of its axis of rotation. This follows directly from equation (22.5). Indeed, if we move the axis of rotation to some other position, the masses m_1, m_2, \dots, m_n do not change but the radii r_1, r_2, \dots, r_n do and this changes the moment of inertia.

Let us calculate the moments of inertia of a system of two particles with respect to two axes which are parallel to each other and perpendicular to the plane of the drawing (Fig. 22.2). The distance between axes A and C equals a . Axis C passes through the centre of mass of the system.

The moment of inertia of the system with respect to the axis passing through point A is

$$I = m_1 R_1^2 + m_2 R_2^2 \quad (22.7)$$

The moment of inertia of the same system but with respect to the axis passing through the centre of mass C is

$$I_0 = m_1 r_1^2 + m_2 r^2 \quad (22.8)$$

According to the Pythagorean theorem

$$R_1^2 = h^2 + (r_1 + x)^2 \quad \text{and} \quad R_2^2 = h^2 + (r_2 - x)^2$$

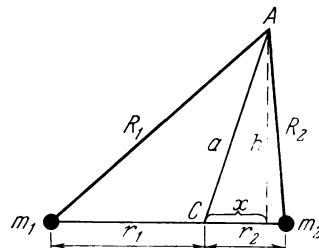


Fig. 22.2

Substituting into equation (22.7), combining with equation (22.8) and after some simple transformations, we obtain

$$I = I_0 + (m_1 + m_2)(h^2 + x^2) + 2x(m_1r_1 - m_2r_2)$$

but $h^2 + x^2 = a^2$ and from the definition of the centre of mass (see Sec. 15.7) it follows that $m_1r_1 - m_2r_2 = 0$. Finally, for nonrelativistic velocities, $m_1 + m_2 = m$ is the mass of the system. Hence

$$I = I_0 + ma^2 \quad (22.9)$$

Thus, the moment of inertia of a system of particles with respect to an arbitrary axis equals the moment of inertia of the system with respect to a parallel axis passing through the centre of mass of the system plus the product of the mass of the system by the square of the distance between the axes (*Steiner's parallel-axis theorem*).

Since ma^2 is a positive value, it follows from the parallel-axis theorem that the moment of inertia of a system has its minimum value about an axis passing through its centre of mass.

2. Let us make use of the parallel-axis theorem to calculate the moment of inertia of a homogeneous bar with respect to an axis perpendicular to the bar and passing through its end (Fig. 22.3). Let the mass of the bar be m and its length l . The moment of inertia is proportional to the mass and to the square of the linear dimensions. Thus

$$I = \theta ml^2$$

where θ is the proportionality factor (an abstract quantity).

First we calculate the moment of inertia with respect to an axis passing through the centre of mass. On the one hand, we can deal with the bar as being made up of two equal pieces with the masses $m_1 = m/2$ and lengths $l_1 = l/2$. Since the moment of inertia of the whole bar equals the sum of the moments of inertia of the two halves, we can write

$$I_0 = 2\theta m_1 l_1^2 = 2\theta \frac{m}{2} \frac{l^2}{4} = \frac{\theta}{4} ml^2$$

On the other hand, according to the parallel-axis theorem,

$$I_0 = I - ma^2 = \theta ml^2 - m \left(\frac{l}{2}\right)^2 = ml^2 \left(\theta - \frac{1}{4}\right)$$

Combining the last two equations we find that

$$\frac{\theta}{4} = \theta - \frac{1}{4}, \text{ from which } \theta = \frac{1}{3}$$

Hence, the moment of inertia of a homogeneous bar with respect to an axis perpendicular to the bar and passing through one end is

$$I = \frac{ml^2}{3} \quad (22.10)$$

The moment of inertia of the same bar with respect to an axis passing through its centre of mass is

$$I_0 = \frac{ml^2}{12} \quad (22.11)$$

3. The moment of inertia of a disk with respect to the axis passing through its centre is proportional to its mass and to the square of its radius. Thus

$$I = \alpha m R^2$$

To determine the proportionality factor α we first find the moment of inertia of the ring shown in Fig. 22.4. It is equal to the difference

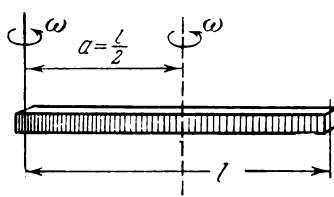


Fig. 22.3

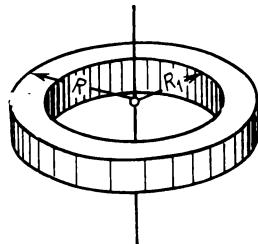


Fig. 22.4

between the moments of inertia of two disks of radii R and R_1 having the masses m and $m_1 = mR_1^2/R^2$. Thus

$$I_{ring} = I - I_1 = \alpha m R^2 - \alpha m_1 R_1^2 = \frac{\alpha m}{R^2} (R^4 - R_1^4)$$

The mass of the ring

$$m_{ring} = m - m_1 = m (R^2 - R_1^2)/R^2$$

If the ring is sufficiently thin ($R_1 \approx R$), its moment of inertia can be determined by equation (22.6). Thus

$$\begin{aligned} I_{ring} &= \Delta m_1 R^2 + \Delta m_2 R^2 + \dots = (\Delta m_1 + \Delta m_2 + \dots) R^2 \\ &= m_{ring} R^2 = m (R^2 - R_1^2) \end{aligned}$$

Combining the two equations for the moment of inertia of a ring, we obtain, after cancelling,

$$\alpha (R^2 + R_1^2) = R^2$$

But then, according to the condition $R_1 \approx R$, $\alpha = 1/2$.

Thus, the moment of inertia of a disk with respect to the axis passing through its centre equals

$$I = \frac{1}{2} m R^2 \quad (22.12)$$

22.4. MOMENT OF FORCE

1. Suppose that a certain force F is acting on a body that is rotating about a fixed axis (Fig. 22.5). Let us calculate the work done by this force as the body rotates through an angle $\Delta\alpha$. According to definition [equation (16.12)] an element of work equals

$$\Delta W = F_t \Delta l$$

But $F_t = F \sin \gamma$, where γ is the angle between the direction (line of action) of the force and the radius from the axis of rotation to

the point where the force is applied. The length of the arc $\Delta l = r \Delta\alpha$. Hence

$$\Delta W = Fr \sin \gamma \times \Delta\alpha \quad (22.13)$$

2. The quantity

$$M = Fr \sin \gamma = Fd \quad (22.14)$$

is called the *moment of force* with respect to the axis of rotation. The shortest distance $d = r \sin \gamma$ from the axis of rotation to the line of action of the force is called the *moment arm*.

The *moment of force with respect to a certain axis is the product of the force by its moment arm.*

The SI unit for the moment of force is the *newton-metre (N·m)*.

3. Substituting equation (22.14) into equation (22.13) for the element of work, we obtain

$$\Delta W = M \Delta\alpha \quad (22.15)$$

i.e. the *work done by a force in rotating a body equals the product of the moment of the force by the angle of rotation.*

4. To calculate the power, equation (22.15) should be divided by the interval of time Δt required for the body to rotate through the angle $\Delta\alpha$. Since $\Delta W / \Delta t = P_{av}$ is the average power, and $\Delta\alpha / \Delta t = \omega_{av}$ is the average angular velocity, we can write

$$P_{av} = M \omega_{av}$$

or, in the limit when $\Delta t \rightarrow 0$, we obtain for the instantaneous power

$$N = M\omega \quad (22.16)$$

The *instantaneous power equals the product of the moment of the force by the angular velocity of rotation.*

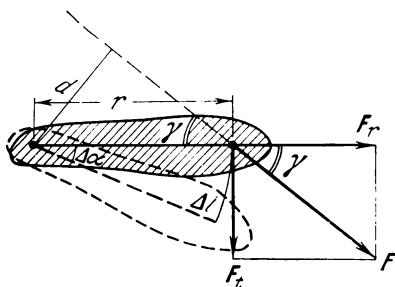


Fig. 22.5

22.5. EQUILIBRIUM CONDITIONS FOR A BODY HAVING AN AXIS OF ROTATION

1. In Sec. 19.6 we derived the equilibrium conditions for a body free to move in any direction. Equilibrium is set up for such bodies if the vector sum of all the forces applied to the body (their resultant) equals zero. Let us find the conditions for the equilibrium of a body which can rotate about a certain given axis.

We shall take moments to be positive if they rotate the body counterclockwise and negative when they rotate it clockwise. Thus in Fig. 22.6, the moments of forces F_1 and F_2 are negative, and the moment of the force F_3 is positive if the axis of rotation passes through point O .

2. If the body is in a state of equilibrium, its potential energy is not changed when it is turned through an infinitely small angle $\Delta\alpha$. Therefore, an element of work ΔW , equal to the change in the potential energy, is equal to zero. Thus $\Delta W = 0$. But

$$\begin{aligned}\Delta W &= \Delta W_1 + \Delta W_2 + \Delta W_3 = M_1 \Delta\alpha + M_2 \Delta\alpha + M_3 \Delta\alpha \\ &= (M_1 + M_2 + M_3) \Delta\alpha\end{aligned}$$

Since $\Delta\alpha \neq 0$, it is evident that

$$M_1 + M_2 + M_3 = 0 \quad (22.17)$$

This is the required condition of equilibrium. Hence, a body having a given axis of rotation is in a state of equilibrium if the algebraic sum of all the moments of force with respect to this axis equals zero.

In other words, equilibrium is set up if the sum of the moments of force tending to turn the body counterclockwise is equal to the sum of the moments tending to turn it clockwise.

3. It is evident from this condition that a small force having a large moment arm can balance a large force having a small moment arm. This property of moments of force finds them practical application for levers, capstans, screw jacks and many other devices that are widely employed in engineering.

22.6. ANGULAR MOMENTUM AND THE FUNDAMENTAL LAW OF DYNAMICS

1. Let several forces be applied to a body rotating about a fixed axis (Fig. 22.6). We shall denote the algebraic sum of the moments of all the forces by M . It was found that the angular velocity of rotation

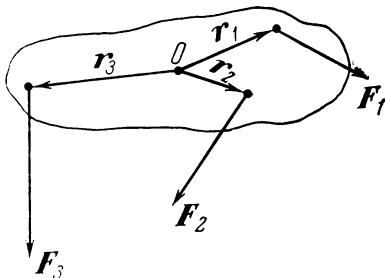


Fig. 22.6

of the body is changed by the action of the total moment of force. This also changes the kinetic energy of the body. If at the instant of time t_1 the body has an angular velocity ω_1 and kinetic energy $K_1 = I\omega_1^2/2$, then at instant t_2 its angular velocity will be ω_2 and its kinetic energy $K_2 = I\omega_2^2/2$. The change in the kinetic energy is

$$\begin{aligned}\Delta K &= K_2 - K_1 = \frac{I\omega_2^2}{2} - \frac{I\omega_1^2}{2} \\ &= \frac{I}{2} (\omega_2 - \omega_1) (\omega_2 + \omega_1) \\ &= (I\omega_{av}) \omega_{av}\end{aligned}$$

where $\omega_{av} = (\omega_1 + \omega_2)/2$ is the average angular velocity. The work done by the total moment of force is expressed, according to equation (22.15), by the formula

$$\Delta W = M\Delta\alpha = M\omega_{av}\Delta t$$

But work is a measure of the change in kinetic energy. Therefore

$$M\omega_{av}\Delta t = (I\omega_2 - I\omega_1) \omega_{av}$$

from which

$$M\Delta t = I\omega_2 - I\omega_1 \quad (22.18)$$

2. The physical quantity equal to the product of the moment of inertia by the angular velocity is called the *angular momentum* (also called the *moment of momentum*):

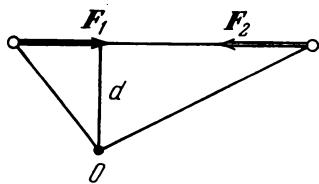


Fig. 22.7

$$L = I\omega \quad (22.19)$$

Note that for a particle moving in a circle of radius r the angular momentum equals the product of the linear momentum by the radius. Thus

$$L = I\omega = mr^2\omega = mrv = pr \quad (22.20)$$

The SI unit of the angular momentum is the *kilogram-metre² per second* ($\text{kg}\cdot\text{m}^2/\text{s}$).

3. Dividing both sides of equation (22.18) by Δt and taking into consideration the fact that

$$I\omega_2 - I\omega_1 = L_2 - L_1 = \Delta L$$

we obtain

$$M = \frac{\Delta L}{\Delta t} \quad (22.21)$$

This is the *fundamental law of dynamics for a rotating rigid body*: the moment of force is equal to the change in the angular momentum per unit time.

4. Note that equation (22.21) concerns only the sum of the moments of the *external forces*. As to the internal forces, the sum of their moments always equals zero. This follows directly from Newton's third law. As proof let us consider a system of two interacting particles (Fig. 22.7). Force $F_2 = -F_1$, and both forces have the same moment arm d . Therefore, $M_1 = F_1d$, $M_2 = F_2d = -F_1d = -M_1$ and the total moment of internal forces is

$$M_{int} = M_1 + M_2 = 0$$

22.7. THE LAW OF CONSERVATION OF ANGULAR MOMENTUM

1. Let the algebraic sum of the moments of the external forces applied to a body equal zero. Among others, this condition always holds true for a closed system to which no external forces are applied. Then it follows from equation (22.21) that

$$\Delta L = 0 \quad \text{and} \quad L = \text{const} \quad (22.22)$$

This result is of great significance and is called the *law of conservation of angular momentum*:

The angular momentum of a closed system remains constant.

2. An interesting consequence stems from the difference between the concepts of linear momentum and angular momentum.

It was shown in Sec. 15.8 that the velocity of the centre of mass of a system of particles cannot be changed by internal forces. In the translational motion of a body, the velocities of all its particles coincide with the velocity of its centre of mass. Consequently, internal forces are incapable of changing the velocity of a body in translational motion.

But the result is entirely different when a body rotates about an axis. Internal forces can change the distances between the separate portions of the body, thereby changing its moment of inertia. The law of conservation of angular momentum, however, asserts that only the product $L = I\omega$ is constant, and not each of its factors. If, due to internal forces, the moment of inertia is reduced, then the angular velocity increases in the same proportion so that the product $I\omega$ remains constant.

3. The conclusion reached here has many practical applications.

Let us consider, for instance, how a tumbler performs a back somersault (Fig. 22.8). After squatting with his arms stretched backward, the acrobat springs upward, throwing his arms forward, then upward and slightly back. This imparts a low velocity of rotation to the tumbler's body with respect to an axis passing through his centre of mass (approximately at the level of his waist). Near the top point of his path, he quickly draws his knees up toward his chin, holding them with his hands. Thus pulling in his arms and legs toward his centre of mass, he substantially reduces his moment of

inertia, thereby increasing his velocity of rotation and enabling him to rapidly complete the somersault. As he lands the tumbler straightens out and throws his arms upward and slightly forward, slowing down his velocity of rotation.

Ice skaters use the same technique in performing a spin about a vertical axis. Stretching his arms to the sides, and straightening out and raising one leg, the skater begins to rotate slowly. Pulling in

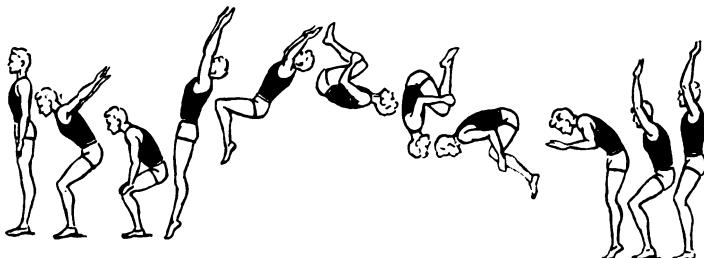


Fig. 22.8

his arms and leg sharply, the skater reduces his moment of inertia and begins to spin at high speed.

4. It can be readily seen that the kinetic energy of rotation is not conserved in the above examples. Let the kinetic energy at the beginning be $K = I\omega^2/2$. After pulling in his arms and legs, the tumbler's moment of inertia is reduced n -fold: $I_1 = I/n$. This means that the angular velocity is increased n -fold: from $I\omega = I_1\omega_1$ it follows that $\omega_1 = n\omega$. Then the kinetic energy becomes

$$K_1 = \frac{I_1\omega_1^2}{2} = \frac{I\omega^2 n^2}{2n} = nK$$

The increase in kinetic energy is $\Delta K = K_1 - K = K(n - 1)$. But to increase the kinetic energy work must be done. What are the forces that do this work?

Since, as a first approximation, the systems dealt with here are closed, external forces can be neglected. Hence, in the given cases (with the tumbler and ice skater) rotational kinetic energy increases due to the work done by internal forces, i.e. in the final analysis, at the expense of the internal energy of the rotating system. The increase in kinetic energy is the result of the work done by the tumbler or skater in pulling in his arms and legs. This conclusion, of course, fully agrees with the law of conservation of energy.

5. It is a consequence of the law of universal gravitation that *each of the planets moves about the sun in an elliptical orbit, having the sun as one focus* (Kepler's first law). According to the law of conservation

of energy (Sec. 19.3)

$$\frac{mv^2}{2} - G \frac{Mm}{r} = \text{const}$$

It is quite evident that the orbital velocity of the planets continuously varies. The velocity of a planet is reduced with an increase of its distance from the sun (Fig. 22.9).

On the basis of many years of observation, Kepler established that the *radius vector of a planet sweeps out equal areas in equal intervals of time* (Kepler's second law). Let us show that this law follows from the law of conservation of angular momentum.

The sun and planets, we know, form a closed system for which the law of conservation of angular momentum is valid. The centre of mass of this system practically coincides with the centre of the sun. Therefore, the velocity of the sun is practically equal to zero. According to equation (22.22), the angular momentum of the system is

$$L = mvr = \text{const} \quad (22.23)$$

The area swept out by the radius vector of a planet in unit time is

$$\sigma = \frac{A}{t} = \frac{1}{2} vr = \frac{L}{2m} = \text{const} \quad (22.24)$$

And this equation represents Kepler's second law.

22.8. ANALOGIES BETWEEN QUANTITIES AND THEIR RELATIONS IN TRANSLATIONAL AND ROTATIONAL MOTION

1. If we compare the relationships between the quantities that characterize the motion of a particle (or a body in translational motion) with the same relationships that describe the rotation of a body about an axis, we will be struck by the deep-seated analogy between these relationships. Let us set down a series of quantities and the equations

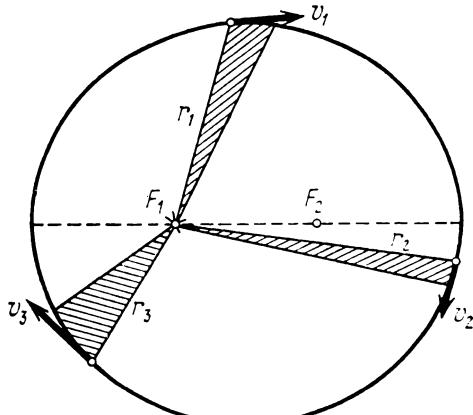


Fig. 22.9

that relate them in a table, restricting ourselves to Newtonian mechanics (Table 22.1).

Table 22.1

Translational Motion	Rotational Motion
Displacement Δl	Angle of rotation $\Delta\alpha$
Velocity v	Angular velocity ω
Mass m	Moment of inertia I
Linear momentum $p = mv$	Angular momentum $L = I\omega$
Force F	Moment of force M
Fundamental equation of dynamics $F = \frac{\Delta p}{\Delta t}$	Fundamental equation of dynamics $M = \frac{\Delta L}{\Delta t}$
Kinetic energy $K = \frac{mv^2}{2} = \frac{p^2}{2m}$	Kinetic energy $K = \frac{I\omega^2}{2} = \frac{L^2}{2I}$
Element of work $\Delta W = F_t \Delta l$	Element of work $\Delta W = M \Delta \alpha$
Power $P = F_t v$	Power $P = M \omega$

It is evident that a simple substitution of analogical quantities in a formula for translational motion will automatically lead to a correct formula for rotational motion. Thus, in the formula $K = p^2/2m$, if we replace the linear momentum p by the analogical angular momentum L , and the mass m by its analogue, the moment of inertia I , we obtain the correct formula for rotational motion: $K = L^2/2I$. In exactly the same way, by analogy, we can write the equation for the work done in winding a spring through a certain angle φ . In Sec. 18.2 we obtained the expression $W = Fl/2$ from the graph as the work done to stretch a spring. Replacing the force by its analogue, the moment of force, and the displacement by the angle of rotation, we obtain $W = M\varphi/2$. This formula could have been derived independently, of course, by means of a graph, but it can be obtained more quickly and simply by analogy.

It will be left as a problem for the student, after introducing the concept of angular acceleration $\beta = \Delta\omega/\Delta t$, to obtain, by analogy, the expressions for the instantaneous angular velocity and for the angle of rotation in uniform rotation, as well as the expression for the fundamental equation of dynamics for a rotating body.

CHAPTER 23

SYMMETRY IN NATURE AND THE CONSERVATION LAWS

23.1. THE CONSERVATION LAWS ARE NATURE'S PRINCIPAL LAWS

1. Three principal laws of nature—the laws of conservation of linear momentum, angular momentum and energy—were dealt with in the preceding chapters. It should be noted that these laws are valid only in inertial frames of reference. Indeed, our derivation of these laws was based on Newton's second and third laws which hold only for inertial reference frames.

Let us recall, also, that linear and angular momentum are conserved only in closed systems (with the sum of all the external forces and respective moments of force equal to zero). Even the fact that it is a closed system is insufficient for the conservation of the energy of a body; it must also be adiabatically isolated (i.e. it must not participate in heat exchange).

2. The conservation laws are valuable for their generality. Neither the details of the specific physical process concerned, nor the features of the various bodies participating in the process, have any influence on the validity of these laws. They are just as applicable to the planets and stars as they are to molecules, atoms and the elementary particles of which atomic nuclei consist. They can be employed in investigating thermal, chemical, electrical and any other processes.

3. During the whole history of the development of physics, the conservation laws were almost the only ones that retained their significance when certain theories were superseded by newer ones.

In relativity theory, for instance, the classical concepts of time and the simultaneity of events, of length and the mass of bodies, undergo essential changes. New statements of the law of addition of velocities and Newton's second law are introduced, the expression for the kinetic energy is changed, etc. The laws of the conservation of linear momentum, angular momentum and energy, however, turn out to be entirely valid in relativity theory. And what is more, two laws of Newtonian mechanics—the law of conservation of mass and the law of conservation of energy—are united in relativity theory into a unified law of conservation of mass and energy.

In the microcosm—the world of atoms, molecules and elementary particles—we find many specific phenomena that sharply differ from the phenomena in the world of large bodies. Owing to the principle of uncertainty, it is impossible here to solve the basic problem of dynamics and to determine uniquely, for example, the path of an

electron in an atom. An entirely different approach is required to phenomena of the microcosm. However, here again the conservation laws remain valid, and in quantum mechanics they play even a more important role than in Newtonian or relativistic mechanics. Frequently, the conservation laws turn out to be the only means of explaining certain phenomena occurring in the microcosm.

It is precisely the *universality of the conservation laws* that makes them especially valuable.

23.2. THE CONSERVATION LAWS ARE FORBIDDENNESS PRINCIPLES

1. A basic difference exists between laws of the type of the fundamental law of dynamics and the laws of conservation. The laws of dynamics give one an idea of the detailed course of the process. If, for instance, we are given the force applied to a particle and the initial conditions, we can find the motion equation, path, magnitude and direction of the velocity at any instant of time, etc. (see Ch. 8). The conservation laws give us no direct indications on the course of the process being considered. They only specify which processes are forbidden and therefore do not occur in nature.

Hence, the conservation laws are manifested as *forbiddenness principles*: if only one of the conservation laws is not complied with by any phenomenon, it is forbidden, and such phenomena are never observed in nature. Any phenomenon that does not violate one of the conservation laws can, in principle, occur. We should note that it is the allowed phenomena that always occur in practice, but with different degrees of probability. Certain of these phenomena occur very frequently, others more rarely, but they can, nevertheless, be observed.

It may seem that the conservation laws leave too much to chance; that there are too many possible versions of what can take place. It is not clear to us, therefore, why, in most cases, only single process is realized in an experiment. It turns out, actually, that the combined effect of several conservation laws often determines, almost uniquely, the possible course of the process.

2. Let us consider the following example. Can a body at rest begin to move at the expense of its internal energy? This process is not inconsistent with the law of conservation of energy. It is only necessary for the generated kinetic energy to be exactly equal to the reduction in internal energy.

As a matter of fact, this process never occurs, because it is inconsistent with the law of conservation of momentum. If the body is at rest its momentum is equal to zero. If it began to move, its momentum would increase by itself, and this is impossible. Therefore, the internal energy of a body cannot change into kinetic energy unless the body breaks up into pieces.

If we assume that the body can break up into pieces, then the forbiddenss imposed by the law of conservation of momentum is removed. The fragments of the body can then move so that their centre of mass remains at rest, and this is all that is required by the law of conservation of linear momentum.

Thus, for the internal energy of a body to change into kinetic energy, the body must be capable of breaking up into pieces. If, however, there is still some law that forbids the break-up of the body, then its internal energy (and rest mass) will be constant.

3. Incidentally, it is precisely the combined effect of the laws of conservation of energy and momentum that can explain why, in producing antiprotons by the collision of fast protons with stationary ones (see Sec. 21.7), it is necessary to impart a kinetic energy of $6E_0 \cong 6$ GeV to the bombarding particles, though an energy of only $2E_0$ is sufficient to produce two new particles—a proton and an antiproton.

23.3. THE CONSERVATION LAWS AND SPACE-TIME SYMMETRY

1. In demonstrating the universal nature of the conservation laws, we arrived at a certain logical inconsistency with the reasoning on which the derivation of these laws was based. Indeed, we obtained the laws of conservation as consequences of Newton's second and third laws. But these laws were established by Newton when he generalized the results of experiments dealing with elastic, gravitational and Coulomb interactions. These interactions are insufficient, however, to describe all phenomena in nature.

Two more kinds of interaction are known today. *Strong interactions* are due to nuclear forces acting between elementary particles within the nucleus, as well as to processes in which heavy elementary particles are produced at very high energies, for example, the proton-antiproton pair creation reaction discussed in Sec. 21.7. *Weak interaction* results in the decay of certain elementary particles, for instance, nuclear beta decay (see Secs. 81.12 and 83.5).

2. It is not at all evident that Newton's laws are valid for these interactions, the more so because the uncertainty principle begins to play a decisive role in them. However, the conservation laws are fully applicable to weak interaction, as is attested to by all known phenomena.

We find, in this way, that the conservation laws are much more universal than Newton's laws from which we derived them. It would seem expedient, therefore, to find certain more general principles on which the conservation laws could be based.

3. It was found that such principles actually exist. They are the *principles of space-time symmetry*. It was shown that the law of conservation of linear momentum follows from the uniformity of space,

the law of conservation of angular momentum from the isotropy of space, and the law of conservation of energy from the uniformity of time.

The fact that the conservation laws are derived from such extremely general concepts is the strongest confirmation of the universal significance of these laws.

23.4. UNIFORMITY OF TIME AND THE CONSERVATION OF ENERGY

As an illustration of the relationship between the properties of space-time symmetry and the conservation laws, let us consider some elementary factors that explain the connection between the law of conservation of energy and the uniformity of time.

1. The uniformity of time, discussed to some extent in Sec. 1.2, is simply an expression of the independence of natural phenomena on the instant of time we choose as the initial one. It follows, in particular, that the rest mass of a closed and adiabatically isolated system is independent of time.

Note that one condition of adiabatic isolation, playing no role in analyzing the properties of space, plays an essential role here, because the rest mass changes in the course of time if heat exchange occurs.

2. If we write equation (16.8) for two instants of time at which we consider a closed and adiabatically isolated system, then, since $m_0 = \text{const}$, we again obtain an equation that coincides with (16.10):

$$\Delta E = u \Delta p \quad (23.1)$$

Since the system is closed, the external force $F = 0$ and therefore $\Delta p = F \Delta t = 0$. Under these conditions, it follows from equation (23.1) that $\Delta E = 0$.

Hence, it follows from the uniformity of time that $\Delta E = 0$ for a closed and adiabatically isolated system, or $E = \text{const}$, i.e. the law of conservation of energy. A more rigid proof would require a much more complex mathematical apparatus than our students probably have at their disposal.

Chapter 24

NONINERTIAL FRAMES OF REFERENCE AND GRAVITATION

24.1. PHENOMENA IN AN ACCELERATED REFERENCE FRAME

1. In the preceding chapters we described phenomena that occurred in inertial reference frames. Let us consider now how bodies act in reference frames that move with acceleration with respect to an iner-

tial frame. For the sake of simplicity, we shall restrict ourselves to the special case in which the noninertial frame moves with respect to the inertial reference frame in a *straight line*, at constant acceleration $w = \text{const}$ and during short time intervals, so that the velocity of the noninertial frame is $v = wt \ll c$.

Notwithstanding the restricted formulation of the problem, we shall obtain, and with the most elementary means, all the laws that have any fundamental significance.

2. Let us imagine that we have a sufficiently long railway car, with a horizontal bar fastened lengthwise. A heavy ball (Fig. 24.1) can

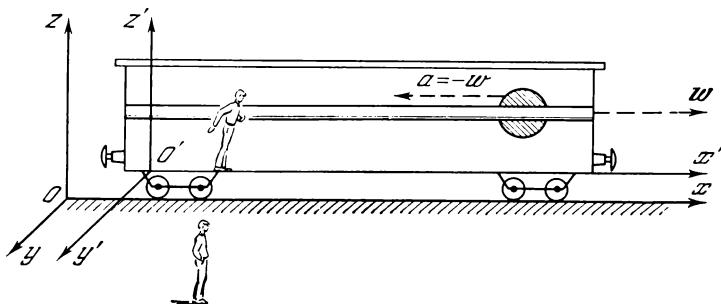


Fig. 24.1

slide without friction along the bar. The frame of reference xyz , attached to the earth, is inertial, as has been mentioned previously, to a quite sufficient extent. Let us consider how phenomena, occurring in the car, are described in reference frame $x'y'z'$, attached to the car.

It has been found that when the car is moving with an acceleration w with respect to the earth, the ball moves along the bar with the acceleration $a = -w$ with respect to the car. Let us describe this phenomenon from the points of view of the two reference frames.

Frame xyz

No forces are applied to the ball and, according to the *law of inertia*, its *velocity does not change*, i.e. it moves without acceleration with respect to the earth. But the car moves with acceleration w with respect to the earth and therefore, *with respect to the car*, the ball moves with the acceleration

$$a = -w$$

Frame $x'y'z'$

No forces are applied to the ball, but still it has the acceleration a with respect to the car. This means that the law of inertia is violated in a reference frame attached to the car: there is acceleration but it is not caused by any forces. The reason is that the car moves with respect to the inertial frame with the acceleration

$$w = -a$$

Thus, the law of inertia is violated in an accelerated frame of reference. For this reason such a frame is said to be *noninertial*.

3. Let us alter the experiment somewhat by attaching the ball to a spring whose other end is fastened to the front wall of the car (Fig. 24.2). While the car is at rest with respect to the earth, or moves

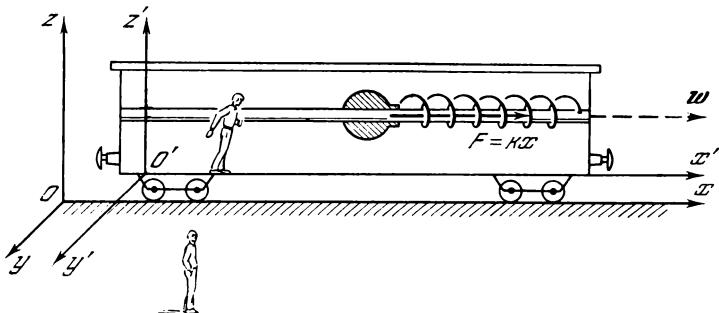


Fig. 24.2

in a straight line with uniform velocity, the spring is not stretched. If the car begins to move with acceleration, the spring stretches and will remain stretched as long as the car is accelerated. The ball will be at rest with respect to the car. Let us also consider this case from the points of view of the two reference frames.

Frame xyz (inertial)

The ball is at rest with respect to the car. Hence, it moves together with the car at an acceleration w with respect to the earth. But, according to Newton's second law, the acceleration is caused by the force $F = mw$. This force is applied to the ball by the stretched spring. Thus

$$F = kx = mw$$

Frame $x'y'z'$ (noninertial)

The ball is at rest with respect to the car even though the force $F = kx$ is applied to it by the stretched spring. Hence, Newton's second law is violated in a reference frame attached to the car. The point is that the car moves with respect to an inertial reference frame with the acceleration

$$w = F/m = kx/m$$

24.2. INERTIAL FORCES

1. As we could see in the preceding section, Newton's laws do not hold in noninertial reference frames. But we can formulate a rule that enables us to *formally* apply Newton's second law (the fundamental equation of dynamics) in noninertial frames as well.

For this purpose let us return to the analysis of the phenomena occurring in the car moving with acceleration (see Figs. 24.1 and 24.2). In the first experiment, when the ball did not interact with

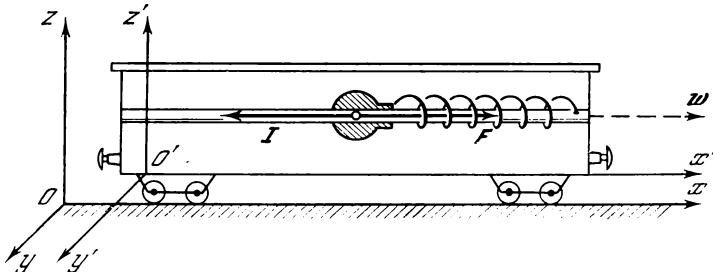


Fig. 24.3

any other bodies, it nevertheless moved with respect to the car with the acceleration $a = -w$. Therefore, the ball acted as if a certain force

$$I = ma = -mw$$

acted on it and imparted the acceleration.

In the second experiment, the stretched spring acts on the ball with the force $F = -kx$. But this force does not impart any acceleration to the ball with respect to the car. It would seem that a certain force

$$I = ma = -mw$$

is applied to the ball, and that it counterbalances force F (Fig. 24.3). The quantity

$$I = -mw$$

where w is the acceleration of the noninertial reference frame (with respect to the inertial frame), is called the *inertial force*.

2. Making use of this concept we can write the fundamental law of dynamics for noninertial reference frames as follows

$$R + I = ma \quad (24.1)$$

where R = resultant of all forces of interaction

I = inertial force

a = acceleration of the body with respect to the noninertial reference frame.

Equation (24.1) is a generalization of the fundamental equation of dynamics for the case of noninertial reference frames:

The vector sum of all the forces of interaction and the inertial force is equal to the product of the mass of the body by its acceleration with respect to the noninertial reference frame. We shall demonstrate the use of this principle by an example.

3. Assume that a certain body of mass m is suspended by a string from the ceiling of a railway car moving with acceleration with respect to the earth. It is found that as long as the car moves with acceleration, the string makes a certain angle α with the vertical. Let us calculate this angle for both inertial and noninertial frames of reference.

Inertial frame

Since the body is at rest with respect to the car, it is moving together with the car at an acceleration w with respect to the earth. This acceleration is imparted to the body by the resultant of two forces: the force of gravity and the tension of the string (Fig. 24.4).

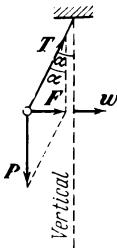


Fig. 24.4

According to the fundamental equation of dynamics

$$\mathbf{F} = \mathbf{P} + \mathbf{T} = mw$$

But $F/P = \tan \alpha$. Substituting the values $F = mw$ and $P = mg$, we obtain

$$\tan \alpha = mw/mg = w/g$$

Noninertial frame

Since the body is at rest with respect to the car, the vector sum of all the forces, including the forces of interaction and the inertial force, equals zero. Consequently, the force of gravity, tension of the string and the inertial force form a closed triangle (Fig. 24.5).

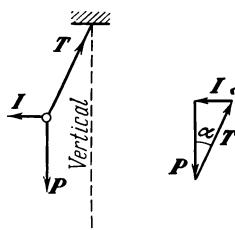


Fig. 24.5

From the drawing we can write: $\tan \alpha = I/P$. Substituting $I = mw$ and $P = mg$, we obtain

$$\tan \alpha = mw/mg = w/g$$

24.3. FEATURES OF INERTIAL FORCES

1. Inertial forces possess some features that distinguish them from forces of interaction: elastic, electrical, gravitational and frictional forces.

The most important of these features are:

(a) An inertial force is produced by the accelerated motion of the reference frame itself, and not by the interaction of bodies. This is why Newton's third law is not applicable to inertial forces.

(b) An inertial force acts on a body only in a noninertial frame of reference. There are no such forces in inertial frames.

(c) Inertial forces are external forces with respect to any system of bodies in a noninertial frame. Hence, there are no closed systems in such frames and, consequently, the conservation laws do not hold (this is discussed in more detail in Sec. 24.4).

(d) Finally, it should be noted that an inertial force, like a gravitational force, is proportional to the mass of the body. Therefore, in an inertial force field as in a gravitational field all bodies move with the same acceleration whatever their mass (see Sec. 24.5 for more detail).

2. There are no inertial forces at all in inertial reference frames, and therefore it is a crude error to attempt to use them there. In inertial frames, the term "force" is used only in the Newtonian sense, as a measure of the interaction of bodies.

In noninertial frames of reference, however, inertial forces act in the same way as forces of interaction. And if the inertial force flings you forward when a railway car is suddenly braked, you will feel the force just as really as you feel the force of gravity. In exactly the same way, an astronaut feels the inertial force pushing him against his cradle in an accelerating rocket at takeoff.

24.4. SPACE AND TIME IN NONINERTIAL REFERENCE FRAMES

1. We have called attention several times to the fact that time is uniform and space is both uniform and isotropic in inertial frames of reference. The geometrical properties of uniform and isotropic space can be described by Euclidean geometry.

One of the basic propositions of Euclidean geometry is the fifth postulate which can be stated as follows: through a point that does not lie on a given straight line we can draw one, and only one, straight line parallel to the given one. From this postulate it follows that the sum of the angles of any triangle equals π radians (180°). Let us recall, finally, that in Euclidean geometry the shortest distance between two points is along a straight line.

To simplify further discussion, we shall restrict ourselves to the analysis of geometrical relationships on a plane, for example xy .

2. In uniform and isotropic space, the length of line segments do not depend upon the region of space they are in. Let us divide the axes of coordinates into equal segments $\Delta x = \Delta y = l$ and draw straight lines, parallel to the axes, through the points of division. Plane xy is thus divided into unit cells (Fig. 24.6) in the form of equal squares.

In exactly the same way, owing to the uniformity of time in an inertial reference frame, the interval of time Δt between two events is independent of the point of space at which these events occur.

3. Space is nonuniform in a noninertial reference frame. We know that the length of a line segment is less in a moving frame than in one in which the line segment is at rest (see Sec. 12.8). Thus

$$\Delta x' = \Delta x \sqrt{1 - v^2/c^2} \quad (24.2)$$

But in motion with constant acceleration (uniformly accelerated motion) $v^2 = 2wx$, where w is the acceleration of the noninertial frame. Substituting into equation (24.2) we obtain

$$\Delta x' = \Delta x \sqrt{1 - 2wx/c^2} \quad (24.3)$$

We see then that in noninertial reference frames, the length of a line segment depends upon the region of space it is located in. The length of the same line segment differs at points with different abscissas.

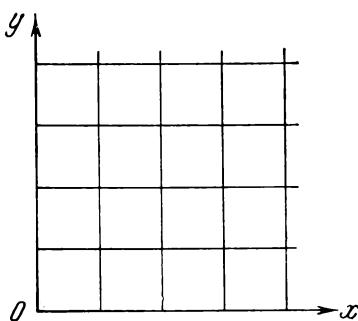


Fig. 24.6

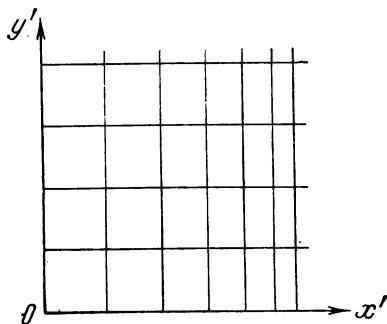


Fig. 24.7

But a line segment along the axis of ordinates retains unchanged length because (in our case) there is no motion along this axis: $\Delta y' = \Delta y$. If we construct a coordinate network in a noninertial frame, we shall see that here plane $x'y'$ is divided into unit cells of oblong shape whose width decreases as we move along the axis of abscissas (Fig. 24.7).

4. Space is not only nonuniform in a noninertial reference frame but is anisotropic.

Indeed, here the two directions are not equivalent along the axis of abscissas. In our example, elements of length decrease along the positive direction of the axis and increase along the negative direction. They remain constant along the axis of ordinates. Consequently, we can distinguish different directions in this space with the aid of physical experiments. This is witness to its anisotropy.

It should be noted that the laws of conservation of linear and angular momentum do not hold in noninertial reference frames due to the nonuniformity and anisotropy of space in these frames. This was mentioned earlier.

5. Finally, it can also be shown that time is nonuniform in noninertial reference frames and, as a result, the law of conservation of energy does not hold for such frames.

In a moving reference frame, an interval of time between two events occurring at the same point (see Sec. 12.9) is

$$\Delta t' = \frac{\Delta t}{\sqrt{1 - v^2/c^2}} \quad (24.4)$$

Making use, as above, of the expression $v^2 = 2wx$, we obtain

$$\Delta t' = \frac{\Delta t}{\sqrt{1 - 2wx/c^2}} \quad (24.5)$$

6. Now, let us show that the nonuniformity and anisotropy of space in noninertial reference frames make it necessary to describe properties of such space by means of a *non-Euclidean geometry*.

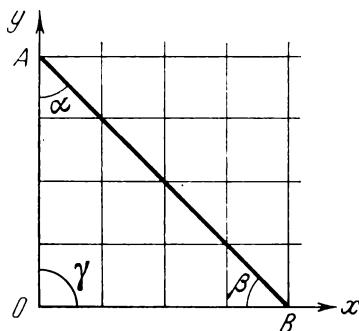


Fig. 24.8

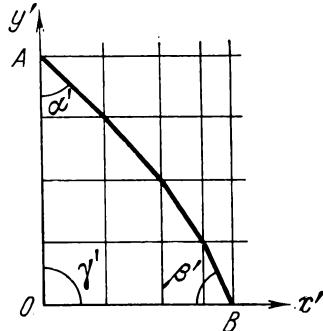


Fig. 24.9

For our proof let us see what a straight line segment, not parallel to the axes of the coordinates, looks like in the given space. First we construct the line segment AB , inclined at an angle of 45° to the axes of coordinates, in an inertial reference frame (Fig. 24.8). In a noninertial frame this segment is converted into a broken line (Fig. 24.9) and, as a matter of fact, if the length element $\Delta x'$ is infinitesimal, into a curve. But then the sum of the angles of the curvilinear triangle thus obtained turns out to be greater than 180° ($\alpha' + \beta' + \gamma' > 180^\circ$). This indicates that geometry is non-Euclidean in the given space.

We should note also that here the shortest distance between points A and B is not along a straight line, but along the arc of some curve.

This also distinguishes the geometry of the given space from the Euclidean type.

7. The possibility that non-Euclidean geometries exist was first mentioned in a hypothesis advanced in 1826 by Nikolai Ivanovich Lobachevsky (1793-1856). He showed that it was possible to devise a consistent geometry in which Euclid's fifth postulate does not hold. This was the basis for a number of subsequent works on non-Euclidean geometries. Lobachevsky also advanced the idea that the geometry of the real world may not be Euclidean.

Contemporaries of Lobachevsky could not comprehend the profoundness of the ideas on which his works were grounded, and many of them simply did not take his ideas seriously. And no wonder. Euclid's geometry, which had been in use for over two thousand years, seemed to be an unshakable pillar of science. Any attempts to devise a non-Euclidean geometry looked like downright nonsense that is at a variance with "common sense". Incidentally, somewhat later, many scientists received the theory of relativity in exactly the same way.

The truth was found, however, in the bold revolutionary ideas of Lobachevsky and Einstein, and not in the "common sense" of conservative scientists. The foresight of Lobachevsky concerning the non-Euclidean nature of the world surrounding us was fully confirmed in the general theory of relativity.

Space whose properties are described by non-Euclidean geometry is often called "curved" space. This term originates from the fact that in a non-Euclidean space the shortest distance between two points is not a straight line, but a certain curve called a *geodesic*. Examples of geodesics are the arc of a great circle on the surface of a sphere and a helix on the surface of a circular cylinder.

24.5. THE PRINCIPLE OF EQUIVALENCE

1. Inertial forces, like forces of gravitation, are proportional to the masses of the bodies to which these forces are applied. Therefore, in an inertial force field, as in a gravitational field, all bodies move with the same acceleration, whatever their mass.

Thus, *in their manifestations, inertial forces are indistinguishable from gravitational forces*. To illustrate this proposition, let us analyze again the phenomenon described in para. 3, Sec. 24.1. Here the ball is subject to the force $F = -kx$ of the elastically stretched spring, but this force does not impart acceleration to the ball with respect to the car. Such a situation is possible only if some other force, in addition to the elastic force, is acting on the ball, and if this second force is equal in magnitude and opposite in direction to the elastic force. However, no other bodies in the car (except the spring) interact with the ball. To what conclusions does this lead us?

First conclusion

Evidently, the car is a noninertial reference frame, and the elastic force is counterbalanced by the inertial force: $F = -I$. To make sure that it really is the inertial force, bodies of various masses are attached to the spring. Since the inertial force is proportional to the mass, the extension of the spring is also proportional to the mass:

$$x = \frac{-F}{k} = \frac{I}{k} = \frac{mw}{k} = \xi m$$

where $\xi = w/k = \text{const}$

Second conclusion

Evidently, the car is an inertial reference frame, but is located in a gravitational field, and the elastic force is counterbalanced by the gravitational force: $F = -P$. To make sure that it really is the gravitational force, bodies of various masses are attached to the spring. Since the gravitational force is proportional to the mass, the extension of the spring is also proportional to the mass:

$$x = \frac{-F}{k} = \frac{P}{k} = \frac{mg}{k} = \xi m$$

where $\xi = g/k = \text{const}$

As we have found, phenomena in an inertial reference frame located in a uniform gravitational field, and those in a noninertial frame moving with an acceleration that is constant in magnitude and direction, occur in exactly the same way. This proposition was first formulated by

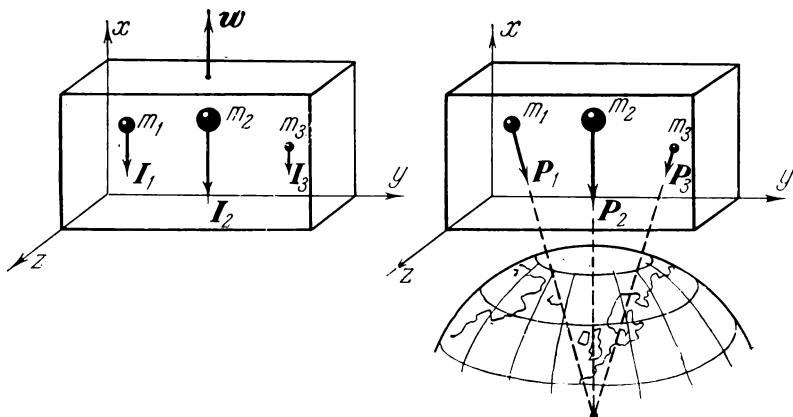


Fig. 24.10.

Einstein and used as the basis for his relativistic theory of gravity (see Sec. 24.6). Einstein called it the *principle of equivalence*.

Special attention should be paid to the precise statement of the principle of equivalence. It holds only for small regions of space: so small that within these regions the gravitational field can be assumed uniform. But in large regions of space, where the nonuniformity

of the gravitational field is clearly manifested, it is impossible to find a noninertial reference frame in which the inertial forces have the same magnitude and direction as the forces in the gravitational field.

The point is that a gravitational force field is a *central field*: the gravitational forces are directed toward the centre of the gravitating mass, for example, the earth. These forces decrease in proportion to the square of the distance. No acceleration of a reference frame can provide such inertial forces. In the accelerated motion of a reference frame, inertial forces will always be parallel to one another (Fig. 24.10).

3. It follows that the principle of equivalence by no means implies that gravitational fields are only "apparent", and that in every case we can choose a reference frame in which there will be no gravitational field. Einstein repeatedly warned against this wrong conclusion. All that is implied is that the properties of space and time in a gravitational field are analogical to those in noninertial reference frames. This is the essence of Einstein's theory of gravity, certain ideas of which will be discussed below.

24.6. AN IDEA OF EINSTEIN'S THEORY OF GRAVITY

1. A gravitational field is set up in the vicinity of a massive body. It is characterized by the gravitational field strength I , which coincides in its physical meaning with the acceleration of a body at a given point of the field, and the scalar φ which is the gravitational field potential (see Secs. 9.6 and 18.8). In a uniform field directed along the axis of abscissas, the potential $\varphi = gx$, where g is the acceleration of gravity and x is the distance of the point in the field from the zero level of the potential which we take at the axis of ordinates [see equation (18.28)].

According to the principle of equivalence, the properties of space and time in a gravitational field should be the same as in noninertial reference frames.

2. Suppose that at a point where the gravitational field potential equals zero, an element of length along the axis of abscissas equals Δx_0 , and that an interval of time at the same point of the field equals Δt_0 . Then, at any other point of the field, where the potential equals φ , the element of length $\Delta x \neq \Delta x_0$ and the time interval $\Delta t \neq \Delta t_0$. To calculate their values we shall use formulas (24.3) and (24.5). Since, according to the principle of equivalence, $g = w$, the product $wx = gx = \varphi$. Substituting into equations (24.3) and (24.5) we obtain

$$\Delta x = \Delta x_0 \sqrt{1 - 2\varphi/c^2} \quad (24.6)$$

and

$$\Delta t = \frac{\Delta t_0}{\sqrt{1 - 2\varphi/c^2}} \quad (24.7)$$

Thus, in a gravitational field, as in noninertial reference frames, time is not uniform and space is non-Euclidean, i.e. neither uniform nor isotropic. This affirms the brilliant foresight of Lobachevsky in advancing the idea that the geometry of the real world may be non-Euclidean.

3. Let us see how, from this point of view, the motion of bodies is described in a gravitational field.

In Newtonian mechanics, a body moves by inertia in the absence of a gravitational field, i.e. in a straight line and at uniform velocity. If the body enters a gravitational field, the inertial properties of its motion are violated by the force. The body acquires acceleration and ceases to move in a straight line at uniform velocity.

According to the general theory of relativity (Einstein's theory of gravity), a *body always moves by inertia* whether it is in a gravitational field or not. Motion by inertia is motion along a *geodesic* (i.e. along the shortest distance), and its motion requires *minimum proper time* (i.e. time by a clock attached to the moving body).

If a body is moving outside the gravitational field, the space found there be uniform and isotropic, and the time uniform. Consequently, the body will move along a straight path and such velocity will be constant: since $\Delta x = \text{const}$ and $\Delta t = \text{const}$, then $v = \Delta x / \Delta t = \text{const}$.

However, if the same body is moving in a gravitational field, where time is nonuniform and space is both nonuniform and anisotropic, the geodesic will no longer be a straight line. It will be a curve whose shape depends upon the structure of the gravitational field. In exactly the same way the velocity will be found to be a variable value since the element of length and interval of time are functions of the gravitational field potential [see equations (24.6) and (24.7)].

On the basis of these ideas Einstein was able to devise a relativistic theory of gravitation, from which Newton's law of gravitation follows as a limiting case for weak fields and slow motion of the interacting bodies.

However, this new viewpoint on the properties of space and time enabled Einstein, not only to find a new theory of gravitation, but to discover three effects which were unknown in Newton's theory.

4. First of all, Einstein showed that the planets revolve about the sun along open curves rather than ellipses. Each of these curves can be approximately represented as an ellipse whose axis slowly rotates in the plane of the orbit.

It is very difficult to detect this effect for planets that are at a great distance from the sun. Mercury is the planet nearest to the sun; therefore it is in a stronger gravitational field and moves along its orbit faster than the other planets. For this reason, relativistic effects are more evident for Mercury. It was found that the rotation

of the major axis of Mercury's orbit should be 43 seconds of arc per century. This feature in the laws of motion of Mercury was observed by astronomers as far back as the middle of the 19th century, but they could not explain it. The most precise astronomic measurements yielded a displacement of 42.6 ± 0.9 seconds of arc per century. This is in excellent agreement with Einstein's equations.

5. The second effect, predicted by the general theory of relativity, is the bending of light rays in a gravitational field. Indeed, light

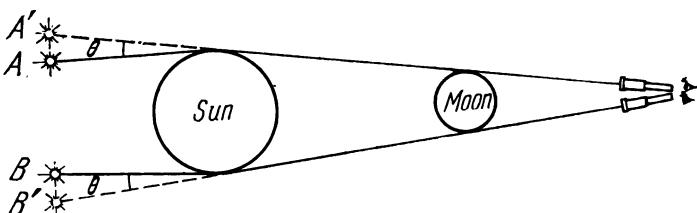


Fig. 24.11

always propagates along a geodesic, and in a gravitational field geodesics are curves of shapes depending upon the structure of the field.

Within the limits of the solar system, the only strong gravitational field is near the surface of the sun. This makes it feasible to observe the bending of light rays in the following way. It is necessary to photograph a portion of the sky near the sun, and then to photograph the same portion again when the sun is in a different part of the sky. The photographs should show a certain shift in the positions of the stars through the angle θ (Fig. 24.11). It is necessary, of course, to photograph the sky during a total eclipse of the sun. Otherwise, weak rays of light coming from stars close to the edge of the solar disk would not be visible on the bright background provided by the sun's light.

Einstein showed that

$$\theta = \frac{4\phi}{c^2} = \frac{4GM}{c^2R} \quad (24.8)$$

where M and R are the mass and radius of the sun. Substituting the values $G = 6.67 \times 10^{-11} \text{ m}^3/\text{kg}\cdot\text{s}^2$, $M = 2 \times 10^{30} \text{ kg}$, $R = 7 \times 10^8 \text{ m}$ and $c = 3 \times 10^8 \text{ m/s}$, we obtain

$$\theta = 0.85 \times 10^{-5} \text{ rad} = 1.75 \text{ seconds of arc}$$

Measurements made by a number of observers yielded results ranging from 1.72 ± 0.11 to 1.82 ± 0.15 seconds which, within the limits of errors of measurement, is in excellent agreement with theory.

6. Finally, owing to the nonuniformity of time, any periodic process in a gravitational field takes place at a slightly lower rate than outside such a field.

Suppose an atom emits electromagnetic waves (either visible light or gamma rays) at a certain frequency v_0 when it is outside of a gravitational field. Then the same atom, when located within a gravitational field with the potential φ , emits waves with a lower frequency v . As we know $v_0 = \frac{\Delta N}{\Delta t_0}$, where ΔN is the number of vibrations. In a gravitational field $v = \frac{\Delta N}{\Delta t}$ from which $\frac{v}{v_0} = \frac{\Delta t_0}{\Delta t}$. Using equation (24.7), we can write

$$v = v_0 \sqrt{1 - 2\varphi/c^2} \quad (24.9)$$

An experiment can be devised to register the change in frequency. Thus

$$\frac{\Delta v}{v_0} = \frac{v_0 - v}{v_0} = 1 - \sqrt{1 - 2\varphi/c^2} = \frac{2\varphi}{c^2(1 + \sqrt{1 - 2\varphi/c^2})}$$

Since, usually, $\frac{2\varphi}{c^2} \ll 1$, then

$$1 + \sqrt{1 - 2\varphi/c^2} \cong 2$$

Hence

$$\frac{\Delta v}{v_0} = \frac{\varphi}{c^2} \quad (24.10)$$

This effect of a change in frequency is extremely small in the gravitational field of the earth. At an altitude of h above the earth's surface, it is only

$$\frac{\Delta v}{v_0} = \frac{\varphi}{c^2} = \frac{gh}{c^2} = \frac{9.81h}{9 \times 10^{16}} = 1.09 \times 10^{-16} h$$

Nevertheless, in 1960, R.V. Pound (b. 1919) and G.A. Rebka, Jr. used the Mössbauer effect (see Sec. 81.11) to detect, experimentally, the change in the frequency of gamma radiation in the earth's gravitational field. They compared the emission of the nuclei of the isotope of iron with an atomic weight of 57 (iron-57) from two sources. One was at the surface of the earth and the other at an altitude of 22.5 m. Theoretically, the change in frequency should be $\frac{\Delta v}{v_0} = 2.46 \times 10^{-16}$. The result of the experiment was $\frac{\Delta v}{v_0} = (2.56 \pm 0.25) \times 10^{-16}$. The ratio of the two values is

$$\frac{\Delta v \text{ (experimental)}}{\Delta v \text{ (theoretical)}} = 1.04 \pm 0.12$$

which, within the possible limits of error of the experiment, unquestionably confirms Einstein's theory of gravity.

24.7. THE TWIN PARADOX

1. Let us imagine that a spaceship takes off from point A on the earth, accelerates, and then travels along the axis of abscissas at the velocity v . When the spaceship reaches point B it turns and travels toward the earth at the same velocity. Upon re-entry it is braked

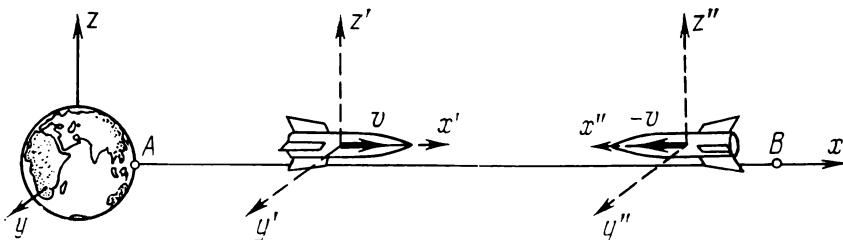


Fig. 24.12

and then lands. Let us denote the reference frame attached to the earth by xyz , the one attached to the spaceship when it is travelling away from the earth by $x'y'z'$ and the one attached to the returning spaceship by $x''y''z''$ (Fig. 24.12).

Let us now compare the time elapsed between takeoff and landing as shown by a clock on the earth and by one in the spaceship. It is assumed, of course, that the clocks run equally well in any of the reference frames.

2. According to the clock that remained on the earth, the time interval between these two events (takeoff and landing) is

$$\Delta t_{\text{earth}} = t_{2A} - t_{1A} = (t_{2A} - t_B) + (t_B - t_{1A}) \quad (24.11)$$

According to equation (12.22), the same time interval as shown by the spaceship clock is

$$\begin{aligned} \Delta t_{\text{space}} &= \Delta t' + \Delta t'' = \frac{t_{2A} - t_B + (x_A - x_B) v/c^2}{\sqrt{1 - v^2/c^2}} \\ &+ \frac{t_B - t_{1A} - (x_B - x_A) v/c^2}{\sqrt{1 - v^2/c^2}} \end{aligned} \quad (24.12)$$

But, since $\frac{x_A - x_B}{t_{2A} - t_B} = -v$ and $\frac{x_B - x_A}{t_B - t_{1A}} = v$ we can substitute into equation (24.12) and obtain

$$\begin{aligned} \Delta t_{\text{space}} &= (t_{2A} - t_B) \frac{1 - v^2/c^2}{\sqrt{1 - v^2/c^2}} + (t_B - t_{1A}) \frac{1 - v^2/c^2}{\sqrt{1 - v^2/c^2}} \\ &= (t_{2A} - t_B) \sqrt{1 - v^2/c^2} + (t_B - t_{1A}) \sqrt{1 - v^2/c^2} \\ &= \Delta t_{\text{earth}} \sqrt{1 - v^2/c^2} \end{aligned}$$

Hence

$$\Delta t_{\text{space}} = \Delta t_{\text{earth}} \sqrt{1 - v^2/c^2} \quad (24.13)$$

3. This problem can be solved in another way, from the viewpoint of an observer travelling in the spaceship. A clock which is at rest with respect to the spaceship will show its proper time (see Sec. 12.9):

$$\Delta t_{\text{space}} = \Delta t'_0 + \Delta t''_0 \quad (24.14)$$

We find the time interval according to the earth clock by using formula (12.21):

$$\Delta t_{\text{earth}} = \Delta t' + \Delta t'' = \frac{\Delta t'_0}{\sqrt{1 - v^2/c^2}} + \frac{\Delta t''_0}{\sqrt{1 - v^2/c^2}} = \frac{\Delta t_{\text{space}}}{\sqrt{1 - v^2/c^2}} \quad (24.15)$$

returning again to equation (24.13).

We see then, that from either point of view, time in a spaceship proceeds at a slower rate than on the earth. If, for instance, $v = 0.9998 c$, the spaceship clock runs at a rate only 1/50 of the earth clock! Therefore, if one twin gets into a spaceship and takes a journey at this velocity into distant outer space, returning one year later (by his clock), he will find that his twin brother who remained on earth has aged by 50 years.

4. The slowing down of time (or time dilation, as it is sometimes called) in a spaceship travelling at very high velocities reveals new possibilities, in principle, of reaching distant stars in a feasible length of time according to the spaceship's own clock. People remaining on the earth will have died, of course, many years before the astronauts return to relate the results of their journey.

It should be noted that we have no idea today of the engineering factors (including the source of energy) that would be required to accelerate a spaceship to ultrarelativistic velocities. Even the acceleration of elementary particles to these velocities involves great difficulties (see Secs. 41.4, 41.5 and 41.6).

5. The discussion above concerns what has been named "the twin paradox" because it seems strange at first sight. No paradox actually exists, and the objections raised against it by some scientists are unfounded. Let us consider several objections.

It is sometimes said that the twin paradox disagrees with the principle of relativity (see Sec. 12.2). This objection is based on a misunderstanding. The astronauts in a spaceship experience g -loads in launching, turning and landing, and weightlessness in free flight. All of this is not experienced by the people that stay on the earth. Hence, the two frames of reference are not equivalent; time proceeds differently in them. The principle of relativity cannot be applied.

Another objection is that the spaceship is not an inertial reference frame during takeoff, turn-around and landing, and that we used formulas from the special theory of relativity which refer to inertial

systems. This observation is reasonable, but it was found that nothing is changed when the noninertial nature of portions of the flight is taken into account. Time also slows down over the parts of the spaceship's path in which its velocity is nonuniform. This was shown in the preceding paragraphs.

6. Doubts are sometimes expressed as to the validity of applying the theory of relativity to the rate of processes taking place in living organisms. For some reason, the feasibility of applying the laws of Newtonian mechanics to these same objects is never questioned. All such doubts are quite unfounded. Science has proved that life amounts to physicochemical processes taking place in extremely complex systems that consist of high polymer compounds—proteins and nucleic acids. Owing to their degree of organization, these systems naturally have their specific regularities, but the general laws of physics, both Newtonian and relativistic, are fully applicable to such objects.

PART THREE ▲ MOLECULAR-KINETIC THEORY OF GASES

Chapter 25 MOLECULAR MOTION

25.1. HOW MOLECULAR SPEEDS WERE MEASURED

1. You know from elementary school physics that all bodies consist of molecules or atoms which are in a state of continuous chaotic motion. How do we go about determining the speed of molecules?

Such an experiment was first conducted in 1920 by O. Stern (1888-1969). The molecular beam method, proposed by him, was subsequent-

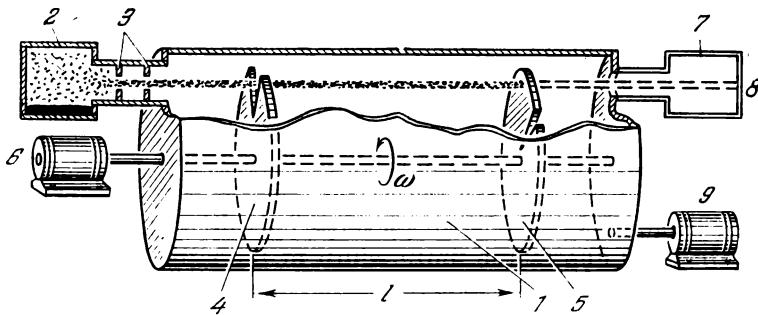


Fig. 25.1

ly employed by various investigators in more highly perfected apparatus. Let us consider the idea of the experiment carried out by B. Lammert in 1929.

2. The apparatus (Fig. 25.1) consists of thick-walled vessel 1 to which "molecular oven" 2 is connected. This oven is a vessel in which mercury (or some other liquid metal) is evaporated. The mercury vapour passes through a system of slits 3 which confine it to

a quite narrow molecular beam. Two disks 4 and 5, having narrow slits (Fig. 25.2) turned with respect to each other through an angle φ , are rotated by motor 6. After passing through the slits in both disks, the beam of molecules enters the catcher, or detector, 7, which is cooled by liquid nitrogen. The molecules are deposited on glass target 8 forming a visible sediment.

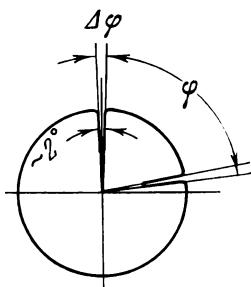


Fig. 25.2

A high vacuum is maintained in the apparatus by means of pump 9 to avoid collisions of the mercury molecules with molecules of air.

3. Obviously, the molecular beam cannot hit the target when the disks are stationary. If the disks are rotated (both disks being fixed on the shaft) then molecules having a definite speed can pass through the second slit. This will occur if during the time it takes the molecules to travel the distance between the disks the second disk turns exactly through the angle φ so that the slit is in the path of the beam.

If the disks rotate with an angular velocity $\omega = 2\pi n$ (where n is the frequency of rotation), then angle $\varphi = \omega t = 2\pi nt$. But $t = l/v$, where v is the speed of the molecules. Therefore $\varphi = 2\pi nl/v$ or

$$v = 2\pi nl/\varphi \quad (25.1)$$

The angle between the slits, the distance between the disks and the frequency of their rotation being known, we can calculate the speed of the molecules.

For example, if $l = 40 \text{ cm} = 0.4 \text{ m}$, $\varphi = 24^\circ = 24\pi/180 \text{ rad}$ and $n = 3000 \text{ rpm} = 50 \text{ rev/s}$, then the speed of the molecules is

$$v = \frac{2\pi \times 50 \times 0.4 \times 180}{24\pi} = 300 \text{ m/s}$$

4. It should be noted that we have measured the velocity with a certain error due to the finite width of the slits. Indeed, suppose a molecule travels alongside the left edge of the slit in the first disk. Through the second slit it can pass either alongside the left edge or the right edge. But in the first case, the system has turned through the angle φ , and in the second, the angle $\varphi_1 = \varphi + \Delta\varphi$. Consequently, the target will be hit by molecules moving at speed v (which will pass alongside the left edge), as well as by those having the lower speed $v_1 = 2\pi nl/\varphi_1$ (which pass along the right edge of the slit), and, of course, by molecules with speeds between v and v_1 . Thus the error in speed measurement is

$$\Delta v = v - v_1 = \frac{2\pi nl}{\varphi} - \frac{2\pi nl}{\varphi + \Delta\varphi} = \frac{2\pi nl\Delta\varphi}{\varphi(\varphi + \Delta\varphi)} = \frac{v\Delta\varphi}{\varphi + \Delta\varphi} \quad (25.2)$$

In the problem calculated above, for $\Delta\varphi = 2^\circ$, we obtain

$$\Delta v = 300 \times 2/26 = 23 \text{ m/s}$$

Thus, in the given experiment we can state only that the speed of the molecules is in the interval between 300 and 277 m/s.

The error can be reduced, of course, by making the slits narrower. But it is impossible, in principle, to completely eliminate the scatter in speeds because the slit cannot be made infinitely narrow.

25.2. MOLECULAR SPEED DISTRIBUTION

1. If all molecules travelled at the same speed, then, for a given angle between the slits and a given distance between the disks, all the molecules would enter the catcher only at a definite speed of rotation of the disks. In this case, at the suitable speed of rotation, the target would be quickly coated with molecules, while at other speeds no molecules would hit the target. Actually, entirely different results were obtained in this experiment. In the first place, molecules entered the catcher at different speeds of disk rotation. This shows that the beam is made up of molecules having different speeds. As a matter of fact it follows from equation (25.1) that molecules travelling at higher speeds enter the catcher at faster disk rotation, and slower molecules at lower speeds of rotation. It was also found that different times were required to obtain a deposit of the same thickness at different speeds of rotation. This last indicates that there are different amounts of fast and slow molecules in the beam.

2. The most difficult part of this experiment is to determine the number of molecules travelling at some definite speed. An exceptionally ingenious way out of this problem was found. Experiments showed that a visible deposit was obtained at a quite definite number of condensed molecules. The greater the number of molecules in the beam, the less the time required to obtain a visible deposit, i.e.

$\frac{N_1}{N_2} = \frac{t_2}{t_1}$. Thus, in the given experiment, we can determine the relative number of molecules that have a speed within the interval from v to $v + \Delta v$, i.e. $\frac{\Delta N}{N \Delta v}$.

The data of one such experiment is listed in Table 25.1 and plotted in Fig. 25.3. The measurements were made with molecules of some substance at a definite temperature of the "molecular oven".

3. The results of the experiment indicate that there is a certain speed at which the maximum number of molecules travel. In our case it lies within the speed interval from 270 to 300 m/s. This speed is said to be the *most probable* one. Let us denote it by u .

Table 25.1

Speed interval, m/s	Relative number of molecules, $\Delta N/N\Delta v$	Speed interval, m/s	Relative number of molecules, $\Delta N/N\Delta v$	Speed interval, m/s	Relative number of molecules, $\Delta N/N\Delta v$
60 to 90	82	270 to 300	368	480 to 510	161
90 to 120	137	300 to 330	361	510 to 540	127
120 to 150	195	330 to 360	342	540 to 600	86
150 to 180	251	360 to 390	313	600 to 660	46
180 to 210	300	390 to 420	278	660 to 720	22
210 to 240	338	420 to 450	237	720 to 840	9
240 to 270	361	450 to 480	198		

Most of the molecules travel at speeds near to the most probable speed. There are not many molecules whose speed is very much higher or very much lower than the most probable speed.

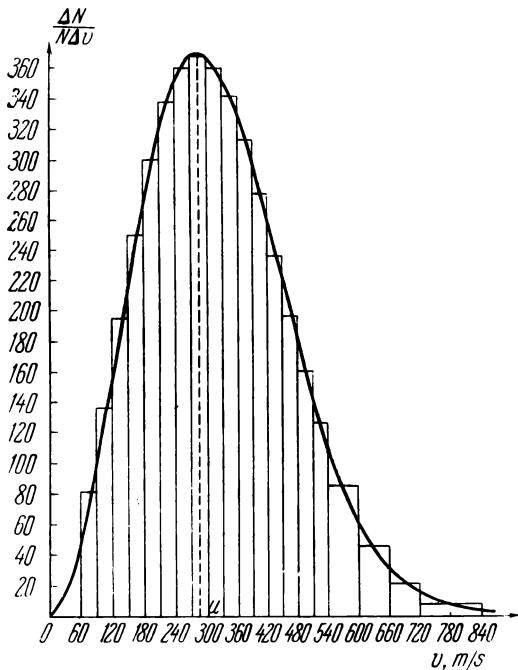


Fig. 25.3

Experimentally obtained molecular speed distribution agrees well with the theoretical distribution derived by James Clerk Maxwell

(1831-1879) as far back as 1859 by applying the laws of the theory of probability to thermal random motion. The law he discovered is called the *Maxwellian molecular speed distribution*. The good agreement of the theoretical law with the experimental data is an excellent confirmation that our conception of the nature of molecular motion in gases is entirely correct.

4. Using either experimental data or the Maxwellian distribution, we can calculate the most probable speed of the molecules of a gas. It was found that this speed depends upon the temperature of the gas and the mass of its molecules.

The most probable speeds of hydrogen molecules, for instance, at 0° and 100 °C are 1510 and 1765 m/s. Under the same conditions, the most probable speeds for oxygen are 378 and 442 m/s, i.e. one-fourth of those for hydrogen. For mercury vapour these speeds are 151 and 176.5 m/s, i.e. one-tenth of those for hydrogen.

It is known that the mass of a molecule of hydrogen is one-sixteenth of that of a molecule of oxygen and 1/100 of that of an atom of mercury (a molecule of mercury consists of one atom; the molecules of hydrogen and oxygen are diatomic). Therefore, at a given temperature of the gases the most probable speed of their molecules is inversely proportional to the square roots of their masses. Thus

$$\frac{u_1}{u_2} = \sqrt{\frac{m_2}{m_1}} \quad (25.3)$$

from which

$$\frac{m_1 u_1^2}{2} = \frac{m_2 u_2^2}{2} \quad (25.4)$$

Hence, the *average kinetic energy of the molecules does not depend upon the kind of gas but only on its temperature*.

25.3. MEAN FREE PATH OF MOLECULES

1. It would seem strange, at first sight, that the diffusion of gases is comparatively slow since the speed of their molecules at room temperatures is several hundred metres per second. For example, it may take an odour dozens of seconds to spread from one end of a room to the other, and not hundredths of a second, as we could expect. The reason is that molecules collide very frequently with one another and, as a result, their paths are very complex and intricate broken lines, rather than long straight lines.

An idea of the nature of this path can be obtained by observing Brownian motion in a microscope. In 1827, Robert Brown (1773-1858) watched the behaviour through a strong magnifying glass of extremely fine pollen, about one micron in size, suspended in water. He observed that these particles are in a state of continuous chaotic

motion; that they seem to be engaged in some fantastic dance. It was found, subsequently, that all microscopic particles suspended in liquids or gases have Brownian motion.

2. As shown independently in 1905 by Einstein and Marian Smoluchowski (1872-1947), Brownian particles behave like gigantic molecules whose average kinetic energy is equal to that of the molecules of the liquid or gas surrounding the particles. Therefore, the nature of the motion of Brownian particles fully corresponds with that of the surrounding molecules, but with one essential difference: the speed of the particles is considerably less than that of the molecules. Indeed, from equality of the average kinetic energies of the particles and molecules, $\bar{K}_{par} = \bar{K}_{mol}$, it follows that

$$\frac{M\bar{v}_{par}^2}{2} = \frac{m\bar{v}_{mol}^2}{2} \quad \text{or} \quad \bar{v}_{par} = \bar{v}_{mol} \sqrt{m/M}$$

where the bar denotes the average speeds, and m and M are the masses of the molecules and Brownian particles. A Brownian particle consists of thousands of millions of molecules: $M/m \cong 10^{10}$, and the

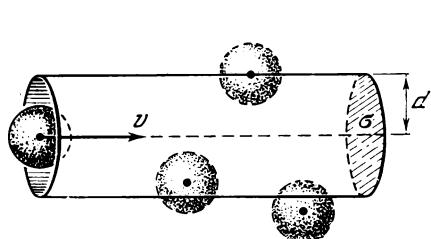


Fig. 25.4

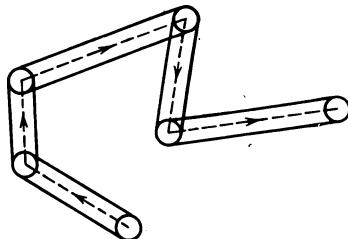


Fig. 25.5

speed of gas molecules is several hundred metres per second. Consequently, the speed of the Brownian particles should be several millimetres per second which is observed in practice.

3. As we see, a molecule is subject to a great many collisions each second. Let us try to calculate the *number of collisions per second* and the *mean free path* of molecules.

Let us assume that molecules of a gas behave like elastic balls in collisions. For simplicity, we shall suppose that only one molecule, the one at the extreme left in Fig. 25.4, is in motion, and that all the others are at rest. In the collision of two molecules, their centres approach each other to the distance d , equal to the diameter of the molecules. It is therefore convenient to assume that the moving molecule has a radius $r = d$ and that the other molecules are point particles (in the sense of point masses or material points).

Upon each collision, the direction of motion of the molecule changes, so that its path is a complicated broken line (Fig. 25.5). For

convenience in calculation let us imagine that this path is straightened out as shown in Fig. 25.4. This cannot affect the calculation of the number of collisions because the shape of the path plays no role here. During the time t , the molecule being considered travels a distance $l = vt$. In this time it will collide with all the molecules whose centres lie within a circular cylinder of length $l = vt$ and with a base of area σ whose diameter $2d$ is twice that of the molecules. The area σ is called the *effective collision cross section*.

The number of molecules in this cylinder is $N = ncl = novt$, where n is the *molecule concentration*, i.e. the number of molecules in unit volume. In one second the molecule experiences z collisions:

$$z = N/t = n\sigma v \quad (25.5)$$

4. If the molecule had been travelling in a vacuum, not experiencing any collisions, it would have covered a distance in one second numerically equal to its speed. But, actually, it is subject to z collisions per second. Therefore, its mean free path is

$$\lambda = \frac{v}{z} = \frac{1}{n\sigma} \quad (25.6)$$

This calculation is very approximate since we made many quite artificial assumptions in its derivation. Rigorous theory, however, taking into account the motions of all the molecules, yields a result differing from the one we obtained by 30 to 40 per cent. This is of no significance in our case because we are interested only in the order of magnitude rather than the exact values.

5. Several procedures exist for determining the mean free path experimentally. It was found that at standard atmospheric pressure, the mean free path of gas molecules is about 0.1 micron = 10^{-7} m, the concentration of the molecules in the gas under these conditions being $n \approx 10^{25}$ molecules per m^3 . Now we can find the effective collision cross section:

$$\sigma = \frac{1}{n\lambda} \approx \frac{1}{10^{-7} \times 10^{25}} = 10^{-18} \text{ m}^2$$

Since $\sigma = \pi d^2$, where d is the effective diameter of the molecule, then

$$d = \sqrt{\sigma/\pi} \approx 10^{-9} \text{ m} = 10 \text{ \AA}$$

Here 1 \AA (angström) = 10^{-10} m is a length unit frequently used in optics and in molecular and atomic physics. In order of magnitude the result obtained here agrees well with size values for molecules obtained by other methods.

25.4. DIFFUSION

1. Diffusion is the process of levelling the concentrations of two substances when they are mixed with each other. Diffusion is observed in gases as well as in liquids and solids.

The following experiment will demonstrate diffusion in gases. A thick-walled steel cylinder is divided by a partition with a valve into two parts (Fig. 25.6). The lower part is filled with oxygen and the upper with hydrogen. Screwed into the side walls are spark plugs between whose points a spark can be produced. At the beginning of the experiment such a spark does not cause an explosion because

the oxygen is separated from the hydrogen. However, if the valve is opened, the gases will soon intermix and a spark will readily blow up the mixture—detonating gas—thus obtained. Note that in diffusion the heavier oxygen rises upward, and the hydrogen, weighing only $1/16$ as much, descends. Consequently, the motion of the gases is contrary to the action of the gravitational field.

This phenomenon is readily explained on the basis of molecular-kinetic theory. The point is that the molecules of the two gases, moving chaotically and colliding with one another,

enter the spaces between neighbouring molecules. This leads to the penetration of one gas into the other. This proceeds gradually until a mixture of the two gases is obtained with the same concentration of the two kinds of molecules.

2. Owing to diffusion, the surrounding atmosphere (or, in any case, its lower layer, the troposphere) is a homogeneous mixture of nitrogen, oxygen and carbon dioxide with small admixtures of inert gases and water vapour. If there were no diffusion, the gravitational field would divide up the atmosphere into layers. The lowest layer would consist of the heaviest gas, carbon dioxide, above it would be oxygen and still higher, nitrogen and the inert gases.

3. Since the speed of molecules increases with the temperature, the rate of diffusion also increases, as is observed in practice. Everybody knows that sugar or salt dissolve much quicker in hot water than they do in cold water. This is another example of the speeding up of diffusion with an increase in temperature. Similar phenomena are observed in gases and solids.

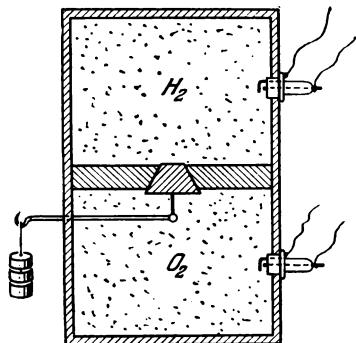


Fig. 25.6

25.5. LAW OF DIFFUSION

1. Let us investigate the mechanism of diffusion in gases in more detail and determine, at the same time, on what the mass of the substances carried over in diffusion depends. To simplify matters we shall consider *self-diffusion*, the case when a vessel is filled with a single gas, but the concentration of the molecules varies at different parts of the vessel.

Suppose, for instance, that the vessel is a tube in which a surplus concentration is maintained at the left end and a certain degree of vacuum at the right end (Fig. 25.7). The molecule concentration n

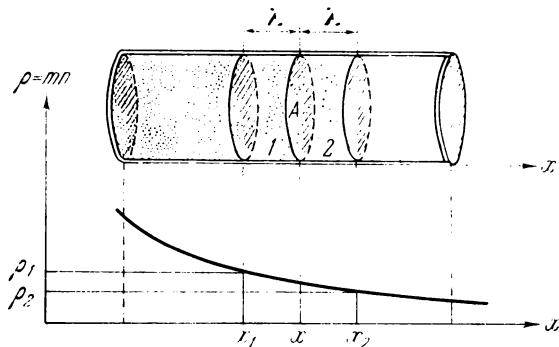


Fig. 25.7

and the proportionate gas density $\rho = m_0 n$ decrease along the axis of abscissas. Let us consider a certain cross-sectional area A in the tube and determine the mass Δm of the substance passing through unit area in unit time, i.e. the quantity $\Delta m/A\Delta t$.

2. Only the molecules located at distances not exceeding the length of the mean free path can pass through the cross section, i.e. molecules within layers 1 and 2 in Fig. 25.7. But, since molecules travel absolutely randomly, all directions in space are the same to them, and, on an average, only one sixth of all the molecules travel in any given direction. Thus, $N/6$ molecules travel along the positive direction of the axis of abscissas, $N/6$ along its negative direction, $N/6$ along the positive direction of the axis of ordinates, etc.

Volume 1 contains $N_1 = n_1 A \lambda$ molecules. Of them one sixth passes through cross-sectional area A , i.e. $N_1/6 = \frac{1}{6} n_1 A \lambda$. In the same way, $N_2/6 = \frac{1}{6} n_2 A \lambda$ molecules pass from volume 2 through the cross section. Thus, during the time Δt , $\Delta N = \frac{1}{6} A \lambda (n_1 - n_2)$ mole-

cules diffuse through the cross section to equalize the concentration at the ends of the tube. Multiplying both sides of this equation by the mass m_0 of a molecule, and, since $m_0 \Delta N = \Delta m$ and $m_0 (n_1 - n_2) = m_0 n_1 - m_0 n_2 = \rho_1 - \rho_2 = -\Delta\rho$, we obtain

$$\Delta m = -\frac{1}{6} A\lambda\Delta\rho \quad (25.7)$$

3. Let us transform this equation to some extent. The quantity $\frac{\Delta\rho}{\Delta x} = \frac{\rho_2 - \rho_1}{x_2 - x_1}$, indicating the rate of change of the gas density along the axis of abscissas, is called the *density gradient*. The fraction $\frac{\Delta x}{\Delta t} = \frac{2\lambda}{\Delta t} = 2v$ is equal to twice the average speed of the molecules. Hence, we can rewrite equation (25.7):

$$\Delta m = -\frac{1}{6} A\lambda\Delta\rho = -\frac{1}{6} A\lambda \frac{\Delta\rho}{\Delta x} \frac{\Delta x}{\Delta t} \Delta t = -\frac{1}{3} \lambda v \frac{\Delta\rho}{\Delta x} A\Delta t$$

Therefore, the mass of the substance carried through unit area in unit time as a result of diffusion is expressed by

$$\Delta m = -\frac{1}{3} \lambda v \frac{\Delta\rho}{\Delta x} A\Delta t \quad (25.8)$$

or

$$\Delta m = -D \frac{\Delta\rho}{\Delta x} A\Delta t \quad (25.9)$$

i.e. the *mass of the diffusing substance is proportional to the density gradient, cross-sectional area of the tube and the time*. This is the law of diffusion.

4. The quantity

$$D = \frac{1}{3} \lambda v \quad (25.10)$$

is called the *coefficient of diffusion*. If we know the coefficient of diffusion and the average speed of the molecules, we can determine their mean free path.

25.6. SEPARATION OF GAS MIXTURES

1. Interesting phenomena are observed in the diffusion of gases through a porous partition. They can be demonstrated by the following experiment (Fig. 25.8). A porous vessel of unburnt clay is connected by a rubber tube to a water manometer. This vessel is covered with a small beaker into which we run hydrogen, a gas whose molecules have much less mass than those of nitrogen or oxygen of which air consists. We shall see that the level of the water in the manometer changes, indicating that the pressure in the porous vessel has become higher than the atmospheric pressure.

This phenomenon can be explained as follows. Both air and hydrogen diffuse through the walls of the porous vessel. Air comes out and hydrogen penetrates into the vessel. However, their rates of diffusion differ. Since the mass of a molecule of hydrogen is only $1/16$ that of oxygen and $1/14$ that of nitrogen, the speed of the hydrogen molecules and, consequently, the coefficient of diffusion of hydrogen, are about 4 times greater than for air. As a result, much more hydrogen penetrates into the porous vessel than the air passing out through its walls. This raises the pressure in the vessel.

Proceeding from these results we can foresee that if we submerge the porous vessel in a heavier gas, for example, carbon dioxide, the rate of diffusion of the air out of the vessel will be much higher than that of the carbon dioxide into the vessel. Consequently, the pressure inside the vessel will drop below the atmospheric pressure, as will be indicated by the manometer. This assumption is confirmed by experiment.

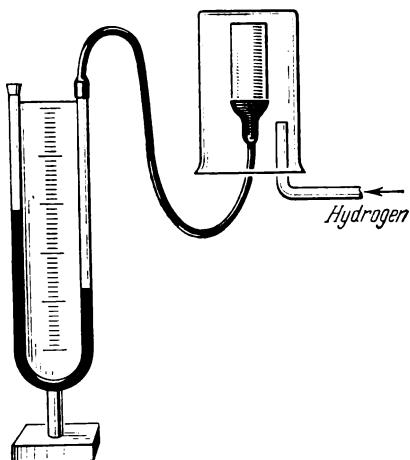


Fig. 25.8

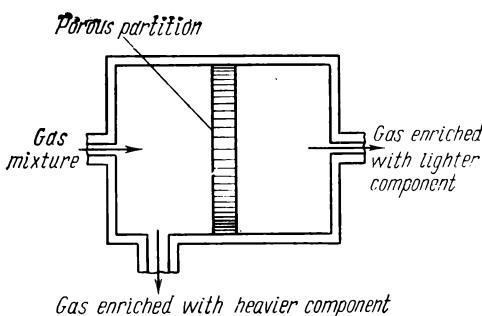


Fig. 25.9

2. The phenomenon described above can be utilized to separate gas mixtures. Let us imagine a vessel divided in half by a porous partition (Fig. 25.9). Into the left half we pump a gas mixture consisting of two components. Since the lighter component diffuses more rapidly through the porous partition than the heavier one, at

the outputs of the instrument we obtain new gas mixtures, one enriched with the lighter component and the other with the heavier one. By passing the mixture we are interested in through a great number of such chambers we can achieve a practically complete separation of one component from the other.

3. Almost all chemical elements are mixtures of their *isotopes*, i.e. atoms having exactly the same chemical properties but different atomic weights.

Natural hydrogen, for instance, consists of 99.985 per cent of the light isotope (protium) and 0.015 per cent of the heavy isotope (deuterium). The masses of their atoms are approximately in the ratio 1 : 2 (more exactly—1.008142 : 2.014735). Artificially, we can also obtain the superheavy isotope (tritium) whose atoms have a mass about three times that of protium. Natural uranium consists mainly of two isotopes. The masses of their atoms are in the ratio, approximately, of 235 : 238. The lighter isotope constitutes 0.715 per cent, the heavier—99.28 per cent of the whole mass of the uranium. The remaining 0.005 per cent consists of other isotopes.

It is impossible to separate isotopes by chemical methods because their properties are exactly the same. To separate them we must make use of some physical phenomenon depending on the mass of the atoms.

One of the most important procedures for separating isotopes is the diffusion method described above.

For example, certain types of nuclear reactors require enriched uranium, containing up to 80 per cent of the light isotope. It is produced from a gas—uranium hexafluoride (UF_6)—which is delivered by pumps through a system of several hundred or even thousand diffusion chambers (depending upon the required degree of enrichment). Thus the installment for the diffusion separation of uranium isotopes is a huge, highly expensive unit.

Chapter 26

AN IDEAL GAS

26.1. GAS PRESSURE

1. We know that when a gas is enclosed in some vessel it exerts pressure on the walls. On the basis of molecular-kinetic theory, this phenomenon is explained as follows. In their entirely random motion the molecules strike the walls of the vessel. The total momentum that the molecules transmit to unit area in unit time is the pressure exerted by the gas.

To determine the pressure we shall make use of formula (17.17) which describes the pressure of a stream of particles striking a wall with elastic collisions (see Sec. 17.5):

$$p = 2nm_0v^2 \quad (26.1)$$

where n is the concentration of the particles (number of particles in unit volume). However, two essential corrections must be made in this equation. In Sec. 17.5 we dealt with a stream of particles that all moved in a single direction with the same velocity. The molecules of an ideal gas travel randomly in all possible directions and their speeds, according to Maxwell's distribution, differ. Therefore, the square of the velocity in formula (26.1) should be replaced by the *average value of the square of the velocity* for all the particles in the vessel:

$$\langle v^2 \rangle = \bar{v}^2 = \frac{1}{N} (v_1^2 + v_2^2 + \dots + v_N^2)$$

The square root of the average velocity squared $\sqrt{\bar{v}^2} = \sqrt{\langle v^2 \rangle}$ is called the *root-mean-square speed* of the molecules.

It is also necessary in equation (26.1) to replace the concentration n by one sixth of the concentration, $n/6$. Indeed, owing to the completely chaotic motion of the molecules and the great number of them in unit volume, we can contend that one sixth of all the molecules is moving in any one direction.

2. In this way, the *pressure of a gas is proportional to the concentration of the molecules and to their average kinetic energy*:

$$p = \frac{2}{3} n \frac{m_0 \bar{v}^2}{2} \quad (26.2)$$

Equation (26.2) can be given another form. Since the molecule concentration $n = N/V$, where N is the total number of molecules in a vessel of volume V , we can substitute into equation (26.2) and obtain

$$pV = \frac{2}{3} N \frac{m_0 \bar{v}^2}{2} \quad (26.3)$$

3. It has been found by experiment that the average kinetic energy of a gas does not depend upon the kind of gas; it is determined only by its temperature (see Sec. 25.2). Consequently, in an *isothermal process* (i.e. a process occurring at constant temperature), the average kinetic energy of the molecules is a constant. But then it follows from equation (26.3) that the product of the pressure of a gas by its volume is, under these conditions, a constant. Thus, we have obtained the following important result:

At a constant temperature ($m_0 \bar{v}^2/2 = \text{const}$) and constant mass of the gas (number of molecules $N = \text{const}$), the product of its pressure

by its volume is a constant (Boyle's law, or the Boyle-Mariotte law)

$$pV = \text{const} \quad (26.4)$$

4. This law is frequently stated as follows: *in an isothermal process, the pressure of a gas is inversely proportional to its volume* (for a fixed mass of the gas).

The graph of an isotherm with the coordinate axes p and V is a parabola (Fig. 26.1). For the same mass of gas, the different curves correspond to the different temperatures.

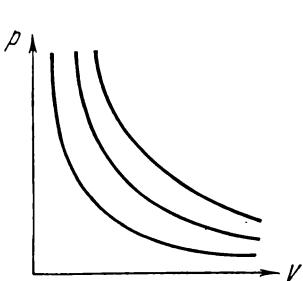


Fig. 26.1

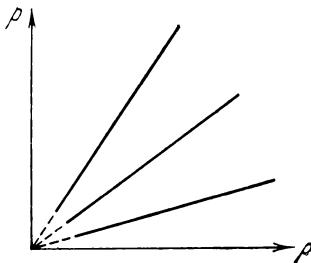


Fig. 26.2

5. The volume is equal to the mass divided by the density: $V = m/\rho$. Substituting in equation (26.4) and, since we assume the mass of the gas to be constant, we can write

$$p/\rho = \text{const} \quad (26.5)$$

Thus, another statement of the Boyle-Mariotte law is: *at constant temperature and fixed mass of a gas, its pressure is proportional to its density*.

A graph of this equation with the coordinate axes p and ρ is shown in Fig. 26.2. The inclination of the curve depends upon the temperature of the gas.

26.2. UNITS OF PRESSURE

1. The SI unit of pressure is the *pascal* (Pa). It is the pressure exerted by a force of one newton equally distributed over a surface with an area of 1 m^2 normal to the force. Thus

$$1 \text{ Pa} = 1 \text{ N/m}^2$$

The cgs unit of pressure is the *dyne/cm²*:

$$1 \text{ Pa} = 1 \text{ N/m}^2 = 10 \text{ dyne/cm}^2$$

Use is often made of a nonsystem unit of pressure—the *bar*:

$$1 \text{ bar} = 10^5 \text{ Pa} = 10^6 \text{ dyne/cm}^2$$

Another nonsystem unit that is widely used is the *technical atmosphere* (at):

$$1 \text{ at} = 1 \text{ kgf/cm}^2 = 98066.5 \text{ Pa} \cong 9.81 \times 10^4 \text{ Pa} = 0.981 \text{ bar}$$

With an error of the order of 2 per cent, it is possible to neglect the difference between the technical atmosphere and bar.

2. Pressure in a liquid or gas can be conveniently compared with the pressure exerted by a vertical column of some liquid. This is

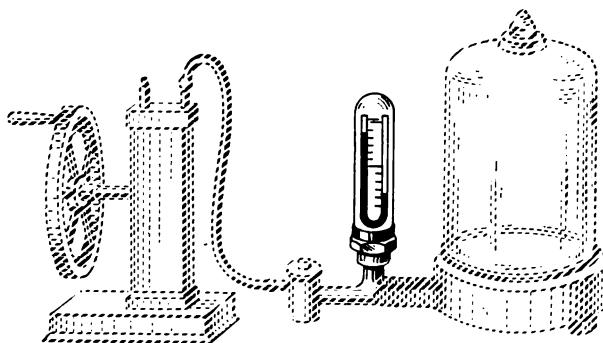


Fig. 26.3

the method used to measure the pressure of the atmosphere by means of a mercury barometer or the degree of vacuum in a vacuum system (Fig. 26.3) with a manometer.

The pressure exerted by a vertical column of liquid is found by dividing its weight by the area of its base. Thus

$$P = \frac{P}{A} = \frac{mg}{A} = \frac{\rho h Ag}{A} = \rho hg \quad (26.6)$$

where ρ = density of the liquid

h = height of the column

g = acceleration of gravity.

An inclined-tube manometer is often used to measure low pressures. Here

$$p = \rho gl \sin \alpha \quad (26.7)$$

where α = angle between the tube and a horizontal plane

l = length of the column of liquid.

3. In practice it is frequently convenient to specify the pressure by the height of the column of liquid in the manometer tube. The corresponding units are:

millimetres of mercury column (mm-Hg or torr):

$1 \text{ mm-Hg} = 10^{-3} \text{ m} \times 13\,595 \text{ kg/m}^3 \times 9.80665 \text{ m/s}^2 = 133.322 \text{ Pa}$
standard, or physical, atmosphere (atm):

$$1 \text{ atm} = 760 \text{ mm-Hg} = 1.01325 \times 10^5 \text{ Pa}$$

(with an error of the order of 3 per cent, the difference between physical and technical atmospheres can be neglected);

metres of water column (m-H₂O):

$$1 \text{ m-H}_2\text{O} = 1 \text{ m} \times 10^3 \text{ kg/m}^3 \times 9.81 \text{ m/s}^2 = 9.81 \times 10^3 \text{ Pa} = 0.1 \text{ atm}$$

26.3. AN IDEAL GAS

1. The Boyle-Mariotte law (called Boyle's law in England and the USA) was repeatedly checked by experiments. It was found that at pressures not exceeding several dozens of atmospheres, and at

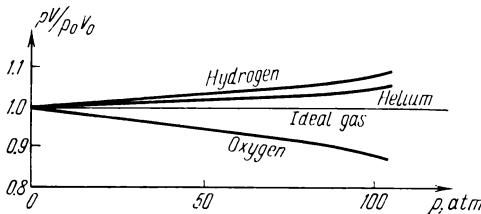


Fig. 26.4

relatively high temperatures (of the order of room temperature and higher), it holds quite well. The deviation of the experimental data from theoretical calculations is within 1 to 3 per cent. However, at pressures of the order of hundreds of atmospheres and higher, the product pV obtained experimentally differs substantially from the theoretical value.

Data is given in Table 26.1 and Fig. 26.4 on the compression of three different gases at constant temperature $t_0 = 0^\circ\text{C}$, at an

Table 26.1

Oxygen			Hydrogen			Helium		
$p, \text{ atm}$	$V, \text{ litres}$	$\frac{pV}{p_0V_0}$	$p, \text{ atm}$	$V, \text{ litres}$	$\frac{pV}{p_0V_0}$	$p, \text{ atm}$	$V, \text{ litres}$	$\frac{pV}{p_0V_0}$
10.07	2.20	0.988	10.07	2.24	1.007	10.04	2.24	1.004
51.4	0.408	0.937	51.9	0.448	1.04	51.1	0.448	1.02
106	0.187	0.884	108	0.224	1.08	105	0.224	1.05

initial pressure of $p_0 = 1$ atm and the same initial volume of $V_0 = 22.4$ litres.

2. An analysis of the test data indicates that at pressures not exceeding several dozens of atmospheres, the relationship between the pressure and volume of the gas is expressed by the Boyle-Mariotte law. It proves expedient under these conditions to disregard the slight deviation of the experimental data from the theoretical value $pV = \text{const}$, and to introduce the concept of an *ideal gas* which behaves exactly according to the Boyle-Mariotte law.

Hence, an *ideal gas* is one in which the pressure in an isothermal process is exactly inversely proportional to its volume (for a fixed mass of gas).

The concept of an ideal gas is an idealization, of course, as are the concepts of a point particle, point charge, etc. Such idealizations prove highly fruitful in many cases because they enable the solution of a problem to be considerably simplified without appreciably impairing the accuracy of the solution.

3. Let us try to find what the molecular structure of an ideal gas would be. To do this we shall compare the density of the gas with that of a liquid (or a solid) at the same pressure and temperature.

At standard atmospheric pressure and a temperature of 100 °C, the density of water is 958.4 kg/m³ and that of steam is 0.598 kg/m³, i.e. only 1/1600 as much. But, as we know, density is the product of the mass of the molecules by their concentration: $\rho = m_0 n$. Consequently, there are 1600 times as many molecules in unit volume of water than in unit volume of steam.

Since the volume is proportional to the cube of the linear dimensions, the distance between the molecules of steam is $\sqrt[3]{1600} \approx \cong 12$ times more than the distance between water molecules under the same conditions. Similar calculations for other substances yield the same result: the *distances between the molecules of gases are dozens of times greater than those of liquids or solids* (at not very high pressures).

4. Further on we shall show that molecular forces change very rapidly with the distance. The forces of molecular attraction, for instance, are inversely proportional to the *seventh power* of the distance (see Sec. 31.4). Therefore, the *forces of interaction between the molecules of liquids and solids are tens of millions of times greater than those of gases* (at not very high pressures). This means that the forces of molecular interaction can be *completely neglected* under these conditions.

We thereby justify the implicit assumption we made in deriving formula (26.2), where we took into consideration only the interaction between the molecules and the walls of the vessel, but not that between the molecules themselves. Therefore, the Boyle-Mariotte

law holds for gases, in which the forces of molecular interaction play no appreciable role.

We should note that the size of the molecules was also disregarded, since their total volume (calculated, for instance, from their effective dimensions) is very much less than the volume of the vessel.

Thus, from a molecular point of view, an *ideal gas is a system of molecules which do not interact with one another and which can be regarded as a first approximation as point particles.*

26.4. TEMPERATURE

1. The concept of temperature originated from the sensation of "hot" or "cold". Such a criterion is very personal, however, because our sensations depend both on the state of our surroundings and on the state of our body. Thus if two persons enter the same room, one having just taken a hot bath and the other coming in from the frost, the first will feel cold and the second, warm.

It was necessary, therefore, to find an impersonal criterion to characterize the temperature, and also to establish a unique method for measuring this quantity.

2. From the point of view of thermodynamics, *temperature is a quantity indicating the direction of heat exchange*. Indeed, experiments show that heat exchange always takes place so that energy in the form of heat is transferred only from hotter to colder bodies.

This phenomenon can be demonstrated by the following experiment. We drop a hot piece of iron into a vessel containing a certain amount of water. After some time we shall find that the temperatures of the two bodies have become equal; to the touch, the water and the iron will have the same warmth. But the iron was cooled, correspondingly reducing its internal energy, and the water was heated so that its internal energy increased.

3. By definition (see Sec. 21.3) the quantity of heat is the energy transmitted by heat exchange. Therefore, when two bodies at different temperatures are in contact, a *quantity of heat is transmitted from the body with the higher temperature to the one with the lower temperature.*

The equality of temperatures is determined by the following condition: the *temperatures of two bodies are equal if no heat exchange occurs between them.*

This must be taken into account in measuring temperatures. A thermometer always indicates its own temperature. Therefore, to measure the temperature of a body correctly, the thermometer should be brought into contact with the body and held there until its reading is stabilized. This means that heat exchange has ceased between the body and the thermometer, and they have the same temperature.

4. From the point of view of molecular-kinetic conceptions, *temperature is a quantity characterizing the average kinetic energy of translational motion of the molecules of an ideal gas*. This definition makes sense because we have found that at the same temperature, the average kinetic energies of the molecules of different gases coincide (see Sec. 25.2).

On the basis of the thermodynamic meaning of temperature, we can reduce the measurement of the temperature of any body to the measurement of the average kinetic energy of the molecules of an ideal gas which is in thermodynamic equilibrium with the given body. However, any measurement of the velocities (or kinetic energies) of molecules is extremely difficult. In practice, therefore, instead of the kinetic energy of the molecules, a proportionate quantity is measured, for instance, the pressure of the gas.

26.5. ABSOLUTE TEMPERATURE AND THE EQUATION OF STATE OF AN IDEAL GAS

1. Let us define the *absolute temperature* as a *physical quantity proportional to the average translational kinetic energy of the molecules of an ideal gas*. On the basis of this definition and the laws of classical molecular theory, the relationship between the temperature and average kinetic energy of the molecules of such a gas is

$$\bar{\varepsilon}_k = \frac{3}{2} kT \quad (26.8)$$

where $\bar{\varepsilon}_k$ = average translational kinetic energy of the molecules
 T = absolute temperature
 k = proportionality factor called *Boltzmann's constant* (see Sec. 26.9).

To enable us to make use of the concept of the absolute temperature stated above, we must first establish a method for measuring this quantity, then establish a unit of measurement and, finally, determine the zero level of temperature. This will be done in the following sections.

2. As has been shown, the average kinetic energy of the molecules of an ideal gas can be expressed in terms of the pressure of the gas and the concentration of its molecules [see equation (26.2)]:

$$\bar{\varepsilon}_k = \frac{m_0 \bar{v}^2}{2} = \frac{3}{2} \frac{p}{n} \quad (26.9)$$

Combining equations (26.8) and (26.9), we can write

$$p = nkT \quad (26.10)$$

i.e. the *pressure of an ideal gas is proportional to its absolute temperature and to the concentration of its molecules*.

3. Since $n = N/V$, we obtain

$$\frac{pV}{T} = Nk = \text{const} \quad (26.11)$$

This equation is called the *equation of state of the gas*. It is formulated as follows:

For a fixed mass of gas (i.e. for a fixed number of molecules N), the product of its pressure by its volume is proportional to its absolute temperature.

26.6. THE CONSTANT-VOLUME GAS THERMOMETER

1. The *constant-volume gas thermometer* (Fig. 26.5) is an instrument for measuring the absolute temperature. It consists of bulb 1, filled with gas (commonly hydrogen), which is connected by tube 2, with

an expanded end, and rubber hose 3 to manometer tube 4. The bulb with hydrogen is brought into contact with the body whose temperature is to be measured. By raising or lowering manometer tube 4 we can make the level of the mercury in the expanded end of tube 2 coincide with a fixed reference mark, thus keeping the gas confined in bulb 1 at a constant volume when we are measuring its pressure.

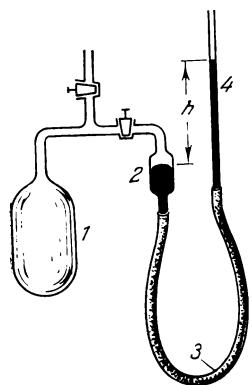


Fig. 26.5

2. The temperatures of two bodies are compared as follows. First we bring the gas thermometer into thermodynamic equilibrium with the first body, and then with the second, measuring the pressure each time from the height of the mercury column. Then, on the basis of the equation of state of the gas

(26.11), and the condition that the volume of gas is constant, we can come to the conclusion that

$$\frac{T_1}{T_2} = \frac{p_1}{p_2} \quad (26.12)$$

Thus, we have reduced temperature measurement to a measurement of gas pressure.

In everyday practice, of course, temperature is not measured by constant-volume gas thermometers. Other instruments, such as mercury and alcohol thermometers, resistance thermometers and thermocouples are used. However, all of these instruments are graduated by means of a gas thermometer which serves, in this way, as a standard instrument.

3. The choice of hydrogen, and not some other gas, as the thermometric body was not a matter of chance. It was found that the read-

ings of a hydrogen thermometer are more accurate than when the bulb of a gas thermometer is filled with any other gas. They are closer to the readings that a gas thermometer with an ideal gas would provide.

Table 26.2 lists the errors of gas thermometers in measuring the temperature of boiling water. The relative error is

$$\delta = \frac{\Delta p}{p_{id}} = \frac{p - p_{id}}{p_{id}}$$

where p_{id} = reading of a gas thermometer with an ideal gas

p = reading of a gas thermometer filled with the given gas. We see that the deviations are very small, only hundredths of one per cent. The error of the hydrogen thermometer is only two-thirds that of the helium thermometer, and only one-sixth that of a thermometer filled with oxygen (or air).

26.7. THE DEGREE AND THE KELVIN.

THE PRACTICAL AND ABSOLUTE TEMPERATURE SCALES

1. The unit of temperature measurement, commonly used in practice, is the *Celsius degree* ($^{\circ}\text{C}$), which is one hundredth of the temperature interval between the melting point of ice and the boiling point of distilled water at the standard pressure, $p_0 = 101\ 325\ \text{Pa}$. The unit of absolute temperature is the *kelvin* (K). The Celsius degree and the kelvin are equal in magnitude. An exact definition of the kelvin will be given in Sec. 36.6.

Thus

$$1\ \text{K} = \frac{T_b - T_m}{100} \quad (26.13)$$

Using this definition of the unit of temperature measurement we can determine experimentally the absolute temperatures for the melting point of ice and the boiling point of water. It follows from equation (26.12) that

$$\frac{p_b}{p_m} = \frac{T_b}{T_m} \quad \text{or} \quad \frac{p_b - p_m}{p_m} = \frac{T_b - T_m}{T_m} \quad (26.14)$$

where p_m = pressure indicated by a gas thermometer submerged in a mixture of water and thawing ice

p_b = pressure indicated by the same thermometer when submerged in boiling distilled water.

According to the test data $p_b/p_m = 1.3661$, then

$$\frac{p_b - p_m}{p_m} = 0.3661$$

Table 26.2

Gas	$\delta = \Delta p/p_{id}$
Oxygen	0.00073
Hydrogen	0.00012
Helium	0.00018

By the definition of the kelvin [equation (26.13)], $T_b - T_m = 100$ K. Substituting into equation (26.14), we can write

$$T_m = 273.15 \text{ K}$$

Therefore, the melting point of ice at standard pressure is 273.15 K; the corresponding boiling point of distilled water at standard pressure is 373.15 K.

2. Having determined the reference (fixed) points of the absolute temperature scale, we can proceed to graduate the gas thermometer. For this purpose we rewrite formula (26.12) as follows:

$$T = 273.15 \frac{h}{h_m} \quad (26.15)$$

where h = height of the mercury column of the gas thermometer at the temperature being measured

h_m = height of the same mercury column at the melting point of ice (this value is determined experimentally for each constant-volume gas thermometer).

This procedure enables us to graduate the scale of a gas thermometer. From the height of the mercury column in the manometer tube (in millimetres) we can readily determine the absolute temperature.

3. The absolute temperature scale is not commonly used to make practical measurements of temperature. More convenient for this purpose is the International Practical Temperature Scale, better known as the Celsius scale. In this scale, the boiling point of water (steam point) has a temperature of 100 °C and the melting point of ice (ice point), 0 °C.

As mentioned before, the degree Celsius and the kelvin have the same magnitude. Only the level from which the temperatures are read has been shifted. A comparison of the two scales is shown in Fig. 26.6. As we can see

$$t = T - 273.15 \quad (26.16)$$

where t = temperature on the Celsius scale
 T = the same temperature on the absolute scale.

For calculations of an accuracy within 0.055 per cent, it may be assumed that

$$t = T - 273$$

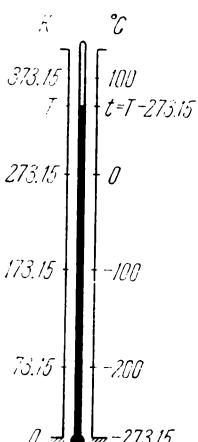


Fig. 26.6

26.8. ABSOLUTE ZERO

1. The zero level of temperature on the absolute scale is called absolute zero. It corresponds to a temperature of -273.15°C on the International Practical Scale.

Absolute zero is the lowest possible temperature. There is no experiment enabling us to obtain a temperature equal to absolute zero. It is all the more impossible to reach temperatures below absolute zero.

The reaching of low temperatures, near to absolute zero, is of exceptional interest because, at such temperatures, the properties of substances are essentially changed. A number of remarkable phenomena are observed, such as the superfluidity of helium, superconductivity, etc. (see Sec. 75.10). Today, by methods too complicated to be dealt with here, a temperature only 0.0001 K above absolute zero has been reached.

2. It should be noted that at temperatures near to absolute zero, the average kinetic energy of the molecules is no longer proportional to the absolute temperature. The reason is that the quantum properties of molecules begin to be manifested at these temperatures. In particular, the average kinetic energy of the molecules does not approach zero as the temperature is lowered, as could be inferred from equation (26.8). It approaches the minimum energy of a localized particle as expressed by equation (16.26).

Hence, at low temperatures, *gases degenerate*, and equation (26.8) is no longer valid. The *degeneracy temperature* can be assessed from the relationship

$$\frac{3}{2}kT \geq \frac{\hbar^2 n^{2/3}}{2m_0}$$

from which

$$T_{deg} = \frac{\hbar^2 n^{2/3}}{2km_0} \quad (26.17)$$

Thus for helium at a concentration $n \approx 10^{25} \text{ m}^{-3}$, we obtain

$$T_{deg} \approx \frac{10^{-68} \times 10^{50/3}}{3 \times 1.38 \times 10^{-23} \times 0.7 \times 10^{-26}} \approx 2 \times 10^{-3} \text{ K}$$

As we can see, the degeneracy temperature of a molecular gas is very low, so that the use of the model of a classical ideal gas is justified, even at very low temperatures. The degeneracy temperature for liquid helium ($n \approx 10^{28} \text{ m}^{-3}$) is

$$T_{deg}^{liq} \approx 10^2 T_{deg}^{gas} \approx 0.2 \text{ K}$$

The electron gas in metals, however, is found to be degenerated at any, even the highest, temperatures (see Sec. 75.5), and its properties differ essentially from those of an ordinary ideal gas.

3. Thus, at temperatures near to absolute zero, the average kinetic energy of the molecules is no longer proportional to the absolute temperature, but is expressed by a considerably more complex function. But the statement, sometimes found in books on popular science, that all molecular motion ceases at absolute zero is incorrect in principle. At the present time scientists have shown that molecular motion never ceases, even at absolute zero, where the molecules have *zero-point vibrations*. These vibrations correspond to *zero-point energy* which is the minimum energy possessed by molecules of the given substance and which cannot be taken away from the body with any cooling procedure, down to absolute zero (see Sec. 70.5). Zero-point vibrations are detected by the scattering of light in a substance at temperatures near to absolute zero.

The concept of an ideal gas makes no sense at absolute zero. Hence the statement, sometimes encountered, that at absolute zero the pressure of an ideal gas is zero, is incorrect. Such an extrapolation of formula (26.10) is invalid. This formula holds only at temperatures far from absolute zero, namely, at temperatures well above the degeneracy temperature.

26.9. AVOGADRO'S NUMBER AND BOLTZMANN'S CONSTANT

1. Equation (26.10) leads to many interesting consequences. First of all it follows that under the same conditions (i.e. at the same pressures and temperatures), the concentration of molecules of different gases is the same. In other words: *unit volumes of various gases under the same conditions contain the same number of molecules* (Avogadro's law).

To compute the concentration of molecules we must know Boltzmann's constant. A number of methods are available for measuring this quantity (see, for instance, Sec. 28.12). The most precise experimental value at present is

$$k = (1.380622 \pm 0.000059) \times 10^{-23} \text{ J/K}$$

Boltzmann's constant being known, we can calculate the number of molecules per cubic metre of gas at standard conditions ($T_0 = 273.15 \text{ K}$ and $p_0 = 1 \text{ atm} = 101\ 325 \text{ Pa}$):

$$N_L = n_0 = \frac{p_0}{kT_0} = \frac{101\ 325}{1.38 \times 10^{-23} \times 273.15} = 2.687 \times 10^{25} \text{ m}^{-3}$$

This quantity is known as *Loschmidt's number*:

$$N_L = (2.68709 \pm 0.00009) \times 10^{25} \text{ m}^{-3}$$

2. Then, if the density of a gas is known, we can calculate the mass of a molecule, recalling that the density is the product of the

mass of one molecule by the molecule concentration. Thus

$$m_0 = \frac{\rho_0}{n_0} = \frac{\rho_0}{2.687 \times 10^{25}}$$

The density of hydrogen under standard conditions is $\rho_0 = 0.0899 \text{ kg/m}^3$, that of oxygen is 1.4291 kg/m^3 and that of helium is 0.1785 kg/m^3 . Then the masses of molecules of hydrogen and oxygen are $3.345 \times 10^{-27} \text{ kg}$ and $53.145 \times 10^{-27} \text{ kg}$, and the mass of an atom of helium is $6.643 \times 10^{-27} \text{ kg}$.

3. By a decision of the Fourteenth General Conference on Weights and Measures, held in Paris in October 1971, another fundamental unit was added to the International System of Units (SI). It is the unit of the *amount of substance*. The amount of substance is the physical quantity determined by the number of specific structural elements (molecules, atoms, ions, etc.) of which the substance consists.

Since the masses of single structural elements (for example, molecules) of various substances differ, then the same amount of various substances have different masses. For instance, 10^{25} molecules of hydrogen and 10^{25} molecules of oxygen are considered to be the same amount of substance, though their masses differ, being 33.45 g and 531.45 g. In this way, the Fourteenth General Conference has finally cleared up the question of mass not being a measure of the amount of substance.

The unit of measurement of the amount of substance is the *mole* (mol) which is the amount of substance of a system which contains the same number of specific structural elements (molecules, atoms, ions, etc.) as there are atoms in 0.012 kg (exactly) of the carbon isotope C¹² (see Sec. 80.1).

Also introduced, besides the mole, are its multiples and submultiples, in particular, the *kilomole* (kmol):

$$1 \text{ kmol} = 10^3 \text{ mol}$$

All further calculations are to be based on the kilomole.

4. The number of atoms contained in one mole (or one kilomole) of matter is called *Avogadro's number*:

$N_A = (6.022169 \pm 0.000040) \times 10^{23} \text{ mol}^{-1} = 6.02217 \times 10^{23} \text{ kmol}^{-1}$
From Avogadro's and Loschmidt's numbers we can find the volume of a mole (*molar volume*) under standard conditions:

$$V_m = \frac{N_A}{N_L} = \frac{6.022 \times 10^{23} \text{ mol}^{-1}}{2.687 \times 10^{22} \text{ litre}^{-1}} = 22.41 \text{ litre/mol} = 22.41 \text{ m}^3/\text{kmol}$$

Here and in the following, the subscript "m" means that the given quantity is given on a per-mole or per-kilomole basis.

The mass of one mole (*molar mass*) is

$$M = \rho_0 V_m = m_0 N_A$$

where ρ_0 = density of the substance under standard conditions
 m_0 = mass of a structural element (molecule, atom, ion, etc.). In the following we shall denote the mass of a kilomole by M . This should not lead to any confusion because the numerical value of this quantity will always be followed by the unit: kg/mol or kg/kmol.

At this point we shall introduce two new dimensionless quantities: the *relative atomic mass* A_r , and the *relative molecular mass* M_r , of a substance. They are defined as the ratios of the average masses of the atom and molecule of a natural mixture of the isotopes of a substance to 1/12 of the mass of an atom of the isotope C¹² (see Sec. 80.1). Thus

$$M_r = \frac{m_0}{1.6603 \times 10^{-27} \text{ kg}}$$

Note that the relative molecular mass is equal in magnitude to the mass of a kilomole of the substance.

5. The product of Avogadro's number by Boltzmann's constant is called the *universal, or molar, gas constant*:

$$R = N_A k = 8.31434 \pm 0.00035 \text{ J/mol-K} = 8.3143 \times 10^3 \text{ J/mol-K}$$

This quantity is used to transform the equation of state (26.10) of a gas as follows:

$$p = nkT = \frac{n}{N_A} N_A kT = \frac{nm_0}{N_A m_0} RT = \frac{\rho RT}{M}$$

Thus, the equation of state will take the form

$$p = \frac{\rho RT}{M} \quad (26.18)$$

Multiplying both sides of the equation by the volume of the gas, and recalling that $m = \rho V$ is the mass of the gas, we obtain the *Mendeleev-Clapeyron equation*:

$$pV = \frac{m}{M} RT \quad (26.19)$$

26.10. MOLECULE DISTRIBUTION IN A FORCE FIELD

1. Suppose that a gas, enclosed in a certain vessel, is placed in a force field, for instance, in a gravitational field (Fig. 26.7). Let us try to find the influence of the gravitational field on the distribution of the gas molecules.

If there were no gravitational field, the molecules would become uniformly distributed in the vessel in the course of time. The molecule concentration n would be the same at all points in the vessel.

Therefore, the density of the gas $\rho = m_0 n$ and the pressure $p = k n T$ would also be the same throughout the vessel.

On the contrary, if there were no thermal motion of the molecules, but only a gravitational field, the molecules would be held by the gravitational forces against the bottom of the vessel. In other words, they would be in a state of minimum potential energy. This is a consequence of the general law: any system of particles always comes to a state of minimum potential energy.

2. We see that the thermal motion of the molecules tends to scatter them as randomly as possible so that they fill the available space as uniformly as possible. A force field, on the contrary, tends to bring as much order as possible in the distribution of the molecules so that they occupy a level with the minimum potential energy. Influenced by these two factors, the gas molecules are distributed throughout the vessel, but their concentration varies at various points. The highest concentration of molecules will be where their potential energy is at a minimum, and the lowest concentration where it is a maximum.

3. As an example we can consider the change in the molecule concentration in the atmosphere of the earth (or other planets). Experiments show that the atmospheric pressure and density of the air decrease with an increase in the altitude above the earth (Table 26.3).

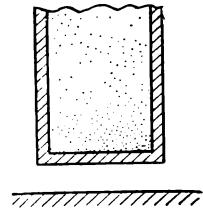


Fig. 26.7

Table 26.3

Altitude above sea level, m	Pressure, Pa	Density, kg/m ³	Temperature, °C	Altitude above sea level, m	Pressure, Pa	Density, kg/m ³	Temperature, °C
0	101 325	1.2250	15.0	8 000	35 600	0.52717	-37.0
1000	89 874	1.1147	8.5	10 000	26 436	0.41271	-50.0
2000	79 495	1.0065	2.0	12 000	19 330	0.31083	-56.5
3000	70 108	0.90913	-4.5	14 000	14 102	0.22675	-56.5
4000	61 640	0.81944	-11.0	16 000	10 287	0.16542	-56.5
5000	54 020	0.73612	-17.5	18 000	7504.8	0.12068	-56.5
6000	47 181	0.65970	-24.0	20 000	5474.9	0.08804	-56.5
7000	41 061	0.58950	-30.5				

Note that in investigating the atmosphere it is necessary to correct for the change in the temperature of the gas with the altitude. This complicates the calculations to some extent.

26.11. BAROMETRIC DISTRIBUTION

1. Let us derive the equation for the distribution of molecules in a force field. For the sake of simplicity, we shall assume the temperature of the gas to be the same at all points.

First we shall introduce the concept of the *probability* of molecules being located at a given point of a force field where the potential energy of the molecules equals U . The probability w is equal to the ratio of the molecule concentration n at the given point of the field to the concentration n_0 at the point of the field where the potential energy equals zero. Thus

$$w = n/n_0 \quad (26.20)$$

2. An analysis of the action of the force field and the thermal motion of the molecules, carried out in the preceding section, shows that the probability of a molecule being located at a given point in the vessel depends upon its potential energy at this point and the temperature of the gas. The higher its potential energy U , the less the probability that the molecule is at this point. The higher the temperature, the greater the probability that the molecules are more uniformly distributed, i.e. the greater the probability for finding the molecule at the given point in the vessel.

Let us introduce the auxiliary variable x , equal to the ratio of the potential energy U of a molecule to the average energy of its thermal motion $\epsilon \cong kT$. Thus $x = U/kT$. Then it turns out that the probability of finding a molecule with the given potential energy is a certain function of this variable:

$$w = f(x) \quad (26.21)$$

The potential energy of a molecule at a certain altitude $h = h_1 + h_2$ is equal to the sum of the energies: $U = U_1 + U_2$. The variable $x = \frac{U}{kT} = \frac{U_1}{kT} + \frac{U_2}{kT} = x_1 + x_2$. The probability is $f(x) = f(x_1 + x_2)$.

One of the laws of probability theory is that the probability of a complex event, consisting of two independent events, is equal to the product of the probabilities of each of the events taken separately. Thus $w = w_1 w_2$ or

$$f(x_1 + x_2) = f(x_1) \times f(x_2) \quad (26.22)$$

It can be shown that the only function satisfying this equation is an exponential function of the form

$$f(x) = a^{-\alpha x} \quad (26.23)$$

This can be checked by making use of one property of an exponential function: in multiplying powers with the same base, the expo-

nents are added together, i.e.

$$a^{-\alpha(x_1+x_2)} = a^{-\alpha x_1} \times a^{-\alpha x_2} \quad (26.24)$$

The base $a > 1$ can be chosen quite arbitrarily. This will only affect the magnitude of factor α . Commonly used is the base of natural logarithms, the irrational number $e = 2.71828\dots$. For $a = e$, the factor $\alpha = 1$.

3. Thus, the probability of finding a molecule at a point in the field where its potential energy equals U is expressed by

$$w = e^{-\frac{U}{kT}} \quad (26.25)$$

The minus sign in the exponent is due to the following. As mentioned previously, the higher the potential energy of a molecule at a given point, the less the probability that it will be found there. Consequently, the required probability is a decreasing function. For a base $a > 1$, an exponential function decreases if its exponent is negative.

4. Combining equations (26.20) and (26.25), we obtain the expression for the molecule concentration:

$$n = n_0 e^{-\frac{U}{kT}} \quad (26.26)$$

Accordingly, we obtain for the gas density

$$\rho = \rho_0 e^{-\frac{U}{kT}} \quad (26.27)$$

and the pressure

$$p = p_0 e^{-\frac{U}{kT}} \quad (26.28)$$

5. If we consider the distribution of molecules in the gravitational field of the earth (or some other planet), then at $h \ll R$, where R is the radius of the planet, we can assume, according to equation (18.21), that $U = m_0gh$, where m_0 is the mass of the molecule, g is the acceleration of gravity and h is the altitude above the surface of the planet. Then the pressure equals

$$p = p_0 e^{-\frac{m_0gh}{kT}} \quad (26.29)$$

A graph of this relationship is shown in Fig. 26.8. It is called the *barometric distribution*.

A similar equation is obtained for the gas density. This result agrees well with the results of test data in investigating the earth's atmosphere.

Precisely barometric distribution is the cause of the lack of an atmosphere on the moon and the highly rarified one on Mars. Owing to the small mass of these celestial bodies, they have relatively weak

gravitational fields. The force of gravity on the moon is 1/6 of that on the earth; on Mars it is 0.38 of the earth's gravity. But, as is evident from equation (26.29), with a small g there is a sharp increase in the probability that the molecules will be at high altitudes

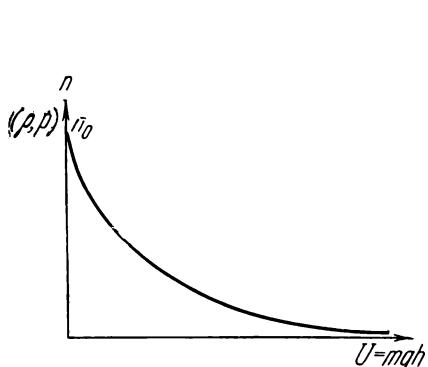


Fig. 26.8

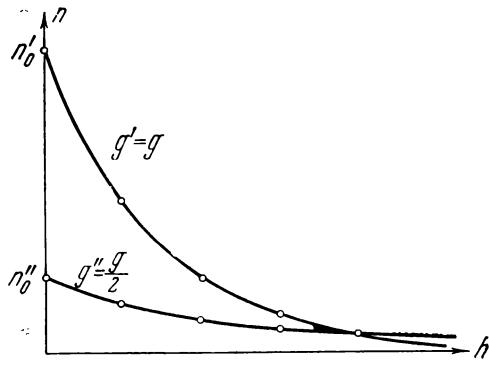


Fig. 26.9

(Fig. 26.9), where the gravitational field is considerably weaker than at the surface of the planet. The velocity of the molecules here may be higher than the escape velocity, and then the gas is dispersed into outer space.

Chapter 27

AN IDEAL GAS AND THE FIRST LAW OF THERMODYNAMICS

27.1. INTERNAL ENERGY OF A MONATOMIC IDEAL GAS

1. An ideal monatomic gas is the simplest of thermodynamic systems, and its internal energy can be readily determined.

Let us recall that we dealt with the molecules of an ideal monatomic gas as if they were point particles whose interaction is so small that it can be neglected. The absence of the forces of interaction means that the potential energy of molecular interaction is constant (Sec. 19.5). The sum of the rest energies of the molecules themselves is also constant, because the molecules do not change in thermal processes. Assuming in equation (20.2) that $\mathcal{E}_0^{(1)} + \mathcal{E}_0^{(2)} + \dots + \mathcal{E}_0^{(N)} = \text{const}$, we find that the *internal energy of an ideal monatomic gas is the sum of the translational kinetic energies of its molecules plus a certain constant.*

It is customary in thermodynamics to denote the internal energy by U rather than \mathcal{E}_0 as in Chapter 20. Hence

$$U = K_1 + K_2 + \dots + K_N + \text{const} \quad (27.1)$$

where K_1, K_2, \dots, K_N are the translational kinetic energies of the separate molecules. Recalling that the average kinetic energy of a molecule is $\bar{\epsilon}_k = (K_1 + K_2 + \dots + K_N)/N$, and, according to equation (26.8), $\bar{\epsilon}_k = \frac{3}{2} kT$, we can write

$$U = N\bar{\epsilon}_k + \text{const} = \frac{3}{2} Nkt + \text{const} \quad (27.2)$$

where N is the total number of molecules in the gas. The constant addend in this equation is often neglected because it plays no role in further calculations.

2. It follows from equation (27.2) that the *internal energy of an ideal gas is fully determined by its absolute temperature*. It does not depend upon the process by means of which the gas reached a state with this temperature.

Accordingly, the change in the internal energy of an ideal gas, in going from a state at temperature T_1 to a state at temperature T_2 , is determined only by its initial and final states, but not by the kind of process by means of which the gas goes from the first to the second state. Thus

$$U_2 - U_1 = \frac{3}{2} Nk(T_2 - T_1) \quad (27.3)$$

27.2. WORK DONE IN THE EXPANSION OF AN IDEAL GAS

1. Suppose that an ideal gas fills some cylinder which is closed by a movable piston (Fig. 27.1). When the piston is displaced an infinitesimal distance Δx , the element of work done is $\Delta W = F\Delta x$. From the definition of pressure we know that $F = pA$, and, since $A\Delta x = \Delta V$ is an infinitesimal change in the volume of the gas, an element of work is

$$\Delta W = p\Delta V \quad (27.4)$$

In expanding, a gas does *positive* work against the external forces ($\Delta V > 0$). When a gas is compressed, the work is *negative* because $\Delta V < 0$, the work being done by the external forces compressing the gas.

2. The work done in changing the volume of a gas by the finite amount $V_2 - V_1$ is found as follows (see Sec. 18.1). First we divide the total change in volume into the elements of volume $\Delta V_1, \Delta V_2, \dots, \Delta V_N$. Next we determine the average pressure at each

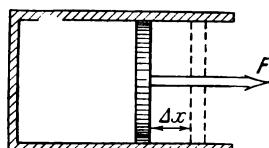


Fig. 27.1

change in volume:

$$P_{av}^{(1)} = \frac{p_1 + p_2}{2}; \quad P_{av}^{(2)} = \frac{p_2 + p_3}{2}; \quad \dots; \quad P_{av}^{(N)} = \frac{p_N + p_{N+1}}{2}$$

Then the total work done is

$$W = \Delta W_1 + \Delta W_2 + \dots + \Delta W_N \\ = P_{av}^{(1)} \Delta V_1 + P_{av}^{(2)} \Delta V_2 + \dots + P_{av}^{(N)} \Delta V_N \quad (27.5)$$

The smaller the chosen element of volume ΔV , the more exactly the calculated value will coincide with the true value of the required work.

3. Work is graphically represented as an area bounded below by the axis of abscissas, above by the pressure vs volume curve and at the sides by the extreme ordinates (Fig. 27.2).

4. It should be noted that the work done when the volume of a gas is changed is determined, not only by the initial and final states of the gas, but also by the *kind of process* by means of which the gas

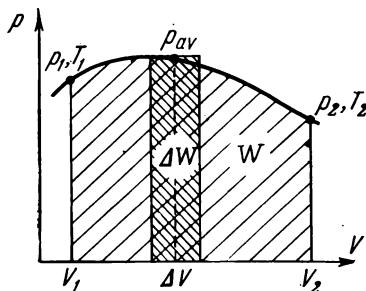


Fig. 27.2

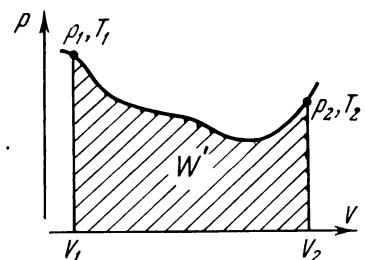


Fig. 27.3

goes from its initial state to its final one. To prove this let us compare Figs. 27.2 and 27.3. Here the initial and final states of the gas are the same in each case, but the work done, W and W' , is different. Obviously $W' < W$ because the intermediate states in the two processes differ.

This means that the force exerted by gas pressure is not a conservative one, and the work done by this force cannot be dealt with as a difference in potential energies (see Sec. 18.6).

27.3. THE FIRST LAW OF THERMODYNAMICS AND THE SPECIFIC HEAT OF A GAS

1. The change in the total energy of a stationary gas reduces to the change in its internal energy: $\Delta E = \Delta U = U_2 - U_1$. Substituting into equation (21.9) we obtain

$$Q = U_2 - U_1 + W \quad (27.6)$$

For a monatomic gas, taking equation (27.3) into account, this expression of the first law of thermodynamics can be written as

$$Q = \frac{3}{2} Nk(T_2 - T_1) + W \quad (27.7)$$

2. The quantity of heat added to a gas consists of two parts. One—the change in internal energy—depends only on the initial and final states of the system and not on the nature of the process by means of which the system is taken from the initial to the final state. The second part—work—depends upon the nature of the process taking place with the gas. Naturally, the amount of heat depends on the process being considered and not only on the initial and final states.

Thus, the quantity of heat and work are quantitative characteristics, not of the state of the system, but of the process it is undergoing. Just as we cannot speak of the "store of work" in a system, there is also no "store of heat". The quantity of heat and work are quantitative characteristics of two different forms in which the energy of a system is changed, and not some special kinds of energy. That is why the term "heat energy", often found in the literature, has no physical meaning.

3. *Specific heat* is the ratio of the quantity of heat added to a system to its mass and change in temperature. Thus

$$c = \frac{Q}{m(T_2 - T_1)} \quad (27.8)$$

where m = mass of the system

T_1 = initial temperature

T_2 = final temperature.

The specific heat is equal in magnitude to the quantity of heat that must be added to a body of unit mass to raise its temperature one degree.

The SI unit of measurement of specific heat is the $\text{J}/(\text{kg}\cdot\text{K})$. Also employed in practice are the nonsystem units $\text{cal}/(\text{g}\cdot^\circ\text{C})$ and $\text{kcal}/(\text{kg}\cdot^\circ\text{C})$:

$$1 \text{ cal/g}\cdot^\circ\text{C} = 1 \text{ kcal}/(\text{kg}\cdot^\circ\text{C}) = 4.1868 \times 10^3 \text{ J}/(\text{kg}\cdot\text{K})$$

Specific heat, naturally, like the quantity of heat, depends on the process by means of which the system goes from the initial to the final state, and not just the states themselves. It is therefore necessary in specifying the specific heat to stipulate the kind of process for which this value was obtained (i.e. measured or calculated).

The specific heats of ideal gases will be calculated in the following sections for various processes.

27.4. ISOCHORIC PROCESSES

1. A process that takes place at constant volume of the gas is said to be *isochoric* (or *isovolumic*). It follows from the equation of state (26.11) of a gas, under the condition $V = \text{const}$, that *in an isochoric process, the pressure of an ideal gas is proportional to its absolute temperature (Charles' law)*:

$$\frac{p_2}{p_1} = \frac{T_2}{T_1} \quad (27.9)$$

2. Evidently, the gas does no work in an isochoric process because at $\Delta V = 0$, the work $W = 0$ too. Therefore, *in an isochoric process, the quantity of heat added to a gas is completely utilized in changing its internal energy*. According to equation (27.7), we obtain

$$Q_V = U_2 - U_1 = \frac{3}{2} Nk (T_2 - T_1) \quad (27.10)$$

Here we have used notation commonly accepted in thermodynamics: if some parameter is constant in the given process, it is taken as the subindex of the quantity being considered. For example, Q_V denotes the quantity of heat added to the system at constant volume, c_V , the specific heat at constant volume, etc.

3. According to equations (27.8) and (27.10), we can define the *specific heat at constant volume* as

$$c_V = \frac{Q_V}{m(T_2 - T_1)} = \frac{3}{2} \frac{Nk}{m} = \frac{3}{2} \frac{k}{m_0} \quad (27.11)$$

because the quotient of the mass of the gas m divided by the number of molecules N is the mass of one molecule: $m_0 = m/N$.

Using equation (27.11) we can calculate the specific heat of any monatomic gas, i.e. gas whose molecules consist of a single atom. These include the inert gases (helium, neon, argon, etc.) and the vapours of metals.

4. It proves convenient, in many cases, to use the *molar specific heats* C_m instead of ones based on the mass. Thus

$$C_{mv} = Mc_V \quad (27.12)$$

where M is the molecular mass of the gas, i.e. the mass of one kilomole. The mass of a kilomole equals the product of the mass of a single molecule by the number of molecules in a kilomole, i.e. by Avogadro's number:

$$M = m_0 N_A \quad (27.13)$$

Substituting into equation (27.12) the values of c_V and M obtained in equations (27.11) and (27.13), we can write

$$C_{mv} = \frac{3}{2} N_A k = \frac{3}{2} R \quad (27.14)$$

The molar specific heat of monatomic gases, obtained by formula (27.14), is in excellent agreement with experimental data over a wide range of temperatures.

27.5. ISOBARIC PROCESSES

1. An *isobaric process* is one that takes place at constant pressure. Under the condition $p = \text{const}$, it follows from the equation of state (26.11) of a gas that *in an isobaric process, the volume of an ideal gas is proportional to its absolute temperature (law of Gay-Lussac)*:

$$\frac{V_2}{V_1} = \frac{T_2}{T_1} \quad (27.15)$$

2. The work done by a gas in isobaric expansion is very simply determined. Since $p = \text{const}$, then

$$W_p = p(V_2 - V_1) \quad (27.16)$$

where V_1 and V_2 are the initial and final volumes of the gas. On a p vs V graph, the work is represented by the area of a rectangle (see Fig. 27.4).

This work can also be expressed in terms of the change in temperature of the gas. Making use of equation (26.11), we can write

$$pV_2 = NkT_2 \quad \text{and} \quad pV_1 = NkT_1$$

Consequently,

$$W_p = Nk(T_2 - T_1) \quad (27.17)$$

3. The quantity of heat added to an ideal gas goes both into increasing the internal energy and into the external work of expansion. According to equation (27.7), for a monatomic gas,

$$\begin{aligned} Q_p &= U_2 - U_1 + W_p = \frac{3}{2} Nk(T_2 - T_1) + Nk(T_2 - T_1) \\ &= \frac{5}{2} Nk(T_2 - T_1) \end{aligned} \quad (27.18)$$

We see that a greater amount of heat is required for isobaric heating of a gas than for isochoric heating in the same temperature interval. This was to be expected because the internal energy of the gas changes by the same amount, but in an isobaric process the gas also does work of expansion, while no work is done in an isochoric process.

4. Evidently, the specific heat at constant pressure is greater than that at constant volume. Thus, for a monatomic gas,

$$c_p = \frac{Q_p}{m(T_2 - T_1)} = \frac{5}{2} \frac{k}{m_0} \quad (27.19)$$

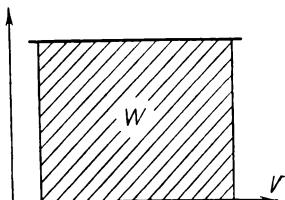


Fig. 27.4

while $c_v = \frac{3}{2} \frac{k}{m_0}$ [see equation (27.11)]. Accordingly, the molar specific heat of such a gas at constant pressure is

$$C_{mp} = \frac{5}{2} R = \frac{3}{2} R + R = C_{mv} + R \quad (27.20)$$

5. The ratio of the specific heats at constant pressure and at constant volume is called the *specific-heat ratio*, or adiabatic exponent:

$$\gamma = c_p/c_v \quad (27.21)$$

Combining equations (27.11), (27.19) and (27.21), we reach a very important conclusion: in all monatomic gases, the specific-heat ratio has the same value. Thus

$$\gamma = 5/3 = 1.67 \quad (27.22)$$

Experiments, conducted in a wide range of temperatures in which a monatomic gas can be dealt with as an ideal gas, confirm this result beyond any doubt.

27.6. ISOTHERMAL PROCESSES

1. As stated in Sec. 26.1 an isothermal process is one that takes place at constant temperature of the gas. It was also stated that the relationship between the pressure and volume of a gas in an isothermal process is expressed by the Boyle-Mariotte law (26.4), and isotherms were plotted on graphs with the coordinate axes p - V and p - ρ .

The work done in an isothermal process is graphically represented by the area under the p - V curve. It is impossible to calculate this work by elementary methods, but for specified values of the pressures and volumes, the numerical value of the work can be found by the method given in Sec. 27.2.

We leave it as a problem for the student to calculate the work done in an isothermal process. The answer can be checked by the following formula, given without proof:

$$W_T = 2.3 p_1 V_1 \log (V_2/V_1) = 2.3 NkT \log (V_2/V_1)$$

2. The internal energy of the gas does not change in an isothermal process. This is evident from equation (27.3) since $T_2 = T_1 = T$. Sometimes, this fact leads to the erroneous conclusion that no heat need be added to the gas in an isothermal process because its temperature does not change. This is entirely wrong! It follows from the first law of thermodynamics (27.6) that

$$Q_T = W_T \quad (27.23)$$

i.e. it is necessary, in an isothermal process, to add a certain amount of heat to the gas, exactly equal to the work done by the gas.

3. Since heat is added to a gas in an isothermal process without changing its temperature, the concept of a specific heat at constant temperature is, generally speaking, meaningless. It is necessary to substitute $T_2 - T_1 = 0$ into the denominator on the right side of equation (27.8), and division by zero results in infinity.

However, if the process is almost isothermal, i.e. it takes place at almost constant temperature, then the specific heat of the gas will be a very large number, because a fraction increases without limit if its denominator approaches zero. In this sense it is said that the specific heat of a gas is infinitely great in an isothermal process.

27.7. ADIABATIC PROCESSES

1. As defined in Sec. 21.5 an adiabatic process is one taking place in a system which has no heat exchange with its environment. According to this definition, $\Delta Q_{ad} = 0$. It follows [see equation (27.8)] that the adiabatic specific heat of a gas equals zero: $c_{ad} = 0$.

For a process to be adiabatic, it is necessary to thermally insulate the gas, i.e. to enclose it in some kind of adiabatic shell (like a Dewar flask). Another possibility is to perform the process so quickly that there is no time for heat exchange between the gas and the environment. This is quite feasible, because heat exchange is much slower than compression or expansion of the gas.

2. In adiabatic expansion the gas does work at the expense of its internal energy (since there is no heat exchange with the environment). In adiabatic compression, the internal energy of the gas is correspondingly increased by the amount of work done on the gas by external forces. Since the internal energy of an ideal gas is fully determined by its temperature, the *temperature of an ideal gas increases in adiabatic compression and decreases in adiabatic expansion*.

This principle is extensively applied in practice. If we open the valve of a cylinder containing compressed carbon dioxide (at a pressure of about 40 atm), the gas begins to expand adiabatically. At this its temperature drops to 80°C below zero, and a part of the gas is converted into a crystalline substance resembling snow. This substance is compacted to form what we call dry ice.

The principle of a diesel internal combustion engine consists in rapid adiabatic compression of the air in the cylinders of the engine, raising its temperature to 500° or 600°C. This leads to self-ignition of the liquid fuel which is injected into the cylinder at the end of compression.

3. Since there is no heat exchange with the environment in an adiabatic process, in calculating the work it is necessary to take $Q_{ad} = 0$ in equation (27.6). According to equation (27.7), we can

write for a monatomic gas

$$W_{ad} = \frac{3}{2} Nk (T_1 - T_2) \quad (27.24)$$

where T_1 and T_2 are the temperatures at the beginning and end of the process. Making use of the equation of state (26.11) of a gas, we obtain

$$W_{ad} = \frac{p_1 V_1 - p_2 V_2}{\gamma - 1} = \frac{3}{2} (p_1 V_1 - p_2 V_2) \quad (27.25)$$

since, according to equation (27.22).

$$\gamma - 1 = \frac{5}{3} - 1 = \frac{2}{3}$$

4. Let us find how the pressure of a gas varies when its volume is changed adiabatically, and plot the adiabatic curve (or, adiabat) on a graph with the coordinates p - V (Fig. 27.5).

Let the gas be in a state specified by the volume V_0 , pressure p_0 and temperature T_0 . Then the gas is allowed to expand to the volume V_1 .

If the gas expands isothermally, its pressure becomes $p_1^{isoth} = p_0 V_0 / V_1$. If, however, the expansion is adiabatic, then the pressure drops much more sharply because, in an adiabatic process, in addition to the increase in volume, there is also a drop in temperature. From the equation of state for the gas $p_0 V_0 / T_0 = p_1 V_1 / T_1$ we obtain

$$p_1^{ad} = \frac{p_0 V_0}{V_1} \frac{T_1}{T_0} = p_1^{isoth} \frac{T_1}{T_0}$$

But, since $T_1 < T_0$ in adiabatic expansion, then $p_1^{ad} < p_1^{isoth}$. For this reason, the adiabat in Fig. 27.5 drops below the isotherm.

Following a similar line of reasoning for the case in which a gas is compressed from a state with the variables p_0 , V_0 and T_0 to one with p_2 , V_2 and T_2 , we shall find that $p_2^{ad} > p_2^{isoth}$, owing to the rise in temperature, and that the adiabat in the graph will be above the isotherm.

It is evident from the graph that the gas does less work in adiabatic expansion than in an isothermal process (Fig. 27.6). However, in adiabatic compression, the external forces do more work on the gas than in isothermal compression (Fig. 27.7).

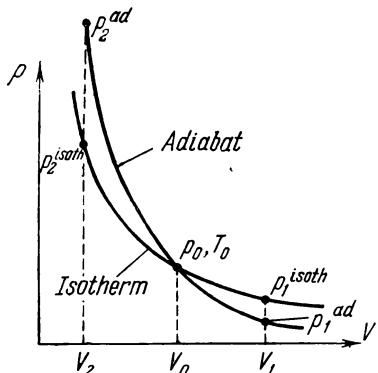


Fig. 27.5

5. The relationship between the pressure and volume of an ideal gas in an adiabatic process is expressed by *Poisson's equation*, which

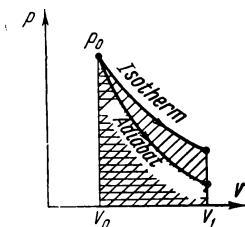


Fig. 27.6

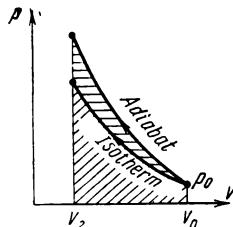


Fig. 27.7

we give without proof:

$$p_1 V_1^\gamma = p_2 V_2^\gamma \quad (27.26)$$

27.8. SPECIFIC HEAT OF A DIATOMIC GAS

1. It was found that the specific heats of diatomic gases (hydrogen, oxygen, nitrogen, etc.), as well as those of polyatomic gases (carbon dioxide, methane, etc.) are considerably higher than the specific heats of monatomic gases. The specific heats, moreover, vary at different temperatures, while those of monatomic gases are independent of the temperature. Though we cannot go into the theory of the specific heats of polyatomic gases in any detail, we shall try to find the dependence of the specific heat of diatomic gases on the temperature, using hydrogen and oxygen as examples. Then we shall set forth the main ideas of the quantum theory of this phenomenon.

2. At standard pressure the boiling point of hydrogen is 20.45 K. Obviously, hydrogen cannot be considered an ideal gas in the vicinity of this temperature. However, at higher temperatures, beginning with about 30 K, this gas at standard pressure will already behave like an ideal gas.

The specific heat of hydrogen in the temperature range from about 30 to 80 K does not differ from that of a monatomic gas. As the temperature is raised, the specific heat gradually increases and the specific heat ratio decreases. The corresponding experimental and calculated data are listed in Table 27.1.

3. The boiling point of oxygen at standard pressure is 90.25 K. Beginning with temperatures of the order of 100 K, oxygen behaves, at standard pressure, like an ideal gas with a molar specific heat at constant pressure $C_{mp} \cong \frac{7}{2} R$, and a specific heat ratio $\gamma \cong 1.4$. The specific heat increases with the temperature (see Table 27.2).

Table 27.1

K	Experimental data		Calculated data		
	c_{mp} , $\frac{\text{kcal}}{\text{kmol}\cdot\text{K}}$	γ	$\frac{c_{mp}}{R}$	$\frac{c_{mV}}{R}$	$\frac{c_{mp} - c_{mV}}{R}$
50	4.96	1.66	2.51	1.51	1.00
100	5.48	1.57	2.76	1.76	1.00
273	6.86	1.41	3.46	2.46	1.00
773	7.06	1.39	3.56	2.56	1.00
1273	7.49	1.36	3.76	2.77	0.99

Table 27.2

K	Experimental data		Calculated data		
	c_{mp} , $\frac{\text{kcal}}{\text{kmol}\cdot\text{K}}$	γ	$\frac{c_{mp}}{R}$	$\frac{c_{mV}}{R}$	$\frac{c_{mp} - c_{mV}}{R}$
150	6.96	1.40	3.50	2.50	1.00
273	7.01	1.39	3.53	2.53	1.00
773	8.02	1.33	4.04	3.05	0.99
1273	8.59	1.29	4.32	3.34	0.98

The curves in Fig. 27.8 show the dependence of the specific heats at constant volume of hydrogen and oxygen on the temperature. The specific heat of a monatomic gas is also shown for comparison.

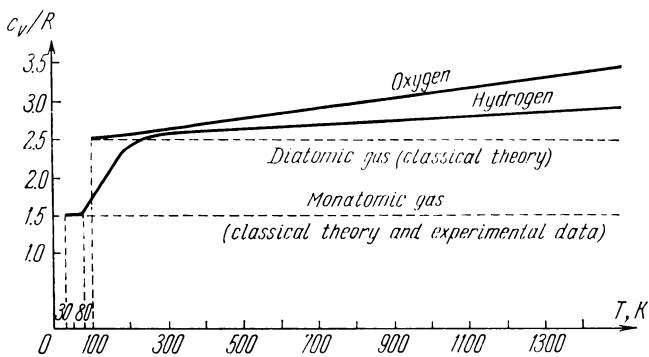


Fig. 27.8

4. As for a monatomic gas, the specific heat at constant volume indicates the quantity of heat that goes to change the internal energy of the gas. But at the same temperature, the average translational

kinetic energies of monatomic and polyatomic gases coincide. Consequently, the internal energy of polyatomic gases is the sum of the energy, not only of translational motion of the molecules, but of other kinds of motion as well: *rotational motion of the molecules and vibration of the atoms in the molecules.*

It was assumed in classical molecular theory that a molecule of a diatomic gas is a rigid system of two spheres between which the distance is constant (like a dumbbell). Such a molecule can have translational motion along three arbitrary coordinate axes and can rotate about two of them, for instance, Oy and Oz (Fig. 27.9).

Rotation about axis Ox does not appreciably increase the energy of the molecule because of its very small moment of inertia with respect to this axis. Therefore, the kinetic energy of a molecule is the sum of three translational kinetic energies and two rotational kinetic energies:

$$\bar{\varepsilon}_k = \frac{m\bar{v}_x^2}{2} + \frac{m\bar{v}_y^2}{2} + \frac{m\bar{v}_z^2}{2} + \frac{I\bar{\omega}_y^2}{2} + \frac{I\bar{\omega}_z^2}{2} \quad (27.27)$$

It is proved in classical molecular theory that each of these motions contributes the same energy, on an average, equal to $kT/2$. Thus

$$\frac{m\bar{v}_x^2}{2} = \frac{m\bar{v}_y^2}{2} = \frac{m\bar{v}_z^2}{2} = \frac{I\bar{\omega}_y^2}{2} = \frac{I\bar{\omega}_z^2}{2} = \frac{1}{2} kT$$

Substituting the last equation into the preceding one, we obtain:

$$\bar{\varepsilon}_k = \frac{5}{2} kT \quad (27.28)$$

From this it follows that the internal energy of a kilomole of a diatomic gas is

$$U_m = \frac{5}{2} N_A kT + \text{const} = \frac{5}{2} RT + \text{const} \quad (27.29)$$

The molar specific heat at constant volume equals

$$C_{mv} = \frac{5}{2} R \quad (27.30)$$

5. Comparing this last result with experimental data (see Fig. 27.8) we can see that the numerical values of the specific heats of diatomic gases does not agree with the theoretical one obtained by equation (27.30). For oxygen, the deviation ranges from about

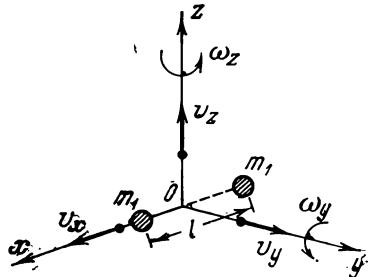


Fig. 27.9

20 to 30 per cent. But the classical theory has even more essential shortcomings. It is unable to explain the following:

(1) Why does the specific heat of diatomic gases (and, in general, polyatomic gases) increase with the temperature?

(2) Why, at low temperatures, does the specific heat of hydrogen coincide with that of monatomic gases (i.e. $C_{mV} = \frac{3}{2}R$ and not $\frac{5}{2}R$ as at higher temperatures)?

(3) Why is this effect observed for hydrogen but is absent in oxygen?

(4) Why do diatomic molecules behave like rigid systems? Why isn't vibration of the atoms in the molecule excited in collisions?

Answers were found to these questions only on the basis of the quantum theory of specific heat.

27.9. QUANTUM THEORY OF THE SPECIFIC HEATS OF GASES

1. The energy of a molecule of a diatomic gas can be defined as the sum of the kinetic energies of its translational and rotational motions, and the energy of the vibration of the atoms within the molecule. Thus

$$\bar{\varepsilon} = \bar{\varepsilon}_{tran}^{kin} + \bar{\varepsilon}_{rot}^{kin} + \bar{\varepsilon}_{vib} = \frac{p^2}{2m_0} + \frac{L^2}{2I} + \bar{\varepsilon}_{vib} \quad (27.31)$$

where p = linear momentum of the molecule

L = its angular momentum

m_0 = mass of the molecule

I = its moment of inertia.

Any change in the internal energy of the gas reduces to changes in translational and rotational energy of the molecules and in the vibrational energy. However, according to quantum theory, these three components of the energy of a molecule are differently affected by a change in the temperature of the gas.

2. A change in the translational kinetic energy reduces to a change in the linear momentum of the molecule. Naturally, this change cannot be less than the uncertainty of the momentum which can be determined from the relation (14.5). Setting the uncertainty of the coordinate in this relation equal in order of magnitude to the mean free path of the molecule $\Delta l \cong \lambda \cong 10^{-7}$ m (see Sec. 25.3) we obtain the uncertainty of the momentum:

$$\Delta p \cong \hbar/\Delta l \cong 10^{-34}/10^{-7} = 10^{-27} \text{ kg}\cdot\text{m/s}$$

This corresponds to the following uncertainty in the value of the kinetic energy of a molecule:

$$\Delta \varepsilon_{tran} = \frac{(p + \Delta p)^2}{2m_0} - \frac{p^2}{2m_0} = \frac{2p\Delta p + \Delta p^2}{2m_0} \gg \frac{3\Delta p^2}{2m_0}$$

The uncertainty in energy corresponds to an uncertainty in temperature ΔT_{tran} which can be found by the relationship

$$\Delta \epsilon_{tran} = \frac{3}{2} k \Delta T_{tran}$$

Hence

$$\Delta T_{tran} = \frac{\Delta p^2}{m_0 k} \quad (27.32)$$

The mass of a molecule of hydrogen is $m_0 \cong \frac{2 \text{ kg}}{6.02 \times 10^{26}} \cong 3.32 \times 10^{-27} \text{ kg}$. Therefore

$$\Delta T_{tran} \cong \frac{10^{-54}}{3.32 \times 10^{-27} \times 1.38 \times 10^{-23}} \cong 2 \times 10^{-5} \text{ K}$$

This means, actually, that the *translational kinetic energy can be changed by any change in the temperature*. Consequently, quantum theory leads, in this case, to the same result obtained by classical statistics.

3. A change in the rotational kinetic energy of a molecule reduces to a change in its angular momentum. It is proved in quantum theory that changes in the angular momentum cannot have any arbitrary value, they must be multiples of Planck's constant. Hence the angular momentum can have only the following sequence of values*:

$$L_1 = \hbar, \quad L_2 = 2\hbar, \dots, \quad L_n = n\hbar$$

Therefore, the rotational kinetic energy of a molecule can only have the corresponding values:

$$\epsilon_{rot}^{(1)} = \frac{L_1^2}{2I} = \frac{\hbar^2}{2I}, \quad \epsilon_{rot}^{(2)} = \frac{L_2^2}{2I} = \frac{4\hbar^2}{2I}, \quad \text{etc.}$$

The minimum change in the rotational kinetic energy of a molecule is $\Delta \overline{\epsilon}_{rot}^{kin} = \epsilon_{rot}^{(2)} - \epsilon_{rot}^{(1)} = 3\hbar^2/2I$. This change in energy corresponds to the change in temperature which follows from equation $\frac{3}{2} k \Delta T_{rot} = \frac{3\hbar^2}{2I}$, i.e.

$$\Delta T_{rot} = \frac{\hbar^2}{Ik} \quad (27.33)$$

4. The moment of inertia of a hydrogen molecule can be determined on the basis of the following. Let us imagine the molecule in the form of a dumbbell (see Fig. 27.9). The distance between the centres of the atoms is $l \cong 0.74 \times 10^{-10} \text{ m}$; the mass of a hydrogen atom is one half of the mass of the molecule and equal to $m_1 = 1.67 \times 10^{-27} \text{ kg}$. Then

$$I_{H_2} = 2m_1 r^2 \cong 5 \times 10^{-48} \text{ kg} \cdot \text{m}^2$$

* This will be dealt with more precisely in Sec. 72.2.

where $r = l/2$. Substituting into equation (27.33) we find that a change in the rotational kinetic energy of a hydrogen molecule takes place if the temperature of the gas is changed by

$$\Delta T_{rot} \cong \frac{10^{-68}}{5 \times 10^{-48} \times 1.38 \times 10^{-23}} \cong 140 \text{ K}$$

Thus, at temperatures of the order of 200 K, the molecules can change their rotational energy in collisions, and therefore $C_{mv} \cong \frac{5}{2} R$.

At temperatures of about 100 K, only a small portion of the molecules, whose energy, according to the Maxwellian distribution (see Sec. 25.2), is above the average value, is capable of changing their rotational kinetic energy in collisions. Therefore, at these temperatures, $\frac{3}{2} R < C_{mv} < \frac{5}{2} R$. At temperatures below 80 K, the percentage of molecules capable of changing their rotational energy in collisions is so small that actually only the translational kinetic energy changes with the temperature. Naturally, at these temperatures, the specific heat at constant volume for hydrogen is the same as for a monatomic gas.

5. The specific heat of oxygen is an entirely different matter. Here the mass of the atom is 16 times as much as that of a hydrogen atom, and the distance between the centres of the atoms in a molecule is $l \cong 1.2 \times 10^{-10} \text{ m}$. Consequently, the moment of inertia of an oxygen molecule is about 40 times as much as that of a hydrogen molecule. According to equation (27.33) the change in the temperature of a gas accompanied by a change in rotational energy is $\Delta T_{rot} \cong \cong 3 \text{ K}$ for oxygen. Since the boiling point of oxygen at standard temperature equals 90 K, then, under the conditions where it can be dealt with as an ideal gas, both the translational and rotational energies of the molecules will increase in collisions. This agrees with the classical theory of the specific heats of diatomic gases.

6. In the preceding discussions we did not take the energy of vibration of the molecules into account, tacitly assuming that it is constant at these temperatures. The reason for this state of affairs is also cleared up by quantum mechanics which shows that, as a first approximation, the vibrational energy is

$$\varepsilon_{vib} = \hbar\omega \left(n + \frac{1}{2} \right)$$

where ω = natural frequency of vibration of atoms in the molecule

n = a whole number: $n = 0, 1, 2, 3, \dots$

Therefore, the minimum change in vibrational energy is $\Delta\varepsilon_{vib} = \hbar\omega$.

The natural frequency of vibration of molecules is determined by optical methods. For hydrogen $\omega_{H_2} = 7.86 \times 10^{14} \text{ rad/s}$, for oxygen $\omega_{O_2} = 2.94 \times 10^{14} \text{ rad/s}$.

Let us calculate the temperature at which the vibrational energy increases due to collisions of the molecules. We obtain $\frac{3}{2} k\Delta T = \Delta \epsilon_{vib} = \hbar\omega$ from which

$$\Delta T_{vib} = \frac{2\hbar\omega}{3k} \quad (27.34)$$

For hydrogen

$$\Delta T_{vib} \cong \frac{2 \times 10^{-34} \times 7.86 \times 10^{14}}{3 \times 1.38 \times 10^{-23}} \cong 3800 \text{ K}$$

and for oxygen

$$\Delta T_{vib} \cong \frac{2 \times 10^{-34} \times 2.94 \times 10^{14}}{3 \times 1.38 \times 10^{-23}} \cong 1400 \text{ K}$$

Consequently, in a temperature range of the order of 1300 to 1500 K, the vibrational energy in a molecule of hydrogen changes only slightly, and the specific heat of hydrogen does not increase as rapidly as that of oxygen (see Fig. 27.8).

In the oxygen molecule, on the other hand, the vibrational energy begins to increase at lower temperatures. At temperatures as low as 700 to 900 K, about half of the molecules are capable of increasing their vibrational energy, and therefore $C_{mv} \cong 3R$. This is greater than the classical value, $2.5 R$, for a diatomic gas. At temperatures of the order of 1400 to 1500 K, all of the molecules should change their vibrational energy in collisions.

We can calculate the limit being approached in this case by the specific heat at constant volume of oxygen. We shall use the same rule we applied in Sec. 27.8 to take into account the translational and rotational kinetic energies. In vibrations, both the kinetic and potential energies of the molecules are changed. These changes turn out to be the same, and each to be equal to $kT/2$. Instead of equation (27.28) we obtain

$$\bar{\epsilon}_k = \frac{7}{2} kT \quad (27.35)$$

Then

$$C_{mv} = \frac{7}{2} R, C_{mp} = \frac{9}{2} R \text{ and } \gamma = \frac{9}{7} = 1.29$$

Comparing these values with those listed in Table 27.2 we see that our result agrees with the experimental data.

7. We can see from the preceding discussion that typical features in the behaviour of the specific heats of diatomic gases can be explained by quantum concepts. Classical statistics, on the other hand, is unable to explain all these phenomena, though it enables us, in certain cases, to obtain the limiting values of the quantities we are interested in.

Chapter 28

THE SECOND LAW OF THERMODYNAMICS

28.1. QUASI-STATIC PROCESSES

1. In the preceding chapters, discussing various processes that take place in an ideal gas, we spoke of its temperature, density and pressure. Here we tacitly assumed that the temperature, density and pressure of the gas are the same at all points within the given volume. These quantities are called *thermodynamic state variables*.

Strictly speaking, the equality of thermodynamic variables at all points of a system is possible only if it is in a state of equilibrium.

This condition is violated if any processes are taking place in the system.

Thus, if we push the piston slightly in a cylinder, the gas near the piston is slightly compressed and, in this region, its pressure, density and temperature are increased. In other parts of the gas, its variables have not had sufficient time to change (Fig. 28.1). After a certain time τ , called the *relaxation time*, equilibrium is restored. This means that the thermodynamic variables take new values, but again the same throughout the gas.

Assume that compression takes place so slowly that it takes much more time than the relaxation time. Then the whole process can be represented as the sum of a great number of very small compressions, in which the time of each micro-compression is still greater than the relaxation time. Then, during each micro-compression, the time will be sufficient for equilibrium to be restored, and the whole process can be conceived as the totality of transitions through a great number of equilibrium states.

Processes in which the system passes through a continuous sequence of equilibrium states are said to be *quasi-static*, i.e. seemingly static (from the Latin word *quasi* meaning "as if").

2. The rate of isochoric, isobaric or isothermal processes is in no way limited; these processes can be performed as slow as desired. Therefore, in a small vessel with walls having high heat conductivity, we can make any of these processes approximate a quasi-static process to any degree of accuracy specified beforehand.

It is much more difficult to realize a quasi-static adiabatic process. On the one hand, it must take place quickly enough to avoid appreciable heat exchange with the environment. On the other hand, the time the process takes should be much longer than the relaxation time.

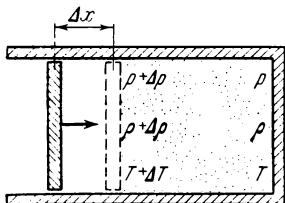


Fig. 28.1

Since these conditions are contradictory, it is difficult to satisfy them simultaneously. Thus processes taking place in practice can only be adiabatic with some approximation if they are to be quasi-static at the same time.

3. Only quasi-static processes are plotted on graphs, because each point of the curve represents some intermediate state of equilibrium. We shall not deal here with the procedure for representing real processes rather than quasi-static ones.

28.2. REVERSIBLE PROCESSES

1. A *reversible process* is one that satisfies the following conditions:

- (a) it can be carried out with equal ease in two opposite directions;
- (b) in each case the system passes through the same intermediate states;
- (c) after carrying out the direct and reverse processes, the system and bodies surrounding it must return to their initial state.

Any process which does not satisfy even one of these conditions, is *irreversible*.

2. *Processes taking place in systems subject to the action of conservative forces (see Sec. 18.5) are reversible.*

Suppose that a perfectly elastic ball drops in a vacuum onto a perfectly elastic plate. Applying the laws of elastic collisions (see Sec. 17.3), we can readily show that the ball, after hitting the plate, bounces back to the initial point, passing, in the reverse direction, through all the intermediate states it passed through in falling. At the completion of the process the ball and all the bodies surrounding it return to their initial state, and this can be repeated as many times as desired.

We can readily see that the vibration of a pendulum in a vacuum or a body on a perfectly elastic spring are also reversible processes.

3. We can, without any particular difficulty, show that *all quasi-static thermal processes are reversible*. As an example, let us consider a quasi-static isothermal process.

Assume that a gas expands quasi-statically and, at the same time, isothermally, and that its volume has increased by the small amount ΔV . Since, in the given process, the system continuously passed through equilibrium states, it is evident that at any instant it could be reversed, i.e. compressed by the same amount ΔV . At this the system would pass through all the same intermediate states that it passed through in expansion.

We know (see Sec. 27.6) that a gas does work in isothermal expansion, receiving a certain quantity of heat, $Q_T = W_T$, from the environment by heat exchange. Suppose that this work was spent in increasing the potential energy of some body of mass m , which was raised to the height h . Then, from the law of conservation of energy,

we can write $Q_T = W_{exp} = mgh$. In the reverse process, the body descends the same height h ; the potential energy of the load is used to do work in compressing the gas again, and the gas gives up to the environment the same quantity of heat it received in expansion.

Thus, after completing the process of expansion of the gas, and then its compression, the gas itself and all the surrounding bodies return to their initial condition, passing through the same intermediate states. Therefore, a quasi-static isothermal process is reversible.

Similar arguments enable the reversibility of any other quasi-static processes to be shown, because in the preceding discussion only the equilibrium method by which the system passed through all intermediate states, i.e. the quasi-static nature of the process, was of importance, and not the kind of heat exchange that occurred between the system and its environment.

28.3. IRREVERSIBILITY OF REAL THERMAL PROCESSES

1. The preceding examples of reversible processes are only idealizations of real phenomena in nature. Actually, there are no strictly conservative systems in nature, because the forces of friction are found in any real system. There are also no strictly quasi-static processes in nature, because all thermal processes take place at a finite rate, and not infinitely slowly. It is clear then that *all real processes in nature are irreversible*.

We should not conclude, however, that the concept of a reversible quasi-static process is an empty one, and therefore unnecessary. Similar to such idealizations as the point particle, point charge, conservative system, etc., the concept of a reversible process is found to be a very convenient idealization of real processes. In many cases it can simplify the solution of a concrete problem. It is advisable, however, first to carefully check to what degree the application of such an idealization is valid, and to what degree the results of the idealized calculations approximate experimental data.

Let us consider several examples that demonstrate the irreversibility of real thermal processes.

2. It was found that in diffusion the concentration is equalized spontaneously without external influence. But such a process will never reverse itself. No matter how long we wait, a mixture of gases will never divide by itself into the initial components, nor will a substance (salt, sugar, paint, etc.) dissolved in a liquid separate out of the liquid by itself.

Of course, any mixture can be divided into its initial components. But, in the first place, the system will not pass again through the same intermediate states it passed through in the diffusion process.

In the second place, the system cannot be returned to its initial state without substantially changing the properties of the bodies that surround it. Thus in separating a gas mixture, obtained by diffusion, into its initial components (see Sec. 25.6), we must expend energy to drive the pumps. In the same way, in separating salt out of an aqueous solution by distillation, we expend energy to evaporate the water. This is associated with a change in the state of the bodies surrounding the given system.

Hence, diffusion is a one-way, and thereby an irreversible process.

3. Experiments show that heat exchange, like diffusion, is a one-way process. In heat exchange, energy is always transmitted automatically from a body at a higher temperature to one at a lower temperature. Consequently, heat exchange is always accompanied by an equalization of temperatures. The reverse process of transmitting energy in the form of heat from cold bodies to hotter ones never occurs by itself.

Another one-way process is the transformation of mechanical energy into internal energy in an inelastic collision or friction. The mechanical energy of the colliding or rubbing bodies is converted into internal energy due to which their temperature is raised. However, no matter how long we wait, we will never observe the reverse process: the spontaneous conversion of internal energy into mechanical energy.

28.4. IRREVERSIBILITY AND STATISTICS

1. On the face of it, the irreversibility of thermal processes would seem to be a paradox. Indeed, all thermal processes, in the final analysis, reduce to mechanical ones: the motion and interaction of molecules. But mechanical processes are reversible; why are thermal processes irreversible? This apparent contradiction can be resolved by means of molecular statistics, in which techniques of probability theory are applied to study processes in systems made up of a huge number of particles. To introduce the ideas of molecular statistics we shall first consider the simpler example of a model consisting of a vessel containing two different kinds of balls.

2. Suppose that two layers of balls of the same size and mass, but painted in two colours, lie on the bottom of the vessel: ten black balls below and ten white balls above (Fig. 28.2). If we shake the vessel, the balls are mixed together and in each row we shall have a set of white and black balls, for example, as shown in Fig. 28.3. No matter how much we continue to shake the vessel, the balls will practically never return to their initial state. Thus, the *process of mixing balls is irreversible*—in shaking the vessel an ordered arrangement spontaneously goes over to a disordered one, and the reverse process practically never occurs by itself.

Why is this process irreversible? How do the balls "know" that they have to arrange themselves at random and not in order? We

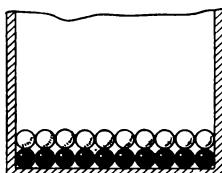


Fig. 28.2

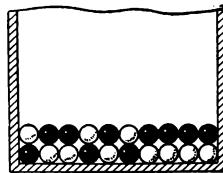


Fig. 28.3

can answer these questions if we can calculate the number of possible combinations in which ten white and ten black balls can be arranged in two rows.

3. To simplify calculations let us designate the balls, for instance, the white ones by capital letters, and the black ones by lower-case letters:

White	A, B, C, D, E, F, G, H, I, J
Black	a, b, c, d, e, f, g, h, i, j

Let there be three white balls and seven black balls in the top row (Fig. 28.3). Naturally, there will be the opposite amounts in the bottom row—seven white and three black. Thus the choice of some distribution in the top row will automatically determine their distribution in the bottom row. Let us calculate how many combinations of balls correspond to this distribution. From ten white balls we can choose three balls in different ways: ABC, ABG, EFJ, etc. Their number equals the number of possible combinations of ten balls in threes. This number is denoted by C_{10}^3 and equals

$$C_{10}^3 = \frac{10 \times 9 \times 8}{1 \times 2 \times 3} = 120$$

In a similar way we can calculate the number of combinations of seven black balls out of ten:

$$C_{10}^7 = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4}{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7} = \frac{10 \times 9 \times 8}{1 \times 2 \times 3} = 120^*$$

Since the black balls can be chosen independently of the white balls (the only requirement being that their total number is ten), the total number of combinations for the distribution of three white

* Note that in all cases $C_m^n = C_m^{m-n}$. We leave it to the student to check this property of combinations by examples. A general proof is given in mathematic courses.

and seven black balls is equal to the product of the combinations of white and black balls, i.e. $C_{10}^3 \times C_{10}^7 = (120)^2 = 14\ 400$.

In a similar manner we can calculate the number of possible combinations of an arrangement with two white and eight black balls, with four white and six black balls, etc. The results of these calculations are listed in Table 28.1.

Table 28.1

State No.	Balls in top row		Number of combinations corresponding to given state	Percentage
	Black	White		
1	10	0	$C_{10}^{10} \times C_{10}^0$	= 1
2	9	1	$C_{10}^9 \times C_{10}^1 = (C_{10}^1)^2 = 10^2$	= 100
3	8	2	$C_{10}^8 \times C_{10}^2 = (C_{10}^2)^2 = \left(\frac{10 \times 9}{1 \times 2}\right)^2$	= 2 025
4	7	3	$C_{10}^7 \times C_{10}^3 = (C_{10}^3)^2 = \left(\frac{10 \times 9 \times 8}{1 \times 2 \times 3}\right)^2$	= 14 400
5	6	4	$C_{10}^6 \times C_{10}^4 = (C_{10}^4)^2 = \left(\frac{10 \times 9 \times 8 \times 7}{1 \times 2 \times 3 \times 4}\right)^2 = 44\ 100$	23.87
6	5	5	$C_{10}^5 \times C_{10}^5 = \left(\frac{10 \times 9 \times 8 \times 7 \times 6}{1 \times 2 \times 3 \times 4 \times 5}\right)^2 = 63\ 504$	34.37
7	4	6	$C_{10}^4 \times C_{10}^6 = (C_{10}^6)^2$	= 44 100
8	3	7	$C_{10}^3 \times C_{10}^7 = (C_{10}^7)^2$	= 14 400
9	2	8	$C_{10}^2 \times C_{10}^8 = (C_{10}^8)^2$	= 2 025
10	1	9	$C_{10}^1 \times C_{10}^9 = (C_{10}^1)^2$	= 100
11	0	10	$C_{10}^0 \times C_{10}^{10}$	= 1
Total			184 756	100

4. Next we shall shake the vessel and, after each shake-up, photograph the redistribution of balls that is obtained. If we take a very large number of photographs, for instance, a million, there will be 5 or 6 showing all the white balls in the bottom row and the black ones in the top row. About the same number of photographs will show the opposite arrangement. But the number of photographs with an arrangement of five white and five black balls in each row will be about 344 000!

Therefore, a system of ten white and ten black balls can be in eleven different states. Each of these states, however, can be realized by a different number of combinations of the balls.

The thermodynamic probability of a given state of a certain system of bodies is the number of combinations of the separate elements of the system by means of which the state is realized.

We see in Table 28.1 that the highest thermodynamic probability corresponds to a uniform distribution of black and white balls in the two rows (state No. 6) and states Nos. 5 and 7, which are near to the uniform distribution. The lowest thermodynamic probability corresponds to a completely ordered arrangement of the balls (states Nos. 1 and 11), as well as to states Nos. 2 and 10, which are near to complete order.

5. It can be shown that the time a certain system is in a given state is proportional to the thermodynamic probability of that state. This clears up the question of why the process of mixing balls is irreversible. When the vessel is shaken, the balls become randomly arranged and, in principle, any combination of them may appear. But the most commonly obtained arrangements are states having a high probability. If we shake the vessel with the balls continuously for 24 hours, the system will be in state No. 6 for about 8 hours, in state No. 5 or No. 7 for about 6 hours, and in state No. 1 or No. 11 for only about 0.5 second!

Hence, the irreversibility of the process of mixing balls is determined by the thermodynamic probability of the states in which the given system can exist. The process of transition from a state with a low thermodynamic probability to one with a higher probability is spontaneous.

The reverse process of transition from a disordered (and therefore most uniform) distribution of the elements of a system (balls) to an ordered state practically never occurs by itself precisely because the probability of such a process is entirely negligible.

We obtain a similar result in analyzing the phenomena in any system consisting of a large number of uniform elements, for instance, in a system comprising a great number of molecules moving randomly.

28.5. DIFFUSION AND THERMODYNAMIC PROBABILITY

1. In many aspects, the diffusion phenomenon is similar to the mixing of balls which was discussed in the preceding section. We can

therefore use the same method to explain the irreversibility of diffusion. Let us imagine a vessel divided into halves by an impenetrable partition. The two halves of the vessel are filled with the same number of molecules of different gases (Fig. 28.4). If we now remove the partition we shall obtain spontaneously, due to diffusion, a uniform mixture of the

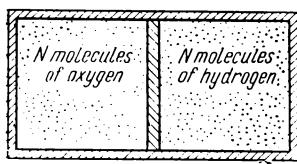


Fig. 28.4

gases. However, no matter how long we wait, the mixture will not separate spontaneously into its initial components.

Let us compare the probability of "molecular disorder", corresponding to a uniform mixture of the gases, with the probability

of "molecular order" corresponding to the initial distribution of the two gases in different halves of the vessel.

2. The initial distribution can be realized by only a single combination; its thermodynamic probability is therefore equal to unity. A uniform mixture can be obtained in a vast number of ways. It is equal to the product of the number of combinations of N molecules of hydrogen with $N/2$ molecules in each half of the vessel by the same number of combinations of N molecules of oxygen with $N/2$ molecules in each half. Thus the thermodynamic probability of this state is

$$W_{unif} = (C_N^{N/2})^2 \quad (28.1)$$

3. Under ordinary conditions there is an enormous amount of molecules, even in small volumes of a gas. For example, one cubic millimetre contains about 10^{16} . At $N \approx 10^{16}$, the thermodynamic probability of a state corresponding to a uniform mixture of the gases is expressed by the immense number

$$W_{unif} \approx 10^{10^{15}}$$

The reason for the irreversibility of diffusion is made clear by this figure. The thermodynamic probability of a state with a uniform mixture is incommensurably greater than one in which the two components are separated. Therefore, a system spontaneously, due to random motion of its molecules, passes over to the most probable state which corresponds to a uniform mixture of the gases. The reverse process in which the mixture separates spontaneously into the initial components is so improbable that it practically never occurs.

28.6. THERMODYNAMIC PROBABILITY IN OTHER THERMAL PROCESSES

1. The concept of thermodynamic probability can be applied to explain the irreversibility of the process of converting mechanical energy into internal energy in an inelastic collision. However, unlike diffusion, it is very difficult to calculate the probability in this case, and we shall restrict ourselves to a qualitative evaluation of this quantity.

Suppose a body travelling at velocity v collides inelastically with a wall. In the first state, all the molecules of the body, though participating in random thermal motion, travel in the same direction at the same velocity v . Following the inelastic collision, the body stops, but its kinetic energy is not lost; it is expended in increasing the internal energy of the body and the wall. Thus

$$K = U_2 - U_1 + Q \quad (28.2)$$

where K = kinetic energy before the collision

$U_2 - U_1$ = change in the internal energy of the body

Q = quantity of heat transferred to the wall, air, etc.

2. The kinetic energy of a body is a measure of a mechanical process, namely, the process of ordered transfer of all the molecules in the same direction at the same velocity. Such a state can be realized in only one way, and its thermodynamic probability is therefore equal to unity.

The conversion of the kinetic energy into internal energy corresponds to a transition from ordered motion to random motion. We cannot readily calculate the number of combinations corresponding to different distributions of the molecules according to all the possible magnitudes and directions of their velocities that characterize random thermal motion. But we can contend with certainty that the number of these combinations is extremely large owing to the immense number of molecules. Consequently, the thermodynamic probability of the second state is expressed by a very large number.

3. Hence, the process of converting mechanical energy into internal energy is a transition from a state with a low thermodynamic probability to a more probable one. This process is spontaneous for an enormous number of molecules. The reverse process of the conversion of internal energy into mechanical energy would imply the transition from a more probable to a less probable state. For a large number of molecules, such a process is so highly improbable that it practically never occurs. This explains the irreversibility of the conversion of mechanical energy into internal energy.

4. The student can readily find on his own that all the arguments used in analyzing the phenomenon of inelastic collisions can also be applied to deal with the process of energy conversion in friction, the heating of conductors by an electric current, etc.

Similar reasoning can clear up the reason for the irreversibility of the heat exchange process. However, we shall not discuss this question here, but will defer it to Sec. 28.8, applying a somewhat simpler line of reasoning.

28.7. THERMODYNAMIC PROBABILITY AND ENTROPY

1. *The thermodynamic probability serves to indicate the direction of thermal processes.* By comparing the probabilities of two states of a thermodynamic system, we can establish, at once, the direction of the process that is feasible in the given system. It will correspond to a transition from a less probable to a more probable state. However, the calculation of the thermodynamic probability is a very complex task, since it is extremely difficult, practically, to determine the number of different combinations of molecules for some state of the system. For this reason, a different physical quantity, introdu-

ced by Rudolf Clausius (1822-1888) and named *entropy* by him, is used in thermodynamic calculations.

Entropy, like thermodynamic probability, is a physical quantity indicating the direction in which processes proceed in nature.

2. L. Boltzmann (1844-1906) found that *entropy is proportional to the logarithm of the thermodynamic probability*. Thus

$$S = K \log W \quad (28.3)$$

where S = entropy

W = thermodynamic probability

K = proportionality factor.

M. Planck (1858-1947) showed that $K=2.3 k$, where k is Boltzmann's constant. If a system goes from one state to another, the entropy varies with the thermodynamic probability. Thus

$$\Delta S = S_2 - S_1 = K \log W_2 - K \log W_1 = K \log (W_2/W_1) \quad (28.4)$$

where the subindex 1 refers to the first state and subindex 2 to the second.

Let us find the principal properties of entropy, this physical quantity that plays such an exceptionally vital role in thermodynamics.

3. To begin with, the *entropy of a system is fully determined by the state the system is in*; it does not depend upon the process by which the system reached this state. In this respect, entropy resembles energy, and differs from the work or quantity of heat. The reason is that each state of a system is determined by a certain number of combinations of the molecules. Consequently, each state corresponds to a definite thermodynamic probability and, according to equation (28.3), each value of the probability corresponds to a definite value of the entropy.

4. Next, the *entropy of a system consisting of two (or several) independent parts is equal to the sum of the entropies of these parts*. In this respect as well entropy is like energy because, in the absence of interaction between two parts of a system, its energy equals the sum of the energies of the parts.

To prove this we shall first calculate the thermodynamic probability. Suppose that the state of the first part of the system is realized by means of $W^{(1)}$ combinations of its molecules, and that of the second by $W^{(2)}$ combinations. Since the two parts are independent, each combination of molecules of the first part can correspond to $W^{(2)}$ combinations of the second part. Then the number of combinations by means of which the whole system can be realized is equal to $W^{(2)}W^{(1)}$. But this is its thermodynamic probability. Thus

$$W = W^{(1)} \times W^{(2)} \quad (28.5)$$

Taking the logarithm of equation (28.5) and multiplying by K , we obtain

$$K \log W = K \log W^{(1)} + K \log W^{(2)}$$

or, according to equation (28.3),

$$S = S^{(1)} + S^{(2)} \quad (28.6)$$

5. Finally, it is evident that *entropy increases in irreversible processes*, such as diffusion or the conversion of mechanical energy into internal energy. It was shown in the preceding sections that the thermodynamic probability increased in these processes. Therefore, according to equation (28.4), the entropy also increases: if $W_2 > W_1$, then $S_2 > S_1$ as well.

28.8. ENTROPY AND HEAT EXCHANGE

1. Suppose a body receives or delivers a certain amount of heat Q in an isothermal process; then its entropy changes. Clausius showed that in this case the change in entropy is

$$\Delta S = S_2 - S_1 = \frac{Q}{T} \quad (28.7)$$

where S_1 = entropy of the initial state

S_2 = entropy of the final state

T = temperature at which heat exchange occurs

Q = delivered or received quantity of heat.

According to the definition of the "quantity of heat" (see Sec. 21.3), the following sign rule is valid: $Q > 0$ if a body receives energy in heat exchange, and $Q < 0$ if the body loses energy, giving it up to another body.

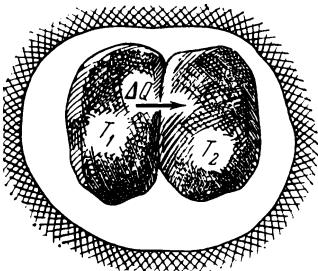


Fig. 28.5

2. Let us calculate the change in entropy of a system of two bodies if there is heat exchange between them. For the sake of simplicity we assume that the bodies are surrounded by an adiabatic shell so that heat exchange with the environment is excluded (Fig. 28.5).

According to equation (28.6) we can write

$$\Delta S = \Delta S^{(1)} + \Delta S^{(2)} \quad (28.8)$$

where $\Delta S^{(1)}$ and $\Delta S^{(2)}$ are the changes in the entropy of the two bodies.

Let the temperature of the first body before heat exchange be T_1 and that of the second, T_2 , with $T_2 < T_1$. As we know from experience, energy (in the form of heat) is always delivered in heat ex-

change from a hotter body to a colder one. Then, according to our sign rule for the quantity of heat, $Q^{(1)} = -Q$ and $Q^{(2)} = Q$, because the first body gives up energy and the second body receives it. As a result of heat exchange, the temperatures of the bodies change and become: $T_1 - \Delta T_1 = T_{av}^{(1)}$ for the first body and $T_2 + \Delta T_2 = T_{av}^{(2)}$ for the second. Then the change in entropy of the whole system, according to equations (28.7) and (28.8) is

$$\Delta S = \frac{Q^{(1)}}{T_{av}^{(1)}} + \frac{Q^{(2)}}{T_{av}^{(2)}} = -\frac{Q}{T_{av}^{(1)}} + \frac{Q}{T_{av}^{(2)}} = \frac{Q(T_{av}^{(1)} - T_{av}^{(2)})}{T_{av}^{(1)} T_{av}^{(2)}} \quad (28.9)$$

But the absolute temperatures $T_{av}^{(1)}$ and $T_{av}^{(2)}$ are positive numbers; according to the conditions for heat exchange $T_{av}^{(1)} - T_{av}^{(2)} > 0$ and $Q > 0$. Thus, in equation (28.9) the right-hand side is a significant positive number. It follows that $\Delta S > 0$, i.e. the *entropy of a closed and adiabatically isolated system increases as a result of heat exchange between its component parts.*

3. Note that the process of heat exchange being considered is *irreversible*: energy (in the form of heat) flows by itself from a hotter body to a colder one, thereby equalizing the temperature in all the parts of the system. The reverse process of heat transfer from a cold body to a hot one never takes place by itself.

Therefore, we see that *in irreversible processes, the entropy of a closed and adiabatically isolated system increases:*

$$\Delta S_{irrev} > 0 \quad (28.10)$$

4. Finally, we shall show that in reversible processes the entropy does not change. For this purpose we shall investigate, as an example, the quasi-static isothermal and adiabatic processes of the expansion or compression of a gas.

Suppose the gas expands in a quasi-static isothermal process. In this expansion the gas does work in lifting a load and receives a certain quantity of heat from the environment. The change in entropy of the whole system is

$$\Delta S = \Delta S_e + \Delta S_g + \Delta S_m \quad (28.11)$$

where the subindexes e , g and m refer to the environment, gas and the mechanical system. But $\Delta S_e = -Q/T$ and $\Delta S_g = Q/T$, because the environment gives up energy (in the form of heat) and the gas gains the same quantity. As to the change in the entropy of the mechanical system, it is equal to zero, because the thermodynamic probability and, consequently, the entropy do not change in any processes in conservative systems. Thus $\Delta S_m = 0$. Substituting into equation (28.11), we can write

$$\Delta S = -\frac{Q}{T} + \frac{Q}{T} = 0$$

It is also evident that in a quasi-static adiabatic process the entropy of the system does not change. Under adiabatic conditions $Q = 0$. It follows, according to equation (28.7), that $\Delta S = 0$.

5. But *quasi-static processes* (both isothermal and adiabatic) are reversible. Consequently, we have shown in these examples that

$$\Delta S_{rev} = 0 \quad (28.12)$$

Hence, we come to a very important conclusion: the *entropy does not decrease in any possible processes*:

$$\Delta S \geqslant 0 \quad (28.13)$$

Here the equal sign refers to reversible, and the greater-than sign to irreversible processes.

28.9. THE SECOND LAW OF THERMODYNAMICS

1. In considering real thermal processes we found that they have a definite direction inherent in them. Namely, they all proceed so that their result is an *equalization of thermodynamic variables*: pressure, temperature, density, chemical composition, etc. At the same time, the direction, and thereby, the irreversibility, of thermal processes in no way follow from the fundamental laws of nature—the laws of conservation of linear momentum, angular momentum and energy (in the form of the first law of thermodynamics).

As a matter of fact, the first law of thermodynamics does not forbid a process in which a certain quantity of heat is transferred from a colder body to a hotter one. The first law makes only one requirement: the quantity of heat given up by one body must exactly equal the quantity received by the other body. In other words, the total energy of a closed and adiabatically isolated system must remain unchanged. But the direction of energy transfer in heat exchange, from the hot to the cold body or vice versa, is not specified by the first law of thermodynamics.

In exactly the same way, the first law does not explain the irreversibility of the conversion of mechanical energy into internal energy in an inelastic collision, friction, etc. The first law makes only one requirement, the sum of the mechanical and internal energies of a closed and adiabatically isolated system must remain constant. But here again the law of conservation of energy allows for the possibility of converting internal energy into mechanical energy just as readily as the opposite. It is impossible to predict the true direction of the process on the basis of the first law of thermodynamics.

2. Thus, we come to the conclusion that direction and, consequently, irreversibility of real thermal processes are not determined by the conservation laws, but by some other law of nature. It is called the *second law of thermodynamics*.

There are various formulations of the second law of thermodynamics. They are equivalent, and any one can be used to derive the others. The most general statement of the law is:

In no process that takes place in a closed and adiabatically isolated macroscopic system does the entropy of the system decrease, i.e. $\Delta S \geq 0$.

3. The second law of thermodynamics is the basis for the principle formulated by Clausius: *no process is possible whose sole result is the transfer of energy in the form of heat from a cold body to a hotter one* (see also Sec. 29.5).

Indeed, were such a process possible, it would be accompanied by a decrease in the entropy (see Sec. 28.8). This would contradict the second law of thermodynamics formulated above.

4. It is necessary to point out the significance of the words "sole result". This means that it is impossible to transfer energy in the form of heat from a cold body to a hot one in the case that no other processes are taking place in nature. If, however, another process (called a *compensating process*, see Sec. 29.4) occurs together with the heat exchange, then the forbidding imposed by the second law of thermodynamics is removed and it becomes possible to transfer energy in the form of heat from a body with a lower temperature to one with a higher temperature. The conditions under which such a process takes place are discussed in Sec. 29.7.

28.10. THE STATISTICAL SENSE OF THE SECOND LAW OF THERMODYNAMICS. FLUCTUATIONS

1. The conservation laws have an unlimited field of application. They are equally valid for microparticles and for macroscopic systems. *The second law of thermodynamics*, on the other hand, is applicable only to macroscopic systems made up of an immense number of molecules.

As we could see from the preceding, the increase in entropy is due to the increase in thermodynamic probability. But molecular statistics by no means excludes, *in principle*, the possibility of a process in which the thermodynamic probability of a certain system decreases, though such processes are extremely rare when there is a large number of molecules in the system. Consequently, we cannot exclude the possibility, at least in principle, of processes in which entropy decreases, no matter how low the probability of their being realized in practice.

2. This leads to the necessity of revising the formulation of the second law of thermodynamics. Taking the ideas of molecular statistics into account, we can state that:

In a closed and adiabatically isolated macroscopic system, the most probable process is one accompanied with an increase in entropy.

In macroscopic systems made up of immense numbers of molecules, the second law of thermodynamics leads to just as reliable results as the conservation laws. But in microsystems consisting of a comparatively small number of molecules, deviations of the true values of physical quantities from their average values are quite possible. Such deviations are called *fluctuations*.

28.11. BROWNIAN MOTION AND FLUCTUATIONS

1. Brownian motion (see Sec. 25.3) is a result of fluctuations in pressure. On the basis of the laws of dynamics it is impossible to explain why Brownian motion is observed with particles approximately 0.5 to 1 micron in size, less with somewhat larger particles, while ones larger than 5 microns never participate in this motion.

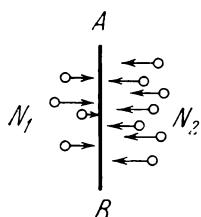


Fig. 28.6

For simplicity let us suppose that a Brownian particle has the form of a long thin thread *AB* (Fig. 28.6).

Let us calculate the probability of the pressure fluctuations we are interested in at various sizes of the particle. We shall assume that the thread is 5×10^{-7} m long and that the distance between molecules of the gas is about $10 \text{ \AA} = 10^{-9}$ m.

Then there will be about 1000 molecules on the two sides of the thread.

If the number of molecules is the same at the right- and left-hand sides of the thread, say $N_1 = N_2 = 500$, the particle will be in equilibrium. But if there are $N_1 = 470$ at the left and $N_2 = 530$ at the right, the pressure at the right will be $530/470 = 1.13$ times higher. In other words, it will increase by 13 per cent. Let us compare the probability of such a fluctuation with that of the equilibrium state.

The thermodynamic probability of an equilibrium state is equal to the possible number of combinations of 1000 molecules with 500 molecules on each side of the thread:

$$W_{eq} = C_{1000}^{500} = \frac{1000 \times 999 \times 998 \times \dots \times 503 \times 502 \times 501}{1 \times 2 \times 3 \times \dots \times 498 \times 499 \times 500}$$

The thermodynamic probability of the fluctuation being considered is

$$W_{fl} = C_{1000}^{470} = \frac{1000 \times 999 \times \dots \times 533 \times 532 \times 531}{1 \times 2 \times 3 \times \dots \times 468 \times 469 \times 470}$$

Their ratio is

$$\begin{aligned} w &= \frac{W_{fl}}{W_{eq}} = \frac{C_{1000}^{470}}{C_{1000}^{500}} \\ &= \frac{(1000 \times 999 \times \dots \times 532 \times 531) (1 \times 2 \times 3 \times \dots \times 469 \times 470 \times 471 \times \dots \times 500)}{(1 \times 2 \times 3 \times \dots \times 469 \times 470) (1000 \times 999 \times \dots \times 531 \times 530 \times \dots \times 501)} \end{aligned}$$

Cancelling like factors we obtain

$$w = \frac{471 \times 472 \times 473 \times \dots \times 498 \times 499 \times 500}{501 \times 502 \times 503 \times \dots \times 528 \times 529 \times 530}$$

2. It is difficult to solve this equation by elementary methods but we can readily determine its limits. For this purpose, we first determine that $471/501 = 0.9401$ and $500/530 = 0.9477$. Therefore

$$0.9401^{30} < w < 0.9477^{30}$$

from which

$$0.157 < w < 0.174$$

More precise calculations yield

$$w \cong 0.171 = 17\%$$

This result is very important, for a Brownian particle about 0.5 micron in size, the probability of a 13 per cent pressure fluctuations is $0.171 \cong 17$ per cent of that of the equilibrium state. Consequently, if such a particle is at rest (equilibrium state) for one hour, for about 10 minutes it will be subject to considerable forces that set it in motion. Assuming that each "jump" takes about one second, we find that a Brownian particle makes about 600 "jumps" per hour.

It will be left as a problem for the student to show that a particle twice as large will make only about 100 jumps per hour.

3. To obtain a similar "jump" of a Brownian particle 5 microns in size, which is surrounded under the same conditions by 10 000 molecules, it would be necessary to have 4700 molecules at the left and 5300 at the right. The fluctuation in pressure would again be 13 per cent. But for the probability we would obtain

$$w = \frac{W_{fl}}{W_{eq}} = \frac{C_{10000}^{4700}}{C_{5000}^{5300}} = 2.2 \times 10^{-8}$$

Hence, the probability of the fluctuation for such a particle is so low that practically it equals zero. To observe even a single "jump", taking about one second, it would be necessary to watch the particle about 4.5×10^7 s or, since 1 year = 3.16×10^7 s, about one and a half years! There is no question then why particles of similar or larger sizes do not participate in Brownian motion.

Though the preceding calculations have been simplified to a great degree, the order of magnitude is quite near to the true one.

28.12. BROWNIAN MOTION AND BOLTZMANN'S CONSTANT

1. Let us try to assess the displacements of a Brownian particle during the time t of observation which is considerably longer than the mean free time τ of a molecule. The mean free time is that required for the molecule to travel over its mean free path. Let us divide the time of observation into a large number of equal intervals $\Delta t = t/N$, so that these intervals are also much longer than the mean

free time. In the following we shall evaluate the magnitude of these intervals.

During the time Δt , the particle moves a distance L , which we shall call a "step" of the particle. Over this path, the particle is subject to an immense number ($\Delta t/\tau \cong 10^{11}$) of collisions with the molecules surrounding it. This means that each subsequent "step" of the Brownian particle is absolutely independent of the direction of the preceding "step" and that the angle between them varies entirely randomly.

Therefore, we are in no way able to predict where the Brownian particle will be located at any instant of time. The motion of a Brownian particle is absolutely chaotic.

2. Moving randomly, the particle sometimes moves away from and sometimes approaches its initial position (Fig. 28.7). And since its

displacement is equally probable in all directions, it may seem that the particle is just "marking time" in the neighbourhood of its initial location. Experiments show, however, that this is not so. A drop of ink, for instance, gradually spreads about in water. The particles of pigment in the ink are the Brownian particles. This means that in a sufficiently long time interval they gradually move away from their initial position. The average distance that a Brownian particle moves away from its initial position during a time interval considerably longer than the mean free time of the molecules is called the *root-mean-square displacement*. The method for calculating this displacement was proposed in 1904-1907 by Einstein and Smoluchowski.

3. To solve this problem, let us refer to Fig. 28.7. Here L is the "step" of the particle. The distances of the particle from its initial position in the sequence of time intervals $\Delta t_1, \Delta t_2, \Delta t_3, \dots, \Delta t_N$ are denoted by $R_1, R_2, R_3, \dots, R_N$ (only $R_1 = L, R_2, R_3$ and R_{12} are designated in the drawing). The respective angles showing the directions of the succeeding displacements of the particle are denoted by $\varphi_1, \varphi_2, \varphi_3, \dots, \varphi_{N-1}$.

Using the law of cosines, we can write the following system of equations:

$$R_2^2 = R_1^2 + L^2 - 2R_1L \cos \varphi_1$$

$$R_3^2 = R_2^2 + L^2 - 2R_2L \cos \varphi_2$$

$$\dots \dots \dots \dots \dots \dots$$

$$R_N^2 = R_{N-1}^2 + L^2 - 2R_{N-1}L \cos \varphi_{N-1}$$

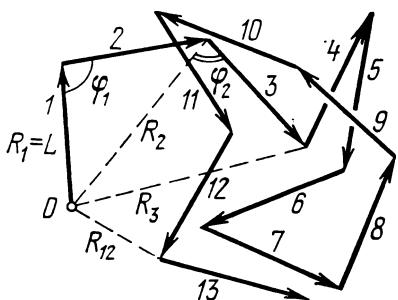


Fig. 28.7

We can readily see that if we add the equations, the squares of the intermediate displacements are cancelled on the left and right sides of the equation that is obtained; only R_N^2 and $R_i^2 = L^2$ are left. Hence

$$R_N^2 = NL^2 - 2L(R_1 \cos \varphi_1 + R_2 \cos \varphi_2 + \dots + R_{N-1} \cos \varphi_{N-1})$$

All the terms in the parentheses are in fact random quantities which may have with equal probability any positive or negative values. Therefore, for a sufficiently large number of observations, the average value of the sum in parentheses equals zero. Now we can write the expression for the root-mean-square displacement which is denoted by Δ^2 , \bar{R}^2 or $\langle R^2 \rangle$:

$$\Delta^2 = NL^2 \quad (28.14)$$

4. Einstein and Smoluchowski found that this quantity can be expressed in terms of the observation time, properties of the liquid and the size of the particles. We shall carry out an approximate, and thereby sufficiently simple, calculation. But first we must evaluate the time interval Δt .

A Brownian particle with a mass of $m \approx 10^{-14}$ kg and radius $r \approx 10^{-6}$ m is an object that is both macroscopic and microscopic at the same time. From the energy point of view it is microscopic because it participates in thermal motion and its average kinetic energy equals the average energy of thermal motion of molecules, i.e. $\frac{3}{2} kT$. But from the point of view of its momentum, the particle

is macroscopic: its linear momentum, $p = \sqrt{2mK} = \sqrt{2m \frac{3}{2} kT}$, is substantially larger than that of a molecule which equals $p_0 = \sqrt{2m_0 \frac{3}{2} kT}$. The ratio of the linear momenta is

$$\frac{p}{p_0} = \sqrt{\frac{m}{m_0}} \cong \sqrt{\frac{10^{-14}}{10^{-28}}} = 10^6$$

Naturally, the collision of a molecule with a Brownian particle resembles the collision of a completely elastic particle with a wall (Sec. 17.3): the molecule bounces off the particle with practically the same velocity in the opposite direction, and the momentum of the Brownian particle is changed by only the very small amount $\Delta p \cong 2p_0$. A very large number of collisions of molecules with a Brownian particle are required for it to "forget" its initial direction of motion and to begin to move in a different direction. The number of such collisions can be taken as

$$N_0 \cong p/p_0 \cong 10^6$$

This enables us to evaluate the time interval Δt during which the particle travels one “step”. Thus

$$\Delta t \cong N_0 \tau \cong 10^6 \times 10^{-10} = 10^{-4} \text{ s}$$

Therefore, during an observation time of the order of $t \cong 1$ hour, the number of “steps” is

$$N = \frac{t}{\Delta t} \cong \frac{10^3}{10^{-4}} = 10^7 \quad (28.15)$$

5. Since a Brownian particle can be regarded as a macroscopic object from the point of view of force action, the total effect of the molecules colliding with it can be equated to the force of viscous friction $F = 6\pi r\eta v$ (Sec. 11.8). But the force of viscous friction slows down a macroscopic object while in the case of a Brownian particle, this force first brakes the particle over the length of a “step” and then accelerates it to its previous velocity in another direction.

Thus, the work done by the average force of viscous friction over the length of a “step” is approximately twice the kinetic energy of thermal motion: $F_{av}L \cong 2\varepsilon_k$.

Substituting the values of the terms we obtain

$$\frac{1}{2} 6\pi r\eta v L \cong 2 \frac{3}{2} kT$$

We multiply both sides of the equation by the observation time and, since $vt \cong NL$, we obtain: $\pi r\eta NL^2 \cong kTt$. Thus

$$\Delta^2 = NL^2 = \frac{kT}{\pi\eta r} t$$

We have thus obtained Einstein's formula. Using it we can determine Boltzmann's constant, and then Avogadro's number (Sec. 26.9) from observations on the behaviour of Brownian particles. Such experiments were conducted from 1908 to 1913 by J. B. Perrin (1870-1942) who obtained the first reliable value of Avogadro's number:

$$6 \times 10^{26} < N_A < 7 \times 10^{26}$$

These classical works played an exceptionally important role in the establishment of statistical physics.

Chapter 29

HEAT ENGINES AND REFRIGERATORS

29.1. HEAT ENGINES AND THE ADVANCEMENT OF ENGINEERING

The history of mankind is closely associated with the development of power engineering. The mastery of a new source of energy, the discovery of new techniques for its conversion and utilization usually constitute a whole age in the history of civilization.

Thus, the vigorous flourishing of industry in the 19th century was associated with the invention of the first heat engine—the steam engine. The development of the internal combustion engine served as the basis for the evolution of automobile transportation and aviation. In literally two decades gas turbines have caused an upheaval in aircraft engineering—low-speed planes with piston engines have been almost completely replaced by jet and turboprop passenger liners flying at speeds approaching that of sound and, in recent years, at supersonic speeds. Space flights, these age-old aspirations of mankind, have at last been achieved by means of jet-propulsion heat engines.

The greater percentage of electric power is produced by steam electric power stations whose generators are driven by steam turbines. In atomic power stations, the energy evolved in nuclear reactions is first converted into steam power which drives a turbine. The turbine, in its turn, drives the rotor of a generator which produces the current.

In elementary school physics courses, students study the layout and principle of various heat engines, such as the steam engine, steam turbine and the internal combustion engine. In this chapter we shall make use of the laws of thermodynamics to investigate the general properties of all heat engines and refrigerators, the purpose of their main units, as well as to find what parameters determine the efficiency of these machines and how the efficiency can be increased.

29.2. THE HEAT ENGINE

1. All heat engines, regardless of their design features, serve the same purpose: the *conversion of internal energy into mechanical energy*. To this end, the energy evolved in the combustion of fuel or in nuclear reactions is transmitted by heat exchange to some gas. Expanding, the gas does work against external forces by driving some mechanism.

2. Obviously, the gas cannot continue to expand without any limit because the engine is of finite size. Consequently, after expanding, the gas must be compressed so that it and all the parts of the engine return to their initial state. Then the expansion can be repeated, etc. Hence, we have reached the conclusion that a *heat engine must operate in cycles*. During a cycle the expansion of the gas is followed by its compression to its initial state.

Real heat engines usually operate on a so-called *open cycle* in which the gas is discharged at the exhaust and a new portion is

compressed. This, however, does not influence the thermodynamics of the process, and in the following we shall deal with a *closed cycle* in which the same portion of gas expands and is compressed.

3. For an engine to do useful work during a cycle it is necessary that the work done in expansion be greater than the work required to compress the gas. Then only will external bodies surrounding the engine obtain more mechanical energy than they give up in compression.

We shall now show that this is possible if the temperature of the gas is lower in compression than in expansion. As proof we shall refer again to Figs. 27.2 and 27.3 (p. 262). The initial and final states of the gas may coincide in expansion and compression. Hence, the work in compression will be less than that in expansion only if the pressure in the compression process is lower than in expansion for all the intermediate states. And this is possible only when the temperature of the gas is lower in compression than in expansion at all intermediate points.

Let us consider Fig. 29.1. Here the work done by the gas in expansion is equal in magnitude to the area under the *p-V expansion curve*. The work done on the gas in compression is the area under the *p-V compression curve*. Finally, the useful work is represented by the hatched area; it is the area bounded by the closed curve of the cycle, i.e. the expansion and compression curves.

29.3. PRINCIPLE AND ENERGY BALANCE OF A HEAT ENGINE

1. Now, on the basis of the preceding, we can discuss the principle and energy processes typical of any heat engine, regardless of its particular design features.

Any heat engine consists of three main parts: the *working medium, or substance*, the *heater (heat source)* and the *cooler (heat sink)*

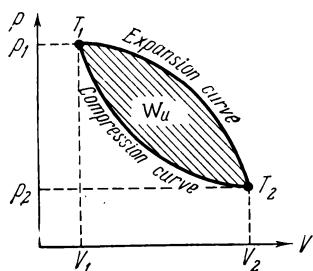


Fig. 29.1

(Fig. 29.2). The working medium is some gas (or steam) which does work by expansion. The working medium receives a certain quantity of heat Q_1 from the *heater*, a body in which a constant high temperature T_1 is maintained by fuel combustion or nuclear reactions. Finally, in combustion, the gas transmits a certain quantity of heat Q_2 to the *cooler*, a body whose temperature T_2 is constant and lower than that of the heater. As a result, the pressure of the gas is lower in compression which, as shown in the preceding section, is a necessary condition for successful operation of the engine.

The cooler may simply be the environment as is the case in internal combustion and jet engines.

2. The energy balance in a cycle can be determined by the first law of thermodynamics [equation (27.6)]. Since the working medium returns to its initial state when the cycle is completed, its internal energy remains unchanged: $\Delta U = U_2 - U_1 = 0$. Then

$$Q_1 = Q_2 + W_u + Q_l \quad (29.1)$$

where Q_l is the energy lost during a cycle as heat exchange with the environment, friction, etc. Therefore, the useful work done by an engine in one cycle is

$$W_u \leq Q_1 - Q_l \quad (29.2)$$

where the less-than sign refers to real engines and the equal sign to an ideal engine in which there are no losses.

3. The efficiency η of a heat engine is the ratio of the useful work done by it to the energy the working medium receives from the heater in one cycle, i.e. to the quantity of heat Q_1 . Thus

$$\eta = \frac{W_u}{Q_1} \leq \frac{Q_1 - Q_l}{Q_1} = 1 - \frac{Q_l}{Q_1} \quad (29.3)$$

Equation (29.3) shows that even an ideal heat engine, operating without losses, has an efficiency which is lower, in principle, than 100 per cent. It could be equal to 100 per cent only if no heat was delivered during the cycle to the cooler, i.e. if Q_2 could be equal to zero. But, as we mentioned previously, this is impossible, because the gas must be cooled in compression. This means that a certain quantity of heat $Q_2 \neq 0$ must be delivered to the cooler.

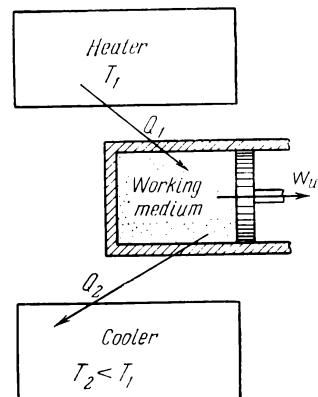


Fig. 29.2

29.4. THE HEAT ENGINE AND THE SECOND LAW OF THERMODYNAMICS

1. An analysis of the principle of operation of heat engines led William Thomson (later Lord Kelvin) (1824-1907) and Planck to a conclusion which can be formulated as follows:

No process is possible whose sole result is the cooling of a heat source and the conversion of the quantity of heat obtained completely into work.

It can be shown that this Kelvin-Planck statement is a consequence of the second law of thermodynamics. Indeed, the conversion of internal energy into mechanical energy is a process in which a system is transformed from a state with random motion of the molecules to one with their ordered motion. But in such a process the entropy of the system must decrease and this disagrees with the second law of thermodynamics. Therefore, the conversion of internal energy into mechanical energy cannot be the *sole* process taking place. It must always be accompanied by another process which leads to an increase in entropy and *compensates for the decrease in entropy* due to the conversion of internal energy to mechanical energy.

2. Though the work done by the isothermal expansion of an ideal gas exactly coincides with the quantity of heat received (see Sec. 27.6), this does not reduce the entropy of the system. The fact is that a parallel process of the expansion of the gas is taking place, and it is accompanied by an increase in entropy. This compensating process removes the restriction imposed by the second law of thermodynamics.

The same is true of the operation of a heat engine. Not only is the heater cooled during the cycle and work is done, but a part of the energy is transmitted to the cooler. It is this compensating process that eliminates the restriction stipulated by the second law of thermodynamics on the conversion of internal energy into mechanical energy.

3. The Clausius principle (see Sec. 28.9), the Kelvin-Planck statement and the entropy increase principle are three equivalent formulations of the second law of thermodynamics.

The first and second laws of thermodynamics are often formulated as laws excluding the possibility of inventing a perpetual motion machine.

First law: a perpetual motion machine of the first kind (*perpetuum mobile I*), i.e. a machine that could do work periodically without receiving energy from other sources, is an impossibility.

Second law: a perpetual motion machine of the second kind (*perpetuum mobile II*), i.e. a machine that could do work periodically only through the cooling of a single source of heat (for instance, the

earth or the ocean, which has immense reserves of internal energy) without transmitting heat to a colder body (for example, the atmosphere) is an impossibility.

29.5. THE CARNOT CYCLE

1. To simplify the analysis of the principle of operation of a heat engine and the calculation of its efficiency, let us consider a thought experiment proposed by S. Carnot (1796-1832).

The working medium in Carnot's idealized engine is an ideal gas. As the cycle starts, the working medium has the same temperature T_1 as the heater. This state is represented in the graph by point 1, (Fig. 29.3). In contact with the heater, the gas expands isothermally, receiving a quantity of heat Q_1 , and reaches state 2. It further expands adiabatically to state 3. In this last process its temperature drops to T_2 , the temperature of the cooler. Expansion is followed by compression of the gas. Being in contact with the cooler, the gas is compressed isothermally by the action of external forces from state 3 to state 4, giving up the quantity of heat Q_2 to the cooler. Then by means of adiabatic compression the temperature of the gas is raised to the initial value. Thus the cycle closes and the working medium is returned to its initial state.

2. The Carnot cycle is an idealization of the cycle of a real heat engine. It is assumed that there are no losses of energy by heat exchange with the environment, that there is no friction in the machine and that processes of gas expansion and compression are quasi-static and therefore reversible. This means that the total change in the entropy of the whole system equals zero.

Now let us see how the entropy of the separate parts of the system changes during the cycle.

The entropy of the working medium naturally remains unchanged because the gas returns to its initial state. Neither does the entropy of the surrounding bodies change because they do not participate in heat exchange, and mechanical processes do not lead to a change in entropy. The entropy of the heater, according to equation (28.7), is reduced by the amount $\Delta S_h = -Q_1/T_1$ and that of the cooler

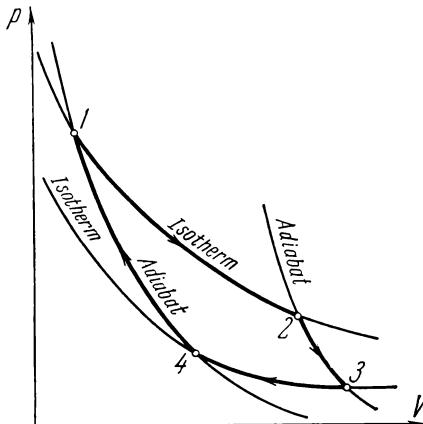


Fig. 29.3

is increased by $\Delta S_c = Q_2/T_2$. Since the change in the entropy of the whole system $\Delta S = 0$,

$$\Delta S = \Delta S_h + \Delta S_c = -\frac{Q_1}{T_1} + \frac{Q_2}{T_2} = 0 \quad (29.4)$$

Hence, $Q_1/T_1 = Q_2/T_2$ or

$$\frac{Q_2}{Q_1} = \frac{T_2}{T_1} \quad (29.5)$$

3. Combining equations (29.3) and (29.5), we find that the efficiency of an ideal engine operating on a reversible Carnot cycle is

$$\eta_{\text{Carnot}} = 1 - \frac{Q_2}{Q_1} = 1 - \frac{T_2}{T_1} = \frac{T_1 - T_2}{T_1} \quad (29.6)$$

This is a result of fundamental importance: the *efficiency of an ideal engine operating on a reversible Carnot cycle is determined only by the temperatures of the heater and cooler.*

29.6. THE EFFICIENCY OF A REAL ENGINE

1. Losses exist in every real engine. Moreover, the processes taking place in it are irreversible. Naturally, the entropy of the system increases during the cycle: $\Delta S > 0$. Equation (29.4) becomes the inequality

$$\Delta S = -\frac{Q_1}{T_1} + \frac{Q_2}{T_2} > 0 \quad (29.7)$$

Therefore, in a heat engine with an irreversible cycle,

$$\frac{Q_2}{Q_1} > \frac{T_2}{T_1} \quad (29.8)$$

2. According to equation (29.3), the efficiency of a real engine is $\eta < 1 - \frac{Q_2}{Q_1}$. If we substitute for the quantity being subtracted, i.e. Q_2/Q_1 , the smaller value T_2/T_1 , the inequality will be even more true. Thus

$$\eta < 1 - \frac{Q_2}{Q_1} < 1 - \frac{T_2}{T_1} \quad (29.9)$$

We see that the *efficiency of a real heat engine is always lower than that of an ideal Carnot engine operating in the same temperature interval, i.e.*

$$\eta < \eta_{\text{Carnot}} \quad (29.10)$$

3. We should note the difference in principle between heat engines and mechanical or electric machines. In improving the design of the latter we try to make their efficiency as near as possible to

the theoretical limiting value of 100 per cent. Though this limit is unattainable under real conditions, we can approach it by reducing losses. When we improve heat engines, we do not try to bring their efficiency to 100 per cent, but to that of a Carnot engine operating in the same temperature interval. It follows that the decisive factor in raising the efficiency of a heat engine is to raise the temperature of the heater and to lower the temperature of the cooler.

Naturally, the reduction of all kinds of losses also raises the efficiency of a heat engine, but the most effective measure is to increase the temperature difference between the heater and cooler.

29.7. THE REVERSE CARNOT CYCLE

1. The Carnot cycle is reversible, and can therefore be run in the opposite direction. Let us see what energy effect is obtained in this case.

Suppose that the working medium, whose state is represented in Fig. 29.4 by point 1, expands adiabatically to the state shown by

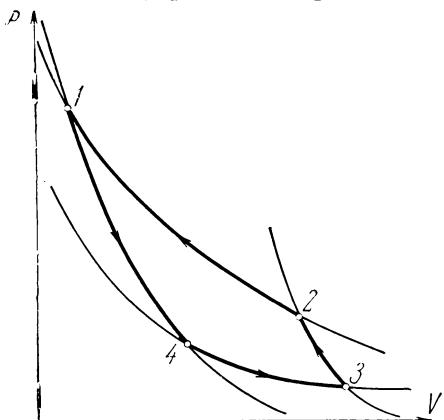


Fig. 29.4

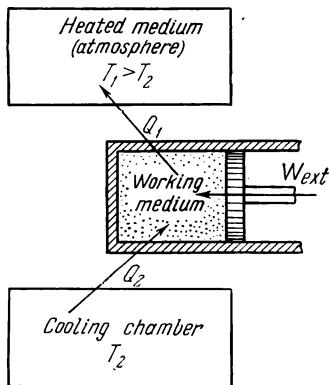


Fig. 29.5

point 4. At this the gas temperature drops to that of the cooling chamber (T_2). Then the working medium expands isothermally to state 3, doing work of expansion and extracting the quantity of heat Q_2 from the cooling chamber. From state 3 the gas is brought to state 2 by adiabatic compression, and its temperature rises to that of the heater (T_1). Finally, the working medium is brought from state 2 to state 1 by isothermal compression, work being done on the gas by external forces. Here the working medium delivers a certain quantity of heat Q_1 to the heater. A schematic diagram illustrating this energy conversion is shown in Fig. 29.5.

2. Since in the reverse cycle, the working medium is compressed at a higher temperature than that at which it expands, compression work is greater than expansion work. Consequently, the positive work done by the external forces during the cycle is

$$W_{ext} = W_{com} - W_{exp} > 0 \quad (29.11)$$

According to equation (27.23), $Q_1 = W_{com}$ and $Q_2 = W_{exp}$. Substituting into the preceding equation we obtain $W_{ext} = Q_1 - Q_2$, or

$$Q_1 = Q_2 + W_{ext} \quad (29.12)$$

3. Hence, in this cycle a certain quantity of heat is transferred from a colder body to one with a higher temperature, i.e. the system acts like a refrigerator. This may appear to contradict the second law of thermodynamics (see Sec. 28.9). But we can readily find that this is not so. The transfer of energy in the form of heat from a colder body to a hotter one is a process leading to a reduction in the entropy. The second law of thermodynamics excludes the possibility of such a process if it is the *sole one* taking place. Here, however, we have another process that takes place in the cycle: the conversion of mechanical energy of the surrounding bodies into internal energy of the heater. This process is accompanied by an increase in entropy.

It was found that the increase in entropy due to the conversion of mechanical energy into internal energy, compensates for the reduction in entropy when heat is transferred from the cooler to the heater. The compensating process eliminates the restriction imposed by the second law of thermodynamics and makes the reverse Carnot cycle possible.

4. Since the Carnot cycle is reversible, the total change in the entropy during a cycle equals zero. Therefore, equation (29.5) holds as well for the reverse Carnot cycle.

If, however, the reverse cycle is made irreversible, then equation (29.7) is changed to some extent. It becomes

$$\Delta S = -\frac{Q_2}{T_2} + \frac{Q_1}{T_1} > 0 \quad (29.13)$$

from which

$$\frac{Q_2}{Q_1} < \frac{T_2}{T_1} \quad (29.14)$$

And again we find that the efficiency of a real cycle is lower than that of the Carnot cycle.

29.8. REFRIGERATORS AND HEAT PUMPS

1. Just as the Carnot engine is an idealization of a real heat engine, the reverse Carnot cycle is the idealization of the cycle of two other thermal devices, the refrigerator and the heat pump.

A refrigerator is used to maintain a temperature in its cooling chamber which is lower than that of the environment. This is possible if the working medium of the refrigerator is put through a reverse cycle similar to the reverse Carnot cycle.

The working medium in a refrigerator is usually the vapour of some low-boiling liquid, such as ammonia or freon. The machine is supplied with energy by an electric power line. This energy is what "pumps the heat" from the cooling chamber to hotter bodies, namely, the environment.

2. The efficiency of a refrigerator is assessed by its *coefficient of performance*:

$$k = \frac{Q_2}{W_{ext}} \quad (29.15)$$

or the ratio of the quantity of heat extracted per cycle from the cooling chamber to the work done by the external forces.

It will be left as an exercise for the student to derive from equations (29.12) and (29.14) the following expression of the coefficient of performance:

$$k < \frac{T_2}{T_1 - T_2} \quad (29.16)$$

where the equal sign refers to a reversible, and the less-than sign to an irreversible cycle.

We should note that the less the difference in the temperatures of the cooling chamber and the environment, the less the mechanical (or electric) energy required to "pump heat" from the colder body to the hotter one. Naturally, the coefficient of performance will be higher, as can be seen from equations (29.15) and (29.16).

Also to be noted is the fact that the coefficient of performance can be higher than 100 per cent. This by no means contradicts the fact that the efficiency of a heat engine is always far below 100 per cent.

3. A refrigerating installation can be used as a *heat pump* for heating purposes.

If a room is heated by ordinary electric heating appliances, the quantity of heat evolved by the heater elements is exactly equal to the electric power consumption. However, if the electric power is used to drive a refrigerating installation, in which the heater is the room being heated and the cooling chamber is the outside atmosphere, more heat is delivered to the room than would be evolved by the direct conversion of electric energy into internal energy by electric heaters or stoves.

Indeed, if there are no losses in the heat pump, the quantity of heat delivered to the room being heated is

$$Q_1 = Q_2 + W$$

where W = electric power consumption

Q_2 = quantity of heat extracted from the outside atmosphere and transferred into the room.

Owing to losses, $Q_1 < Q_2 + W$ in a real installation. But still, notwithstanding the losses, we can obtain that $Q_1 > W$ with a well-designed heat pump.

4. The efficiency of a heat pump is determined by its *coefficient of performance*:

$$K = \frac{Q_1}{W} \quad (29.17)$$

i.e. the ratio of the quantity of heat delivered to the room being heated to the electric power consumption.

It will be left as an exercise for the student to show, on the basis of the preceding equations, that

$$K \leq \frac{T_1}{T_1 - T_2} \quad (29.18)$$

It should be noted that in a reversible cycle the coefficient of heat pump performance is always greater than 100 per cent.

5. Thus, the heat engine, refrigerator and heat pump are three types of thermal devices whose principle of operation is determined by the first and second laws of thermodynamics. Further on we shall come across many other phenomena which we can explain by using these fundamental laws of nature.

Chapter 30

FUNDAMENTALS OF FLUID DYNAMICS

30.1. THERMODYNAMIC PARAMETERS OF MOVING FLUIDS

In this chapter the principal laws of thermodynamics will be applied to study phenomena in moving *fluids*, i.e. gases and liquids.

The state of a moving fluid is determined, not only by its pressure, density and temperature, but by its velocity as well. Therefore, velocity v , pressure p , density ρ and temperature T are the four thermodynamic parameters whose changes determine all processes taking place in a fluid.

Suppose the fluid is flowing in a pipe. Then the thermodynamic parameters will change at each cross section of the pipe (Fig. 30.1).

If the thermodynamic parameters are independent of time at any cross section of the pipe, such motion or flow is said to be *steady*. If the parameters vary with time, the flow is *nonsteady*.

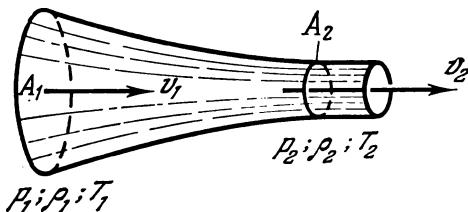


Fig. 30.1.

We shall confine the following discussion to steady fluid flow at velocities much less than that of light ($v \ll c$).

30.2. THE EQUATION OF CONTINUITY

1. The *mass flow rate*, or *mass flux*, is the mass of the substance flowing through a cross section of the pipe in unit time:

$$\mu = \frac{\Delta m}{\Delta t} \quad (30.1)$$

If the velocity of the fluid at a certain cross section of area A equals v , then in the time interval Δt , the volume of fluid passing through this cross section is $\Delta V = Av\Delta t$. The mass of this fluid is $\Delta m = \rho Av\Delta t$. Substituting into equation (30.1) we obtain for the mass flow rate

$$\mu = \rho Av \quad (30.2)$$

2. It follows from the law of conservation of mass that for *steady flow*, the *mass flow rate is the same for any cross section of the pipe*. Indeed, if, for example, more fluid would flow in per second through cross section A_1 in Fig. 30.1 than flows out through section A_2 , then the mass of the fluid within this volume would increase with time. This, in its turn, would lead to increases in the density and pressure of the fluid and would finally change the velocity of flow. But this would not satisfy the conditions of steady flow.

Hence, for steady flow we can write the following equation:

$$\rho_1 A_1 v_1 = \rho_2 A_2 v_2 \quad (30.3)$$

This equation is an expression of the law of conservation of mass in fluid dynamics and is called the *equation of continuity*.

3. The following terms are in use in fluid dynamics. A substance flowing through a pipe is called a *fluid* regardless of its aggregate

state, liquid or gaseous. Distinction is made between a *compressible fluid* (usually a gas) and an *incompressible fluid* which is one of the so-called dropping liquids or even gases in cases when their compressibility can be neglected (see Sec. 30.6).

4. For an incompressible fluid $\rho_1 = \rho_2 = \text{const}$. Hence

$$A_1 v_1 = A_2 v_2 \quad (30.4)$$

30.3. THE MOMENTUM EQUATION

1. In steady flow, the momentum of a certain mass of fluid is a constant value for a given cross section. In passing over to another section, the momentum changes due to the action of pressure forces and forces of internal friction (viscosity).

It is exceptionally difficult to take the forces of internal friction into account. Some remarks on this question will be made at the end of the chapter (see Sec. 30.17). Here, we shall neglect the viscosity of the fluid and shall take only the pressure forces into account. A fluid in which the forces of internal friction are not taken into account is said to be *ideal*, or *nonviscous*.

2. According to the fundamental law of dynamics [equation (7.4)] the change in momentum in unit time equals the acting force. Thus

$$F_1 - F_2 = \frac{\Delta m v_2 - \Delta m v_1}{\Delta t} \quad (30.5)$$

where $F_1 = p_1 A_1$ and $F_2 = p_2 A_2$ are the pressure forces at cross sections A_1 and A_2 . But, according to equations (30.1) and (30.2), $\frac{\Delta m}{\Delta t} = \rho A v$. Substituting into equation (30.5) we obtain

$$p_1 A_1 - p_2 A_2 = \rho_2 A_2 v_2^2 - \rho_1 A_1 v_1^2$$

or

$$A_1 (p_1 + \rho_1 v_1^2) = A_2 (p_2 + \rho_2 v_2^2) \quad (30.6)$$

This is called the *momentum equation* for the flow of an ideal compressible fluid.

30.4. BERNOULLI'S EQUATION

1. Let us apply the law of conservation of energy—the first law of thermodynamics [equation (21.9)]—to gas flow. Dealing with the gas as an ideal compressible fluid, we shall not take the forces of internal friction into account. We shall assume that the processes occurring in the gas are adiabatic.

At cross section A_1 (see Fig. 30.1) a portion of gas of mass Δm has the kinetic energy $K_1 = \Delta m v_1^2 / 2$ and the internal energy $U_1 = \Delta m c_v T_1$. At cross section A_2 the respective values will be $K_2 =$

$= \Delta mv_2^2/2$ and $U_2 = \Delta mc_v T_2$. Then the change in the total energy of the gas as it flows from section A_1 to A_2 can be expressed by

$$\Delta E = (K_2 + U_2) - (K_1 + U_1) = \frac{\Delta mv_2^2}{2} + \Delta mc_v T_2 - \frac{\Delta mv_1^2}{2} - \Delta mc_v T_1$$

The work done by the gas is

$$W = p_2 \Delta V_2 - p_1 \Delta V_1$$

where ΔV_1 and ΔV_2 are the volumes of the given mass of gas at cross sections A_1 and A_2 . Substituting into equation (21.9) and recalling that there is no heat exchange (i.e. $Q = 0$) in an adiabatic process, we can write

$$\frac{\Delta mv_2^2}{2} - \frac{\Delta mv_1^2}{2} + \Delta mc_v T_2 - \Delta mc_v T_1 + p_2 \Delta V_2 - p_1 \Delta V_1 = 0$$

We divide both sides of this equation by Δm , recalling that $\frac{\Delta m}{\Delta V} = \rho$ is the density of the gas. Transferring all the terms with the subindex 1 to one side of the equation and those with subindex 2 to the other, we obtain

$$\frac{v_1^2}{2} + c_v T_1 + \frac{p_1}{\rho_1} = \frac{v_2^2}{2} + c_v T_2 + \frac{p_2}{\rho_2} \quad (30.7)$$

Formula (30.7) is the *energy equation* for the flow of an ideal gas. It is otherwise called *Bernoulli's equation* and was proposed by D. Bernoulli (1700-1782) in 1738.

2. Substituting into equation (30.7) the value for the pressure $p = R\rho T/M$ [see equation (26.18)] and recalling that $c_p = c_v + (R/M)$, we then have

$$\frac{v_1^2}{2} + c_p T_1 = \frac{v_2^2}{2} + c_p T_2 \quad (30.8)$$

Making use of equation (26.18), we can write

$$\frac{v_1^2}{2} + \frac{Mc_p}{R} \frac{p_1}{\rho_1} = \frac{v_2^2}{2} + \frac{Mc_p}{R} \frac{p_2}{\rho_2}$$

But $\frac{Mc_p}{R} = \frac{c_p}{c_p - c_v} = \frac{\gamma}{\gamma - 1}$ because $\frac{c_p}{c_v} = \gamma$. Then

$$\frac{v_1^2}{2} + \frac{\gamma}{\gamma - 1} \frac{p_1}{\rho_1} = \frac{v_2^2}{2} + \frac{\gamma}{\gamma - 1} \frac{p_2}{\rho_2} \quad (30.9)$$

Further on we shall need this expression of Bernoulli's equation.

3. The density and temperature of an ideal incompressible fluid are not changed in its flow. Putting $\rho_1 = \rho_2 = \rho$ and $T_1 = T_2$ in equation (30.7), we have

$$\frac{\rho v_1^2}{2} + p_1 = \frac{\rho v_2^2}{2} + p_2 \quad (30.10)$$

This is Bernoulli's equation for an ideal (nonviscous) incompressible fluid.

30.5. RATE OF PROPAGATION OF DISTURBANCE WAVES IN ELASTIC MEDIA

1. Let us imagine that in some layer of stationary gas, the pressure is suddenly raised by rapid compression or heating. Owing to its surplus pressure this layer begins to expand, transmitting the pressure pulse to adjacent layers. The compression of these layers will be transmitted farther and farther, and a *wave of elastic disturbances* will be originated in the gas.

A similar phenomenon can be observed if we drop a stone into a quiet pond. Here again the disturbances are transmitted from



Fig. 30.2

layer to layer, and a circular wave spreads over the surface of the water (Fig. 30.2).

The *wavefront* is the surface at all points of which the thermodynamic parameters (velocity, density, pressure and temperature) have the same values. The *wavefront velocity* (*wave-propagation velocity*) w is rate of transmission of an elastic disturbance from layer to layer.

2. To calculate the wave velocity it proves convenient to use a reference frame that moves together with the wavefront.

In a reference frame xyz fixed in an undisturbed gas, the process of wave propagation is a nonsteady one because the gas parameters change with time. However, in a frame $x'y'z'$, moving together with the wavefront, the process is a steady one to which the equations of continuity, momentum and energy, discussed in the preceding sections, can be applied (Fig. 30.3).

In a moving frame of reference, the undisturbed gas runs against the wave front with the velocity $u = -w$; the compressed gas lags behind the wavefront and moves to the left with the velocity $u_1 = -v_1 - w$ (Fig. 30.4). Applying the continuity equation (30.4) and the momentum equation (30.6) to the gas flow we obtain

$$\rho u = \rho_1 u_1 \text{ and } p + \rho u^2 = p_1 + \rho_1 u_1^2 \quad (30.11)$$

Solving the first equation for u_1 we obtain $u_1 = \rho u / \rho_1$ and when we substitute into the second equation we obtain the expression for the wavefront velocity:

$$w = |u| = \sqrt{\frac{\rho_1}{\rho} \frac{p_1 - p}{\rho_1 - \rho}} \quad (30.12)$$

3. Two types of waves are observed in gases: shock and sound waves.

A typical feature of a *shock wave* is the fact that the thermodynamic parameters at the wavefront change with a sudden jump and that

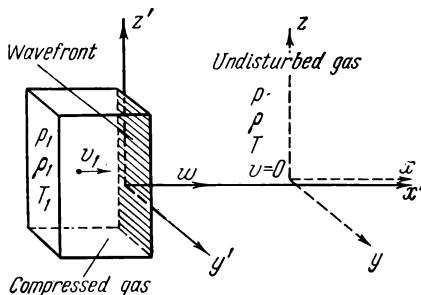


Fig. 30.3

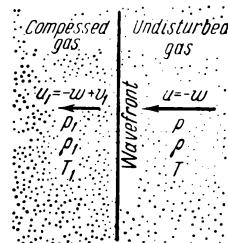


Fig. 30.4

the pressure difference $p_1 - p$ and density difference $\rho_1 - \rho$ may reach very high values. For this reason, a shock wavefront is said to be a *pressure, or compression, shock*.

Sound waves are *waves of weak disturbances*. The difference in pressures and densities are very small: $\Delta p \ll p$ and $\Delta \rho \ll \rho$. It can therefore be assumed in equation (30.12) that $\rho_1 = \rho + \Delta \rho \cong \rho$, obtaining the following expression for the velocity of a sound wave:

$$a = \sqrt{\frac{\Delta p}{\Delta \rho}} \quad (30.13)$$

4. Sound waves propagate at sufficiently high velocities and as the wavefront passes through a gas the compression of the gas can be assumed adiabatic. This enables a more convenient equation to be derived for the velocity of sound waves.

We make use of the expression for the work done in the quasi-static adiabatic expansion of a gas by the small volume ΔV [see equation (27.25)]. Putting $p_1 = p - \Delta p$, $V_1 = V - \Delta V$, $p_2 = p + \Delta p$ and $V_2 = V + \Delta V$, we have

$$\begin{aligned} W &= \frac{p_1 V_1 - p_2 V_2}{\gamma - 1} = \frac{(p - \Delta p)(V - \Delta V) - (p + \Delta p)(V + \Delta V)}{\gamma - 1} \\ &= \frac{-2(V\Delta p + p\Delta V)}{\gamma - 1} \end{aligned}$$

On the other hand, $W = p_{av}(V_2 - V_1) = 2p\Delta V$. Equating these two expressions we obtain

$$2p\Delta V = \frac{-2(V\Delta p + p\Delta V)}{\gamma - 1}$$

or

$$\gamma \frac{\Delta V}{V} = -\frac{\Delta p}{p} \quad (30.14)$$

The minus sign is due to the increase in volume $\Delta V > 0$ in expansion of the gas and the consequent decrease in pressure so that $\Delta p < 0$. The mass of the gas is $m = \rho V$. In changes of volume, the density also changes, but the mass of the gas is constant. Hence

$$(\rho + \Delta\rho)(V + \Delta V) = (\rho - \Delta\rho)(V - \Delta V)$$

Removing the parentheses and adding together like terms, we have

$$2\Delta\rho V = -2\Delta V\rho$$

from which it follows that the relative change in volume is equal (in magnitude) to the relative change in density. Thus

$$-\frac{\Delta\rho}{\rho} = \frac{\Delta V}{V} \quad (30.15)$$

Substituting into equation (30.14) we have

$$\gamma \frac{\Delta\rho}{\rho} = \frac{\Delta p}{p} \quad (30.16)$$

Combining equations (30.13) and (30.16) we find that the velocity of sound is

$$a = \sqrt{\frac{\gamma p}{\rho}} \quad (30.17)$$

or, since the state equation for the gas [see equation (26.18)] is $p = R\rho T/M$, we can write

$$a = \sqrt{\frac{\gamma RT}{M}} \quad (30.18)$$

Thus the velocity of sound in an ideal gas depends entirely on its temperature.

Given that for air $\gamma = 7/5$ (as a diatomic gas) and $M = 29$ kg/mol, it will be left as a problem for the student to show that the velocity of sound is

$$a = 20 \sqrt{T} \text{ (in m/s)}$$

5. After substituting expression (30.17) into Bernoulli's equation (30.9) we have

$$\frac{v_1^2}{2} + \frac{a_1^2}{\gamma - 1} = \frac{v_2^2}{2} + \frac{a_2^2}{\gamma - 1}$$

where a_1 and a_2 are local velocities of sound, i.e. velocities of sound at various points in the stream.

30.6. COMPRESSIBLE FLUID DYNAMICS. MACH NUMBER

1. It was stated above that the compressibility of a gas can be neglected in certain cases, assuming its density to be constant. We shall now introduce a criterion enabling us to specify the conditions for which such a simplification is permissible.

The pressure in a gas evidently will increase by the greatest amount if its motion is slowed down to a stop. Let its velocity be v , pressure p and density ρ before it is slowed down. When it has been stopped its velocity is $v_1 = 0$, pressure $p_1 = p + \Delta p$ and density $\rho_1 = \rho + \Delta\rho$. According to Bernoulli's equation (30.9),

$$\frac{v^2}{2} + \frac{\gamma}{\gamma - 1} \frac{p}{\rho} = \frac{\gamma}{\gamma - 1} \left(\frac{p + \Delta p}{\rho + \Delta \rho} \right)$$

or

$$\frac{v^2}{2} = \frac{\gamma}{\gamma - 1} \left(\frac{p + \Delta p}{\rho + \Delta \rho} - \frac{p}{\rho} \right) = \frac{\gamma}{\gamma - 1} \frac{\Delta \rho}{\rho + \Delta \rho} \left(\frac{\Delta p}{\Delta \rho} - \frac{p}{\rho} \right)$$

But, according to equations (30.13) and (30.17), $\frac{\Delta p}{\Delta \rho} = a^2$ and $\frac{p}{\rho} = \frac{a^2}{\gamma}$. Substituting into the preceding equation and recalling that $\Delta \rho \ll \rho$, we have

$$\frac{v^2}{2} = \frac{\gamma}{\gamma - 1} \frac{\Delta \rho}{\rho} \left(a^2 - \frac{a^2}{\gamma} \right) = a^2 \frac{\Delta \rho}{\rho}$$

2. Thus, the relative change in the density of the gas when its flow is decelerated to a stop is

$$\frac{\Delta \rho}{\rho} = \frac{v^2}{2a^2} \tag{30.19}$$

The ratio of the velocity v of the fluid to the velocity a of sound is called the *Mach number*:

$$M = \frac{v}{a} \tag{30.20}$$

The Mach number serves as a criterion of the compressibility of a fluid. It is evident from equation (30.19) that when M is small, then $\Delta \rho \ll \rho$, and the compressibility of the fluid can be neglected. But if M is large, the density of the fluid changes substantially, and its compressibility cannot be neglected.

30.7. THE MACH CONE

1. Any body travelling in a gas is a source of sound waves. The body, running against the stationary gas, initiates a region of higher pressure in front. This induces a wave of elastic disturbances in the gas. The shape of the wave depends essentially on the velocity of the body.

For the sake of simplicity, let us consider a point source, i.e. a body whose dimensions can be neglected for the purposes of the given problem.

2. Let the velocity of the body be less than that of sound. Then $M = v/a < 1$. The sound wavefront is a sphere whose centre lies

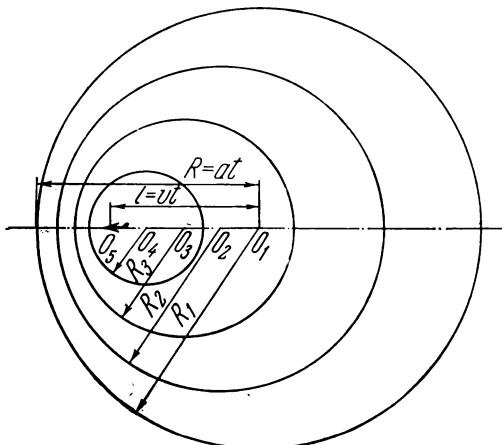


Fig. 30.5

at the point where the source was when it emitted the wave. Let the source be at point O_1 at the initial instant, at point O_2 a second later, at point O_3 two seconds later, etc. (Fig. 30.5). After four seconds the body will be at point O_5 . The wavefront emitted from point O_1 is a sphere of radius $R_1 = 4a$, that from point O_2 a sphere of radius $R_2 = 3a$, that from point O_3 a sphere of radius $R_3 = 2a$, etc. It is evident in Fig. 30.5 that the body travels inside a spherical wave which is always ahead of it. The path of the body in t seconds is $l = vt$. It is less than the radius of a spherical sound wave $R = at$.

The result will be entirely different if the velocity v of the point source is higher than that of sound ($M > 1$). Here the body will overtake the sound wave. The path travelled by the body during the time t , i.e. $l = vt$, is greater than the wave radius $R = at$ (Fig. 30.6). Indeed, suppose the body emits a sound wave at the initial instant

from point O_1 . After one second, overtaking the wavefront, the body will reach point O_2 and emit another wave, etc. At the end of the fourth second the body will be at point O_5 , having overtaken all the waves it emitted during its travel. Then the set of waves emitted by the body constitute a family of spheres whose envelope is a cone with

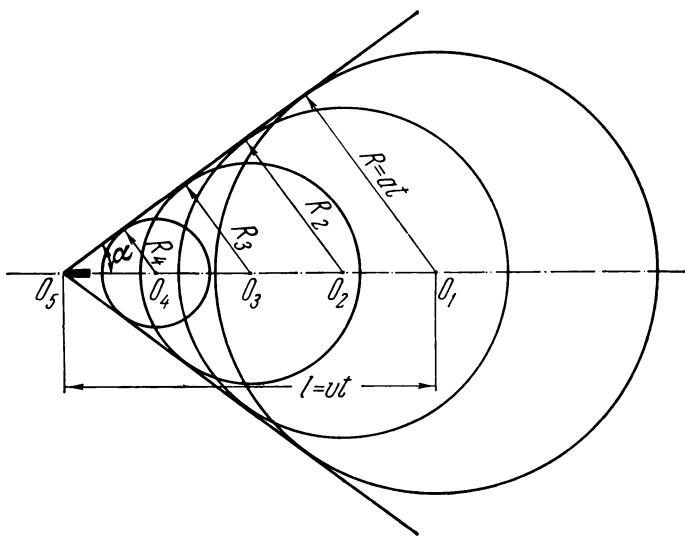


Fig. 30.6

the body at its apex. This is called a cone of weak disturbances or a *Mach cone*, in honour of Ernst Mach (1838-1916) who first described it. The half-angle at the apex of this cone is found, as is evident from Fig. 30.6, by the relation

$$\sin \alpha = \frac{R}{l} = \frac{a}{v} = \frac{1}{M} \quad (30.21)$$

30.8. THE BOW SHOCK WAVE

1. The point source discussed in the preceding section is a highly simplified idealization of phenomena occurring upon the motion of a body in a gas at a velocity exceeding that of sound, i.e. at supersonic velocities. Actually, a *bow shock wave*, or *compression shock* is originated in front of a body of finite size. This is a very thin layer, with a thickness of the order of the mean free path of the molecules, in which the thermodynamic parameters of the gas change discontinuously (in a jump).

2. The density of the gas varies sharply within the compression shock. This leads to a considerable change in the refractive index of light rays passing through the gas, enabling a bow shock wave

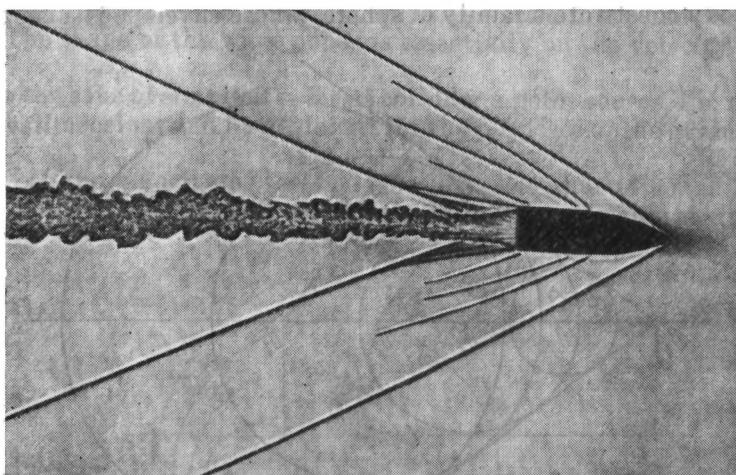


Fig. 30.7

to be photographed with special apparatus. Shown in Fig. 30.7 is a photograph of an oblique shock wave induced in front of a wedge-shaped or conical body (for a Mach number $M \approx 2.5$).

3. The velocity of a bow shock wave is known; it coincides with the velocity of the body. Equations derived in Secs. 30.2, 30.3 and 30.4 can be applied to determine the differences in pressure and density of the gas before and after the jump.

We shall confine ourselves to an analysis of the phenomena associated with a normal pressure shock.

30.9. PHENOMENA OF A NORMAL COMPRESSION SHOCK

1. In a moving frame of reference, fixed to the compression shock

$$\left. \begin{aligned} \rho u &= \rho_0 u_0 \\ p + \rho u^2 &= p_0 + \rho_0 u_0^2 \\ \frac{u^2}{2} + \frac{\gamma}{\gamma-1} \frac{p}{\rho} &= \frac{u_0^2}{2} + \frac{\gamma}{\gamma-1} \frac{p_0}{\rho_0} \end{aligned} \right\} \quad (30.22)$$

where $u_0 = -w$ = approach velocity of the undisturbed gas toward the compression shock

p_0 and ρ_0 = pressure and density of the undisturbed gas
 $u = v - w$ = velocity of the gas behind the compression shock

p and ρ = its pressure and density in this region
 w = velocity of the body (and bow shock wave) with respect to the earth.

Equations (30.22) are the equation of continuity and the momentum and energy equations.

2. The process of compressing the gas at the compression shock is *adiabatic* but it is not *quasi-static* because the gas passes so rapidly through the shock that there is insufficient time to equalize the parameters in neighbouring regions. Consequently, the relationship between the pressure and density of the gas cannot be expressed here by Poisson's equation which describes only quasi-static adiabatic processes.

Eliminating from equations (30.22) the velocities

$$u_0 = \sqrt{\frac{\rho}{\rho_0} \left(\frac{p - p_0}{\rho - \rho_0} \right)}$$

and

$$u = \sqrt{\frac{\rho_0}{\rho} \left(\frac{p - p_0}{\rho - \rho_0} \right)}$$

and performing a number of simple but lengthy calculations, we have

$$\frac{p}{p_0} = \frac{\left(\frac{\gamma+1}{\gamma-1}\right) \frac{\rho}{\rho_0} - 1}{\frac{\gamma+1}{\gamma-1} - \frac{\rho}{\rho_0}} \quad (30.23)$$

This last is the *equation of a shock adiabat*, or the *Hugoniot equation*. A plot of this equation is shown in Fig. 30.8. Poisson's quasi-static adiabat and an isotherm are plotted on the same graph for comparison.

3. Shock compression of the gas is accompanied by a sharp rise in its temperature. As a result, the pressure increases with the density in shock compression much faster than in quasi-static adiabatic compression of the gas.

Thus, according to Poisson's equation, the density of a gas can increase without limit in quasi-static adiabatic compression. In shock compression, however, the increase in density is limited. In-

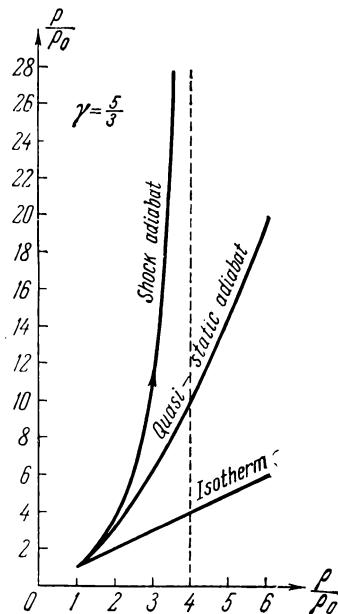


Fig. 30.8

deed, at $\frac{\rho}{\rho_0} = \frac{\gamma+1}{\gamma-1}$ the denominator of equation (30.23) becomes zero. This is meaningless. Hence

$$\frac{\rho}{\rho_0} < \frac{\gamma+1}{\gamma-1} \quad (30.24)$$

The density of a monatomic gas ($\gamma = 5/3$) cannot be increased more than fourfold in shock compression. The density of a diatomic gas ($\gamma = 7/5$) will not increase more than sixfold if we assume that no vibrations of the atoms in the molecule are excited. But if we take

into account the vibrations that are also excited at high temperatures in diatomic molecules (see Sec. 27.9), specific heat ratio is found to be $\gamma = 9/7$, and the density of the diatomic gas can increase eightfold, but not more.

4. Since the process of shock compression is not quasi-static, it is *irreversible*. Consequently, it is accompanied by an increase in entropy. It follows that *shock rarefaction*, i.e. *shock expansion*, of a gas is impossible. Such a process would have to be accompanied by a decrease in entropy, thereby disagreeing with the second law of thermodynamics.

Thus, if a gas undergoes shock compression, it will expand again

by a Poisson adiabat (Fig. 30.9) and not by a shock adiabat. At the end of the expansion process, the gas will not return to its initial state. At the same density, its pressure and temperature will be much higher than in the initial state.

5. We should note that the work done on a gas in its shock adiabatic compression considerably exceeds the work done in its quasi-static adiabatic compression. The reason is that in shock compression the temperature of the gas and, consequently, its internal energy, increase much more rapidly than in a quasi-static process.

30.10. WAVE DRAG

1. Since a compression shock is initiated in front of a body travelling in a gas at supersonic velocities, the body encounters much higher resistance than if it were travelling at a subsonic velocity. One of the main causes for resistance, or drag, is the difference in pressures at the front and rear edges of the body the gas is flowing around (see

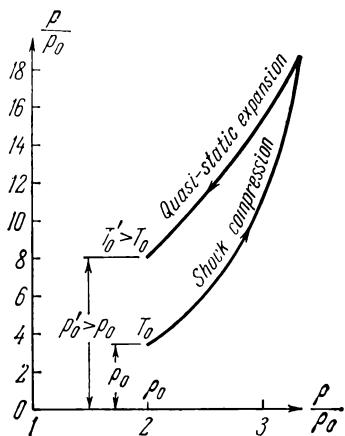


Fig. 30.9

Sec. 11.8). But the pressure at the front edge is sharply increased when a compression shock is initiated. This likewise increases the pressure drag. Since this drag is caused by a shock wave it is called *wave drag, or resistance*.

2. Wave drag can be reduced by imparting a pointed shape to the front, or leading, edge. It may be either conical (*a*) or ogival (*b*) as shown in Fig. 30.10. Supersonic jet-propelled aircraft are of arrow-shaped design with pronounced sweepback of the wings (*a*) (Fig. 30.11). Such a design is called for because a normal compression shock is initiated in front of a blunt leading edge and an oblique one in front of a pointed edge (Fig. 30.7). The velocity changes much more intensively at a normal shock, and it follows from equation (30.6) that the less the rate of flow changes, the less the difference in pressures. This is why the wave drag is considerably less for an oblique pressure shock than for a normal one.

3. The initiation of a compression shock is accompanied by an increase in entropy, and the kinetic energy of the moving body is

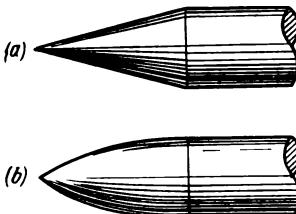


Fig. 30.10.

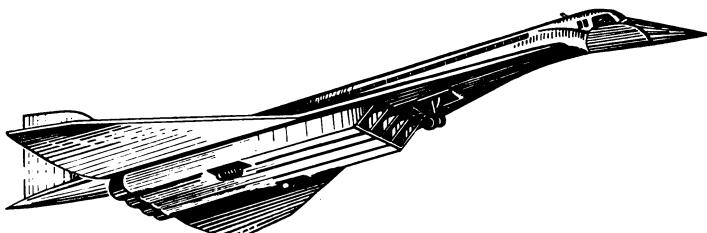


Fig. 30.11

irreversibly converted into internal energy of the gas. This phenomenon is used in braking a spaceship in the atmosphere.

This particular principle is used in a soft landing of spacecraft returning, for example, to the earth after orbiting the moon. The high precision with which spacecraft are controlled enables them to enter the earth's atmosphere at a quite definite angle. This provides for the required degree of deceleration without causing impermissible *g*-loads.

When a spaceship is braked, the surrounding atmosphere is heated to a very high temperature, up to tens of thousands of degrees. For this reason, the spaceship must be protected by a heat-insulating shield of some heat-resistant material with low heat conductivity. Special plastics are employed for this purpose.

30.11. NOZZLES

1. A pipe in which a fluid expands so that its internal energy is converted into kinetic energy is called a *nozzle*. We wish to find the conditions under which the fluid flowing out of the nozzle will attain maximum velocity. This problem is very simple for an incompressible fluid: equation (30.4) indicates that the maximum velocity is obtained at the narrowest part of the pipe. For a gas this depends upon whether the rate of flow is subsonic ($M < 1$) or supersonic ($M > 1$).

2. Let us consider two nearby cross sections whose areas differ only slightly: $A_1 = A - \Delta A$ and $A_2 = A + \Delta A$ with $\Delta A \ll A$. Accordingly, $\Delta p \ll p$, $\Delta \rho \ll \rho$, and $\Delta v \ll v$. The equation of continuity (30.3) can be written for this case as

$$(A - \Delta A)(\rho - \Delta \rho)(v - \Delta v) = (A + \Delta A)(\rho + \Delta \rho)(v + \Delta v)$$

After removing the parentheses and adding together like terms, we have

$$2A\rho\Delta v + 2\rho v\Delta A + 2Av\Delta \rho + 2\Delta A\Delta \rho\Delta v = 0$$

Dividing the equation by $2A\rho v$ we obtain the term $\frac{\Delta A}{A} \frac{\Delta v}{v} \frac{\Delta \rho}{\rho}$,

which is very small (thus, if $\frac{\Delta A}{A} \approx \frac{\Delta v}{v} \approx \frac{\Delta \rho}{\rho} = 10^{-3}$, then their product will be of the order of 10^{-9}) and can therefore be neglected. Finally

$$\frac{\Delta A}{A} + \frac{\Delta \rho}{\rho} + \frac{\Delta v}{v} = 0 \quad (30.25)$$

3. The Bernoulli equation (30.9) for our case will be

$$\frac{(v - \Delta v)^2}{2} + \frac{\gamma}{\gamma - 1} \frac{p - \Delta p}{\rho - \Delta \rho} = \frac{(v + \Delta v)^2}{2} + \frac{\gamma}{\gamma - 1} \frac{p + \Delta p}{\rho + \Delta \rho}$$

Eliminating the parentheses and adding together like terms, we obtain

$$2v\Delta v = \frac{2\gamma}{\gamma - 1} \frac{p\Delta \rho - \rho\Delta p}{\rho^2 - \Delta \rho^2}$$

But $\Delta \rho \ll \rho$, and therefore the term $\Delta \rho^2$ in the denominator can be neglected. Then

$$v\Delta v = \frac{\gamma}{\gamma - 1} \left(\frac{p}{\rho} \frac{\Delta \rho}{\rho} - \frac{\Delta p}{\rho} \right)$$

According to equation (30.13) $\Delta p = a^2 \Delta \rho$, and from equation (30.17) it follows that $p/\rho = a^2/\gamma$. Substituting into the preceding equation we have

$$v\Delta v = \frac{\gamma}{\gamma - 1} \left(\frac{a^2}{\gamma} \frac{\Delta \rho}{\rho} - a^2 \frac{\Delta \rho}{\rho} \right)$$

or, finally,

$$\frac{\Delta p}{\rho} = - \frac{v \Delta v}{a^2} \quad (30.26)$$

4. Combining equations (30.25) and (30.26) we obtain an expression that relates the change in the rate of flow with the change in the cross-sectional area of the nozzle:

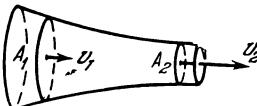
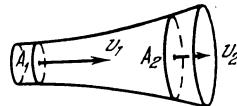
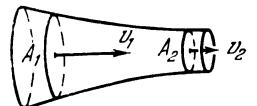
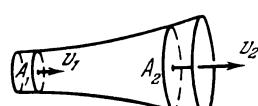
$$\frac{\Delta A}{A} + \frac{\Delta v}{v} \left(1 - \frac{v^2}{a^2} \right) = 0$$

or

$$\frac{\Delta v}{v} (1 - M^2) = - \frac{\Delta A}{A} \quad (30.27)$$

This expression leads to a very important consequence: if $M < 1$, then the sign of Δv is opposite to that of ΔA ; if $M > 1$, then the sign of Δv coincides with that of ΔA . This means that the velocity of a subsonic stream increases when the nozzle narrows (converges) and is reduced when the nozzle widens (diverges). It is just the opposite with a supersonic stream which is accelerated in a diverging nozzle and slowed down in a converging nozzle (Table 30.1).

Table 30.1

	Converging nozzle $A_2 < A_1$ and $\Delta A < 0$	Diverging nozzle $A_2 > A_1$ and $\Delta A > 0$
Subsonic stream $M < 1$	 <p>$v_2 > v_1$ and $\Delta v > 0$ Stream is accelerated</p>	 <p>$v_2 < v_1$ and $\Delta v < 0$ Stream is decelerated</p>
Supersonic stream $M > 1$	 <p>$v_2 < v_1$ and $\Delta v < 0$ Stream is decelerated</p>	 <p>$v_2 > v_1$ and $\Delta v > 0$ Stream is accelerated</p>

The maximum velocity at the narrowest part (throat) of the nozzle does not exceed the velocity of sound.

30.12. ANALOGY BETWEEN A NOZZLE AND A HEAT ENGINE

1. Suppose that a nozzle is connected to some reservoir (for instance, a steam boiler or the combustion chamber of a gas turbine) in which an ideal gas is maintained at constant pressure p_0 , density ρ_0 and temperature T_0 . The corresponding parameters of the surrounding atmosphere are denoted by p_{at} , ρ_{at} and T_{at} . Gas flowing out of the reservoir and through the nozzle is accelerated. Its velocity as it flows out of the nozzle is v . This velocity can be determined by Bernoulli's equation (30.8), the velocity of the gas in the reservoir being $v_0 = 0$. Then

$$\frac{v^2}{2} + c_p T_{at} = c_p T_0$$

from which we can write the expression for the kinetic energy of unit mass of the gas:

$$\frac{v^2}{2} = c_p (T_0 - T_{at}) \quad (30.28)$$

2. The maximum velocity and the corresponding maximum kinetic energy for unit mass of the gas could be obtained under the condition that $T_{at} = 0$, i.e. if the gas escaped into a vacuum where the temperature was absolute zero. Thus

$$\frac{v_{\max}^2}{2} = c_p T_0 \quad (30.29)$$

In this case the internal energy of the gas would be completely converted into kinetic energy.

3. Now let us determine the ratio of the true value of the kinetic energy of unit mass of a gas to the maximum value:

$$\eta = \frac{v^2}{v_{\max}^2} = \frac{T_0 - T_{at}}{T_0} \quad (30.30)$$

This ratio is equal to the efficiency of the Carnot cycle for a heat engine operating within the same temperature interval. It is the maximum coefficient of utilization to the internal energy of the gas in the reservoir.

30.13. THE LAVAL NOZZLE

1. To obtain a supersonic stream from the end of a nozzle it must be designed with a special shape. As a matter of fact the gas flows from the reservoir at a low velocity (see Eq. 15.8). To accelerate the stream it is necessary, according to equation (30.27), to have a converging part at the beginning of the nozzle. At a sufficiently high difference in pressures, the velocity of the stream at the narrowest

cross section will reach the local velocity of sound. If the nozzle diverges after the throat, then, according to equation (30.27), the stream will continue to accelerate.

A nozzle designed for such operation is called a *Laval*, or *converging-diverging*, nozzle (Fig. 30.12). Such nozzles are used in steam and gas turbines, and in jet-propulsion engines and rockets.

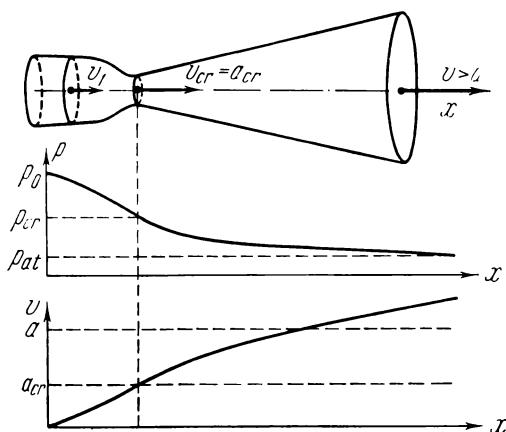


Fig. 30.12

2. As a gas flows through a Laval nozzle, its pressure, density and temperature continuously decrease; its internal energy correspondingly decreases. At the expense of this factor, the kinetic energy of the stream continuously increases, and its velocity can considerably surpass the velocity of sound (see Fig. 30.12).

30.14. THE JET ENGINE

1. The Laval nozzle is the main element of any jet engine. To obtain a large reactive thrust it is necessary for the jet outflow velocity to be as high as possible [see equation (15.8)]. And a high, that is, supersonic gas outflow velocity can be attained only by means of a Laval nozzle.

There are many designs of jet engines. We shall confine ourselves to a discussion of the simplest type—the athodyd, or ramjet engine (rj). A schematic diagram of such an engine is shown in Fig. 30.13.

2. A ramjet engine operates efficiently and generates a sufficiently large reactive thrust if it is travelling at a supersonic velocity. For simplicity let us take a reference frame fixed to the engine. In this frame, a stream of air runs against the entry nozzle at a velocity

higher than that of sound. One or several compression shocks are formed in region 1, behind the entry nozzle, and the velocity of the air drops to a subsonic value. Flowing along the diverging channel, the subsonic stream is further decelerated. Its velocity is reduced, and

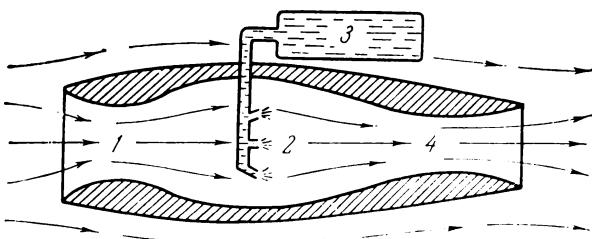


Fig. 30.13

its pressure, density and temperature increase. In region 2, called the combustion chamber, fuel from tank 3 is injected into the compressed air. The energy evolved in fuel combustion increases the internal energy of the gas (mixture of air and the products of combustion). From the combustion chamber, the gas passes into the

Laval nozzle 4. If the thermodynamic parameters of the gas in the combustion chamber have been properly chosen, then at the narrowest part of the nozzle the velocity of the stream will equal the local velocity of sound, and a supersonic stream will begin to flow out of the exit nozzle.

Note that a ramjet engine is very simple in design and highly reliable. However, it can operate only after being accelerated to supersonic velocities. This can be done by means of assisted take-off rockets.

3. The principle of the turbocompressor jet engine is illustrated in Fig. 30.14. It can operate at any velocities, including low ones. Unlike the ramjet engine, here the air is compressed by compressor 1. The energy of the stream behind combustion chamber 2 is partly used to drive gas turbine 3 which runs the compressor and sometimes a propeller as well. The remaining energy is used to produce the reactive thrust in nozzle 4 located behind the turbine.

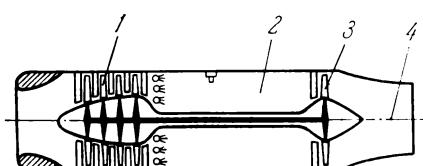


Fig. 30.14

30.15. THE AIRPLANE WING

1. A plate around which a fluid flows, profiled in such a way that the upward force, or *dynamic lift*, on the plate considerably exceeds the drag, is called an *airfoil*, or *wing*. Depending upon the velocity for which the airfoil is designed, it has a cross-sectional shape such as shown in Fig. 30.15a or b. In supersonic aircraft, the leading edge is sharp (Fig. 30.15c) to reduce wave drag.

2. Dynamic lift can be explained qualitatively by Bernoulli's equation. As the gas flows over the airfoil it divides into two streams. The airfoil (or wing) is of asymmetrical shape: on top it is convex, underneath it is concave or flat. Owing to this asymmetry, the rate of flow v_1 above the airfoil is greater than that v_2 below, while the pressure p_1 above is correspondingly less (Fig. 30.16). This difference in pressures produces the dynamic lift.

The dynamic lift can be increased in various ways.

3. In the first place, the airfoil can be made more convex above and below it can be made concave (see Fig. 30.15a). Here the rate of flow is decreased below the airfoil and increased above. The greater the difference in velocities, the greater the pressure difference and,

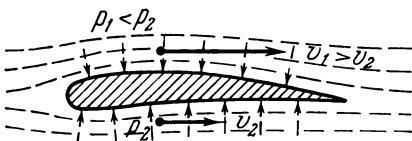


Fig. 30.16

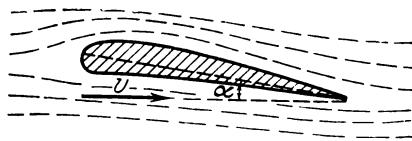


Fig. 30.17

with it, the dynamic lift. But a wing of such a shape is subject to high drag, or resistance, as well. For this reason, airfoils of this type are commonly used on low-speed aircraft.

4. A certain angle can be provided between the chord of the airfoil and the windstream direction. This is called the *angle of incidence*, or *attack*, α (Fig. 30.17). This will also slow down the stream below the airfoil and thereby increase the dynamic lift. But the drag also increases with the angle of incidence. The airstream will break away from the wing (called flow separation) and reduce the lift. This is why modern airplanes do not have an angle of incidence exceeding 3° or 4° .

5. Finally, the dynamic lift can be increased by increasing the velocity of the approaching stream. Indeed, the difference in velo-

cities above the upper and below the lower surfaces of the airfoil is small at small angles of incidence and a small convexity (called *camber*) of the airfoil (see Fig. 30.15b). Therefore, the half-sum of these velocities differs only slightly from the airstream velocity: $(v_1 + v_2)/2 \cong v$. This allows us to change equation (30.10) to the form

$$\Delta p = \frac{1}{2} \rho (v_2 + v_1)(v_2 - v_1) \cong \rho v \Delta v$$

Consequently, at high airstream velocities (*airspeeds*) the dynamic lift will be large even if the airfoil, or wing, is small and has a low camber. This is the shape of wing used in modern high-speed aircraft, because it enables the drag to be substantially reduced, and this is certainly expedient.

30.16. MEASURING THE PRESSURE AND VELOCITY IN A STREAM OF FLUID

1. In this section we shall deal with subsonic flow ($M < 1$) in which the compressibility of the fluid can be neglected. We shall learn how to measure the pressure in a moving fluid with a stationary manometer and how to find the rate of flow from the pressure.

2. We place in the stream a small tube of streamlined shape, with one or several holes along the side of the horizontal leg (Fig. 30.18). Such a streamlined tube does not appreciably distort the stream and therefore has practically no effect on its velocity (and, consequently, its pressure). Owing to the elastic forces, the liquid enters the tube and rises in its vertical leg to a certain height h at which the hydrostatic pressure counterbalances the pressure in the stream. The pressure can be found by the formula $p = \rho gh$. If the pressure in the liquid is low, the manometric tube is inclined, and then the pressure is calculated by the formula

$$p = \rho gl \sin \alpha$$

[see equation (26.7)].

To measure the pressure in a gas, the tube is connected by a hose to an ordinary liquid-column gas manometer or a metallic pressure gauge.

3. Now we put another tube, a so-called *Pitot tube*, at the same place in the stream with its entry end facing upstream. The column of liquid in the manometer tube will rise higher than in the preceding case (Fig. 30.19), thereby indicating an increase in pressure. Indeed, the fluid is stopped in the Pitot tube and the pressure of the liquid column in the manometer tube is counterbalanced by the pressure of *stagnant fluid* and not of moving fluid as in the preceding experiment. The pressure of the stagnant fluid is found by Bernoulli's

equation (30.10), if we put $v_2 = v_{stag} = 0$ and $p_2 = p_{stag}$. Thus

$$p_{stag} = p + \frac{\rho v^2}{2} \quad (30.31)$$

Often the pressure in the stagnant fluid p_{stag} is not very aptly called the *full ram pressure*, the pressure p in the flowing fluid is called the *static pressure*, and their difference $p_{stag} - p = \rho v^2/2$ is called the *dynamic pressure*.

4. The flow velocity at some point in a fluid is measured by a *Pitot-Prandtl tube* (Fig. 30.20). Since the difference in the pressure

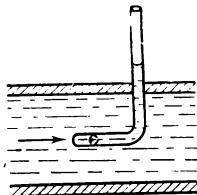


Fig. 30.18

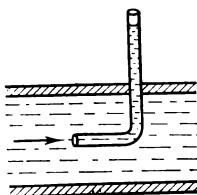


Fig. 30.19

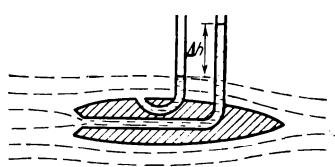


Fig. 30.20

indicated by the two manometric tubes is proportional to the square of the velocity, the manometer can be calibrated directly in velocity units.

To measure the flow velocity of a gas, the Pitot-Prandtl tube is connected to a differential manometer which indicates the difference in pressures and, thereby, according to equation (30.31) the flow velocity.

The scale of the differential manometer is calibrated in velocity units.

It follows from the principle of relativity that the reading of this instrument does not depend on whether the stream of fluid flows over the tube or the tube is moving in a stationary fluid. The application of the Pitot-Prandtl tube to determine the velocity of an airplane with respect to the air or that of a ship with respect to the water is based on this principle.

30.17. VISCOUS FLUID DYNAMICS. FLUID FRICTION IN PIPES

1. Suppose a liquid flows along a pipe of constant cross section. Applying to this case the equation of continuity (30.3), the momentum equation (30.6) and, since $\rho_1 = \rho_2$ and $A_1 = A_2$, we find that the pressure should be the same at all cross sections: $p_1 = p_2$. But this implies a paradoxical phenomenon: liquid can flow along a pipe without any pressure drop, all by itself! And this completely contradicts the results of experiments. This can be readily demonstrated

by means of the simple device shown schematically in Fig. 30.21. When a liquid flows along the pipe, the manometric tubes indicate a difference in pressure, or pressure drop. We find that the pressure at the beginning of the pipe is considerably higher than at its end.

The cause of the disagreement between the theory and an experiment is that in deriving equation (30.6) we neglected the forces of internal friction which, in the given situation, play an essential role.

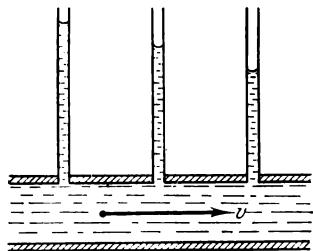


Fig. 30.21

forces of fluid friction at higher Reynolds numbers for either smooth or rough pipe. Empirical formulas, carefully checked experimentally, are used for this purpose.

3. The pressure drop, i.e. the difference in the pressures at two cross sections with the distance l between them, can be calculated by the formula

$$\Delta p = \lambda \frac{l}{d} \frac{\rho v^2}{2} \quad (30.32)$$

where λ is the coefficient of fluid friction.

The following relationship is valid at $Re \leq 2320$ for any kind of pipe, smooth or rough,

$$\lambda = \frac{64}{Re} \quad (30.33)$$

Valid for smooth pipes (for instance, glass tubes) at $Re > 2320$ is the empirical formula

$$\lambda = \frac{0.316}{\sqrt[4]{Re}} \quad (30.34)$$

It should be noted that the transition from laminar flow, for which the coefficient of fluid friction λ is taken according to equation (30.33), to turbulent flow where equation (30.34) is to be used, depends not only on the Reynolds number, but on many other factors. A discussion of this question is beyond the scope of this book.

PART FOUR ▲ MOLECULAR FORCES AND STATES OF AGGREGATION OF MATTER

Chapter 31 MOLECULAR FORCES

31.1. THE DENSITY AND COMPRESSIBILITY OF SUBSTANCES

1. We found in Sec. 26.3 that a comparison of the densities of a substance in the solid, liquid and gaseous states gives us an idea of the distance between molecules in the three states. The distances between molecules in the solid and liquid states were found to differ only slightly, no more than several per cent, while in gases the distances (at not very high pressures) were about ten times as great. Hence, the molecules in liquids and gases are said to be closely packed. Owing to close packing, liquids and solids are much less compressible than gases.

2. Suppose a body is subject to uniform compression (from all sides) or expansion (to all sides). This will change its volume and, consequently, its density. We denote by ϵ the *relative volume change*. Then

$$\epsilon = -\frac{\Delta V}{V} = \frac{\Delta \rho}{\rho} \quad (31.1)$$

in which the minus sign indicates that at an increase in the volume of a body its density, at constant mass, decreases.

Surplus pressure σ with respect to atmospheric pressure is sometimes referred to as *stress*:

$$\sigma = \Delta p \quad (31.2)$$

The *compressibility* β of a substance is the quotient of the relative change in volume divided by the stress. Thus

$$\beta = \frac{\epsilon}{\sigma} = -\frac{\Delta V}{V \Delta p} = \frac{\Delta \rho}{\rho \Delta p} \quad (31.3)$$

The reciprocal is called the *bulk compressibility modulus* or simply *bulk modulus*:

$$K = \frac{1}{\beta} \quad (31.4)$$

3. Combining equations (30.16) and (31.3), we find that the compressibility of a gas in an adiabatic process is

$$\beta_{ad} = \frac{1}{\gamma p} \quad (31.5)$$

It will be left as a problem for the student to show, using the equation of state for a gas, that in an isothermal process

$$\beta_{isoth} = \frac{1}{p} \quad (31.6)$$

At pressures in the gas of the order of several dozens of atmospheres, $p \approx 20 \text{ atm} \approx 2 \times 10^6 \text{ Pa}$, the isometric compressibility of gases is $\beta \approx 5 \times 10^{-7} \text{ Pa}^{-1}$.

The compressibility of liquids and solids is only about one thousandth to one hundred thousandth of that of gases. This is evident from Table 31.1, which lists the density and compressibility of certain liquids and solids at the same value of the surplus pressure ($p \approx 20 \text{ atm}$).

Table 31.1

Substance	Density ρ , kg/m ³	Compressibility β , Pa ⁻¹	Substance	Density ρ , kg/m ³	Compressibility β , Pa ⁻¹
Water (0 °C)	999.9	48.9×10^{-11}	Copper	8930	0.73×10^{-11}
Ice (0 °C)	916.8	40×10^{-11}	Aluminium	2690	1.37×10^{-11}
Ethyl alcohol (15 °C)	790	76×10^{-11}	Diamond	3500	0.23×10^{-11}
Mercury (20 °C)	13545.9	3.8×10^{-11}	Graphite	2300	3.0×10^{-11}

31.2. MOLECULAR FORCES

1. An analysis of data on the compressibility of liquids and solids and also of many phenomena discussed in elementary physics (adhesion of well ground-in surfaces of solids to each other, wetting of solids by liquids, etc.) indicates that there are considerable forces between molecules. They are called *molecular forces*. These forces possess many interesting features.

As the molecules move away from each other, molecular interaction is manifested in the form of forces of *attraction*; when brought together to distances of the order of the linear dimensions of the molecules themselves, interaction takes the form of forces of *repulsion*. Consequently, there is a definite distance d between molecules where the forces of attraction and repulsion balance each other and their resultant equals zero. A system of molecules located at distances d from one another is in a state of stable equilibrium and their interaction energy is minimal (see Secs. 19.5 and 19.6).

Molecular forces are displayed only when the distance between molecules is very small, i.e. of the order of the effective size of the molecules themselves (see Sec. 25.3). When the distances are increased by several times, the forces become practically equal to zero.

This explains why the action of molecular forces can be neglected in most cases in analyzing phenomena in gases where distances between molecules are about tenfold those in liquids and solids.

2. The magnitude of molecular forces does not depend upon the total number of molecules. Thus, the density or elastic properties of liquids and solids do not depend upon the size of the test piece, a drop of water and water in the ocean have exactly the same density and compressibility at the same temperature and external pressure, etc. In this, molecular forces differ in principle from gravitational or Coulomb forces in which the resultant is determined from the action of all the bodies making up the system. The reason why molecular forces act only between adjacent molecules will be cleared up in the following sections.

31.3. THE ELECTRICAL ORIGIN OF MOLECULAR FORCES

1. Careful research conducted by many scientists in the first half of the 20th century has led to the conclusion that *molecular forces are of electrical origin*.

On the face of it such a conclusion may seem strange because we know that molecules are neutral and that electrical interaction is typical only of charged bodies. However, it has been shown that electrical interaction is developed between dipoles: neutral particles with asymmetrical arrangement of the charges, which are equal in magnitude but of opposite sign (see Sec. 10.4). This enables us to understand the nature of the electrical interaction between molecules.

2. The molecules of many substances are of symmetrical structure as shown schematically in Fig. 31.1. Here the positively charged ball represents the nuclei of the atoms making up the molecule, and the negatively charged sphere is the electron cloud of these atoms.

Suppose the molecule is deformed so that its electron cloud is displaced somewhat to one side as shown in Fig. 31.2. Such a deformed molecule can be regarded, to a first approximation, as a dipole with a certain moment $\mathbf{p}_e = ql$. In this case the molecule sets up, outside of itself, an electric field with the strength

$$E \cong \frac{p_e}{\epsilon_0 r^3}$$

[see equation (10.10)].

3. The field of this molecule deforms the nearby molecules. The positive nucleus of the molecule is subject to a repulsive force directed along the field strength vector; the negatively charged electron cloud is subject to forces directed oppositely (Fig. 31.3). These two kinds of forces deform the molecule.

Hence, *molecules having a dipole moment p_e induce a dipole moment p'_e in neighbouring molecules.*

It can be readily shown that the induced dipole moment is proportional to the field strength. Indeed, the dipole moment is $p'_e = el'$, where l' is the arm of the induced dipole. The strength E of the

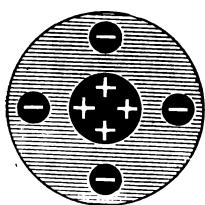


Fig. 31.1

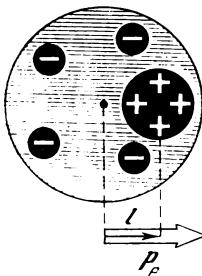


Fig. 31.2

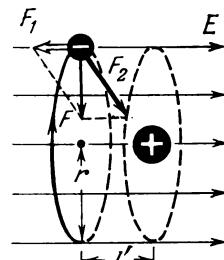


Fig. 31.3

external electric field does not exceed 10^4 to 10^6 V/m and usually is considerably weaker. The strength of the field set up on the orbit by a positively charged nucleus is $E_{nuc} = \frac{e}{4\pi\epsilon_0 r^2} \cong 10^{11}$ V/m (the radius of the atom $r \cong 10^{-10}$ m). Therefore, force F_1 , exerted on the electron by the external electric field, is substantially less than force F_2 , with which the nucleus acts on the electron. For this reason the displacement of the electron cloud is found to be considerably less than the orbit radius.

It is evident from Fig. 31.3 that

$$\frac{l'}{r} = \frac{F_1}{F} \cong \frac{F_1}{F_2}$$

Substituting the values $F_1 = eE$ and $F_2 = eE_{nuc} = \frac{e^2}{4\pi\epsilon_0 r^2}$, we obtain

$$l' \cong \frac{4\pi r^3}{e} \epsilon_0 E$$

Then the induced dipole moment is

$$p'_e = el' = 4\pi r^3 \epsilon_0 E = \alpha \epsilon_0 E \quad (31.7)$$

The quantity $\alpha \cong 4\pi r^3$ is called the *polarizability of the molecule*.

In our case

$$p'_e \cong \frac{\epsilon_0 \alpha p_e}{\epsilon_0 r^3} = \frac{\alpha p_e}{r^3} \quad (31.8)$$

4. If two molecules have dipole moments p_e and p'_e of the same direction and are not far from each other, they attract each other [see equation (10.7)]. Thus

$$f = -\frac{6p_e p'_e}{4\pi\epsilon_0 r^4}$$

If we substitute into this equation the expression for the moment of an induced dipole equation (31.8), we obtain finally

$$f \cong -\frac{\alpha p_e^2}{\epsilon_0 r^7} \quad (31.9)$$

This result is very important: if a molecule induces a dipole moment in its nearest neighbours, forces of attraction between them are inversely proportional to the seventh power of the distance between the molecules.

5. Let us try to estimate the order of magnitude of molecular interaction. The dipole moment has a value of about $p_e \cong 10^{-30}$ C-m, the polarizability is $\alpha \cong 10^{-28}$ m³ and the distance between the molecules is $r \cong 3 \times 10^{-10}$ m. Substituting into equation (31.9), we have

$$|f| \cong \frac{10^{-28} \times 10^{-30} \times 36\pi \times 10^9}{2.2 \times 10^3 \times 10^{-70}} \cong 5 \times 10^{-11} \text{ N}$$

This would seem to be a very small force. But if you take into account the fact that all bodies consist of an immense number of molecules, it becomes clear that their interaction yields an appreciable effect.

We can assess, for instance, the force that must be applied to rupture a bar with a cross-sectional area of 1 cm². It is equal to the force of interaction between two adjacent molecules multiplied by the number of molecules that fit into an area of 1 cm² when sufficiently closely packed. Since the distance between the molecules is of the order of 3×10^{-10} m = 3×10^{-8} cm, the sought for number of molecules is

$$N = \left(\frac{1}{3 \times 10^{-8}} \right)^2 \frac{1}{\text{cm}^2} \cong 10^{15} \frac{1}{\text{cm}^2}$$

Then the rupture stress is

$$\sigma = |f| N \cong 5 \times 10^{-11} \times 10^{15} = 5 \times 10^4 \text{ Pa}$$

It is known from experiments that the tensile strength of steel is about 10^5 Pa. Our tentative and quite crude calculation yielded a result which is correct in order of magnitude.

31.4. THE MOLECULAR FORCE GRAPH

1. In the preceding section we dealt with the simplest case of molecular attraction, the interaction between induced dipoles. Actually, forces of attraction between molecules may also be due to various

other causes, which are beyond the scope of this book. Any rigorous theory of molecular forces must be based only on quantum mechanics. It was found, however, that in all cases the force of molecular attraction was inversely proportional to the seventh power of the distance between the molecules:

$$f_{att} = -\frac{a}{r^7} \quad (31.10)$$

Here the minus sign indicates that we are dealing with a force of attraction, and a is a factor depending upon the structure of the molecules and the kind of attraction force between them.

2. Experiments in the compression of bodies show that forces of repulsion appear between molecules that are brought close together. These forces are also of electrical origin. It was found that here the forces of repulsion between positively charged nuclei begin to play an essential role. Besides, quantum mechanical effects are manifested, which are also beyond the scope of our book.

As molecules are brought closer together, the forces of repulsion vary rapidly, much faster than the forces of attraction. This can be expressed approximately as

$$f = \frac{b}{r^9} \quad (31.11)$$

Here factor b depends upon the same things that factor a does. A more accurate formula is obtained in quantum mechanics but we shall not concern ourselves with it here.

3. Naturally, the molecule is subject to the action of the resultant force

$$f = -\frac{a}{r^7} + \frac{b}{r^9} \quad (31.12)$$

The sign of this force depends upon the distance between the molecules. A certain distance $r = d$ exists at which the repulsion forces counterbalance the attraction forces. It follows from equation (31.12) that $f = 0$ if $d = \sqrt[7]{b/a}$. When $r < d$ the resultant force is positive, i.e. the repulsion force exceeds attraction. When $r > d$, the opposite is true: attraction is stronger than repulsion and the resultant force is negative.

It can be shown that at $r_0 = d\sqrt[7]{9/7} = 1.134d$ molecules are attracted to each other with a maximum force $f_{max} \approx -\frac{a}{11d^7}$ (Fig. 31.4). At $r = 1.5d$ the force of attraction is reduced to about 1/3 of the maximum value; at $r = 2d$, to about 1/16 and at $r = 3d$ to almost as small as 1/250. If the distance is reduced to $r = 0.9d$, the force of repulsion is $f \approx 5|f_{max}|$ and at $0.7d$ it is $f \approx 140|f_{max}|$.

4. Such a strong dependence of the force on the distance signifies that *molecular forces are short-ranged*: molecules interact only with

their closest neighbours. If molecules are represented by closely packed spheres (Fig. 31.5), which, in general, conforms to the packing of molecules in many liquids and solids, the central black molecule in the drawing interacts only with the first layer of molecules surrounding it. With an error not exceeding 6 per cent we can neglect the interaction of the molecule with the second layer and, all the more, with the third and subsequent layers.

This is what explains all the features of molecular forces that were discussed in Sec. 31.2 and, in particular, the independence

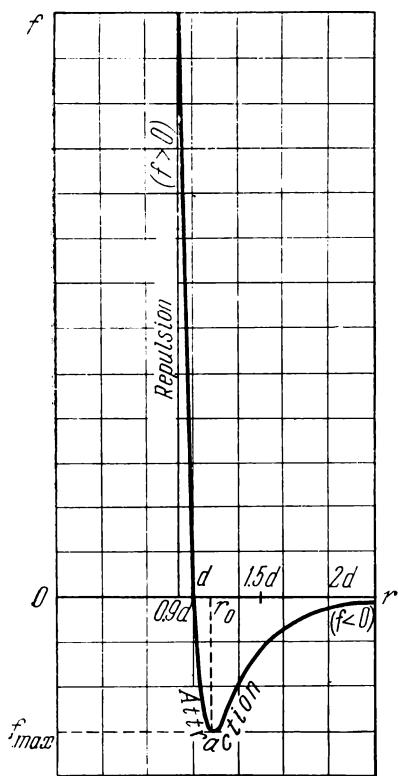


Fig. 31.4

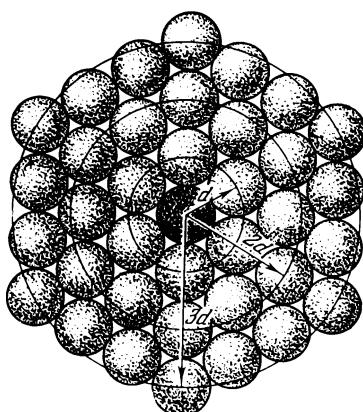


Fig. 31.5

of the magnitude of molecular forces from the total number of molecules in the system and the lack of molecular interaction in gases (except when they are under very high pressures).

31.5. THE POTENTIAL ENERGY CURVE OF MOLECULAR INTERACTION

1. Knowing the nature of molecular forces, we can plot a graph of the energy of molecular interaction, the so-called *potential energy curve* (see Sec. 19.5). We should take into account the properties of such curves, which were discussed in that section.

First of all, according to equation (19.12), when $r < d$, the potential energy curve should make an obtuse angle with the axis of

abscissas because of the positive force of repulsion. For the same reasons, when $r > d$, the curve should be inclined at an acute angle to the axis of abscissas because of the force of attraction acting in this case. Hence the curve has a minimum at $r = d$.

This affirms the statement made in Sec. 31.2 about the system of molecules possessing minimum energy when it is in a state of equilibrium.

2. Since the repulsion force varies more strongly with the distance than the attraction force, the curve is nonsymmetrical. To the left of the minimum ($r < d$) it is very steep. To the right of the minimum ($r > d$) it increases at first, but much less steeply, and then flattens out because the force of interaction is practically zero when $r > 2d$.

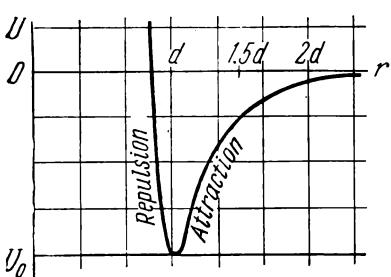


Fig. 31.6

As we know, the zero level of potential energy can be chosen arbitrarily. It proves convenient to assume that the potential energy equals zero when the molecules are infinitely distant from each other. Then it will practically equal zero beginning with $r > 2d$.

3. The potential energy curve of molecular interaction is shown in

Fig. 31.6. Here U_0 is the minimum energy of interaction between molecules, which they would have if located at a distance of d from each other. An analysis of this curve gives us an idea of many properties of solids and liquids. We shall discuss one of these properties, namely, the cause of their thermal expansion, in the next section.

31.6. THERMAL EXPANSION OF SOLIDS AND LIQUIDS

1. Experiments indicate that the volume of solids and liquids increases with temperature. Let us try to find the cause of thermal expansion. Though the structure and properties of solids and liquids differ, as we shall see in the following chapters, this is of no consequence with respect to the mechanism of thermal expansion.

At a given temperature the molecules of a solid or liquid are located at definite distances from one another and vibrate about the equilibrium positions. Let us plot the total energy of a molecule for a number of temperatures of the body on the potential energy curve (Fig. 31.7). Here ϵ_0 is the so-called zero-point energy, the minimum vibrational energy of the molecule at the absolute zero of temperature (Sec. 26.8); ϵ_1 and ϵ_2 are the vibrational energies of the molecule at temperatures T_1 and T_2 . We see in Fig. 31.7 that at absolute zero the molecules vibrate about equilibrium positions between which

the distance is equal to d . The vibrational energy increases with the temperature of the body. Therefore, if at a temperature T_1 the molecule vibrates between points A_1 and B_1 , at a temperature T_2 it will vibrate between points A_2 and B_2 . But, since the potential energy curve is nonsymmetrical, point B is displaced to the right much more than point A is displaced to the left. Hence, the equilibrium position is also displaced to the right with an increase in temperature. In other words, owing to the nonsymmetrical shape of the potential energy curve of molecular interaction, the distance between molecules increases with the temperature.

2. We have thus qualitatively explained the mechanism of thermal expansion. It is no simple matter, however, to obtain a quantitative relation from an analysis of the potential energy curve of molecular interaction. Let us refer, therefore, to the results of an experiment. We find that in a small temperature interval the relative change in volume is proportional to the change in temperature.

Thus

$$\Delta V/V = \beta \Delta T \quad (31.13)$$

where ΔT = change in temperature

β = coefficient of volume expansion, i.e. the relative change in volume per degree of temperature.

Putting $\Delta V = V - V_0$, where V_0 is the volume of the body at the temperature $t_0 = 0^\circ\text{C}$, and $\Delta T = t$ is the temperature of the body on the Celsius scale, we have

$$(V - V_0)/V_0 = \beta t \quad (31.14)$$

from which

$$V = V_0(1 + \beta t) \quad (31.15)$$

By analogy we have for the density

$$\rho = \rho_0(1 - \beta t) \quad (31.16)$$

3. If we heat a rod, its length increases. Using arguments quite similar to those above we can show that in a small temperature interval

$$l = l_0(1 + \alpha t) \quad (31.17)$$

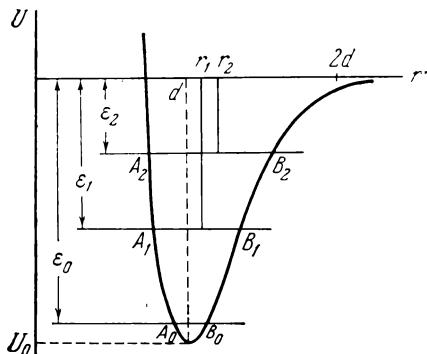


Fig. 31.7

where l_0 = length of the rod at 0°C

l = its length at the temperature $t^\circ\text{C}$

α = coefficient of linear expansion, i.e. the relative change in length per degree.

4. Let us show that for solids the coefficient of volume expansion is thrice the coefficient of linear expansion.

For this we first recall that $V = l^3$. Substituting into equation (31.15), we obtain

$$l_0^3(1 + \alpha t)^3 = V_0(1 + \beta t)$$

Cancelling $V_0 = l_0^3$ and removing the parentheses, we have

$$\beta = 3\alpha + 3\alpha^2t + \alpha^3t^2$$

The coefficient of linear expansion is very small for all solids; its order of magnitude ranges from 10^{-5} to 10^{-6} per 1°C . This means that in a small temperature range ($t \leq 100^\circ\text{C}$) we can neglect all terms containing α^2 and α^3 . Then

$$\beta = 3\alpha \quad (31.18)$$

Chapter 32 LONG-RANGE ORDER

32.1. THE MONOCRYSTAL

1. It has long been known that certain solids are found in nature in the form of *crystals*, i.e. bodies whose faces are regular polygons. The shape of the polyhedrons depends upon the conditions under which the crystals are formed, but the most typical property of a crystal is the fact that the angles between the edges of the faces and between the faces themselves are constant.

Thus, a crystal of common salt may have the form of a cube, parallelepiped or prism and also of irregular configuration (Fig. 32.1). But at each vertex three edges always meet at right angles to each other. In exactly the same way three edges, meeting at a vertex of a crystal of calcite (Iceland spar), form one acute angle of $78^\circ 08'$ and two obtuse angles of $101^\circ 52'$ each (Fig. 32.2).

A certain direction, called the *crystal*, or *optic*, *axis*, can be found in many crystals, such as Iceland spar or quartz. In Fig. 32.3 the optic axis is the direction parallel to the straight line MN . The significance and role of this axis will be dealt with in Chapter 64.

2. A large single crystal of more or less regular shape is called a *monocrystal* (from the Greek word *monos* meaning single). A typical feature of a monocrystal is its *anisotropy*, i.e. the difference in its physical properties in different directions.

3. The anisotropy of the mechanical properties of a monocrystal is manifested primarily by the fact that its strength varies in different directions. If, for instance, we fashion a sphere from some monocrystal and then break it up with a blow, it will split into pieces

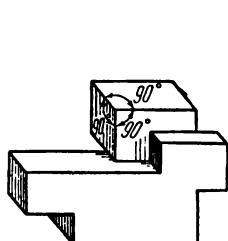


Fig. 32.1

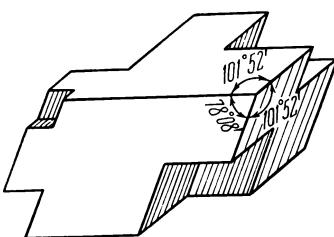


Fig. 32.2

whose faces make with each other the angles typical of the particular substance. This means that the monocrystal is weakest in these directions.

The elasticity of the monocrystal also differs in different directions. And, since the velocity of sound depends upon the elastic properties

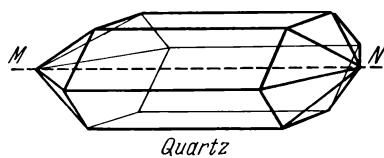
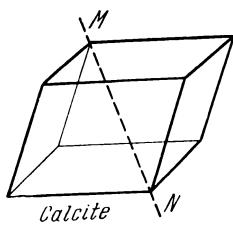


Fig. 32.3

of a substance and its density, the rate of propagation of sound in monocrystals, as confirmed by experiment, differs in different directions.

4. Monocrystals display an anisotropy of thermal properties by having different coefficients of linear expansion and thermal conductivity in different directions. An experiment for measuring the coefficient of linear expansion of small specimens is quite complicated and we shall not discuss it here. The results of such measurements on certain monocrystals are listed in Table 32.1.

The difference in heat conductivity of a monocrystal in different directions can be readily detected experimentally. This can be done by cutting a plate out of a quartz crystal parallel to its optic axis and coating the plate with a thin layer of wax. If we touch the plate

Table 32.1

Substance	Coefficient of linear expansion	
	Along the axis	Across the axis
Quartz	7.5×10^{-6}	13.7×10^{-6}
Cadmium	52×10^{-6}	20×10^{-6}
Zinc (crystal)	60×10^{-6}	13×10^{-6}
Tin (crystal)	30.5×10^{-6}	15.5×10^{-6}

with a red-hot needle, the wax will melt in the vicinity of the point of contact with the quartz. We will find that the melted area has the form of an ellipse with its major axis parallel to the optic axis. This indicates that the thermal conductivity is much higher along the axis than in the perpendicular direction. Their ratio is $K_{||}/K_{\perp} = 1.72$, where $K_{||}$ is the heat conductivity along the optic axis, and K_{\perp} is the same in the perpendicular direction.

32.2. THE POLYCRYSTAL

1. If special measures are taken, monocrystals can be obtained from a melt of a metal. If melted iron, for example, is simply cooled, the body obtained will not possess anisotropy. The reason for this can

be found when the surface of the iron is finely ground, etched with acid and then observed under a microscope (Fig. 32.4). The substance will be found to consist of separate grains with a size of the order of 10^3 to 10^4 Å. Each such grain is a crystal which is of irregular shape because its growth was interfered with by adjacent crystals. The consequent granular structure is said to be *polycrystalline* (from the Greek word *polys* meaning much or many).

Since the crystal axes of all the grains are oriented entirely at random, anisotropy cannot be displaced

over any appreciable distances. As a result, a polycrystal is *isotropic*: its properties are the same, on an average, in all directions.

2. The great majority of solids occurring in nature—rock, sand, metals, salts, etc.—are of polycrystalline structure. The formation



Fig. 32.4

of large monocrystals is an exceptional phenomenon occurring under especially favourable conditions. Also found in nature are solids like glass which do not have a granular structure. They are called *amorphous solids*. The structure and properties of such bodies will be discussed in Sec. 34.6.

32.3. THE CRYSTAL LATTICE. LONG-RANGE ORDER

1. From studies of the macroscopic properties of crystals, it was proposed as far back as the 18th century that a crystal is made up of unit cells of regular geometric shape. But only in 1912 did Max von Laue (1879-1960) confirm this hypothesis by discovering the diffraction of X-rays in crystals. It was found that the particles of which a crystal consists form a regular crystal lattice. In this lattice we can distinguish a certain cell of minimum size that could be transposed parallel to itself, each time a distance equal to the length of its edge, to form, in this manner, as large a monocrystal as desired. This cell is called the *unit cell*, and the length of its edge is equal to the *lattice constant* of the crystal (see Sec. 62.7).

Each vertex of the cell is called a *lattice point*; a straight line passing through the points is called a *lattice line* and a plane passing through the points is called a *lattice plane*. The planes and lines that are dotted most thickly with lattice points form the faces and edges of the crystal.

2. In 1881 E. S. Fedorov (1853-1919) established the fact that in nature there can be only 230 possible different kinds of spatial symmetry (Fedorov groups), which can be divided into 7 crystal systems or into 32 classes. These classes differ in the shape of their unit cells, by the angles between the edges and faces and by the ratio of the lengths of the edges.

The classification proposed by Fedorov was unquestionably affirmed by means of X-ray structural analysis. Not a single crystal has yet been found in nature that does not belong to one of the Fedorov groups.

3. As a rule the centres of atoms, ions or molecules are located at the points of the crystal lattice. As we shall see in the following, not all the atoms are at lattice points. Of importance, however, is the fact that in the crystal lattice we observe a regular distribution of matter, as well as of electric charges and their associated electric fields.

The lattice spacing is very small: in metals, for example, it ranges from 3 to 7 Å, which corresponds to several layers of atoms since the distance between two atoms in a solid is about 1 Å. But even in polycrystals the grains are of a size of the order of 10^3 to 10^4 Å while monocrystals are very much larger.

We must arrive, then, at the conclusion that in crystalline bodies the regular arrangement of the elements of the space lattice is repeated over hundreds, thousands and tens of thousands of layers. This is why we say that in a crystal we observe *long-range order* in the packing of atoms or other particles (ions, molecules, etc.) of which the unit cells of the crystal consist.

32.4. DEFECTS IN PACKING AND THE BLOCK STRUCTURE OF CRYSTALS

1. If a crystal had ideal structure, long-range order would be observed over unlimited distances. However, no such ideal crystals are formed in nature. In any real crystal we find repeated violations of long-range order, which are developed when the substance crystallizes.

One of the most important kinds of violation of regular structure in a crystal is the defect called a *dislocation*.

2. In an *edge dislocation* an extra lattice plane is formed (Fig. 32.5) in some part of the crystal. As a result the crystal is divided into two blocks on the interface of which the dislocation core, or centre, lies. The core is designated by the sign \perp .

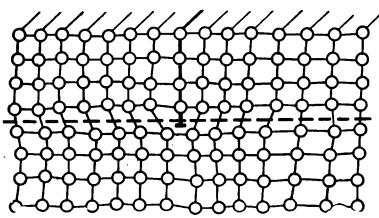


Fig. 32.5

two blocks on the interface of which the dislocation core, or centre, lies. The core is designated by the sign \perp .

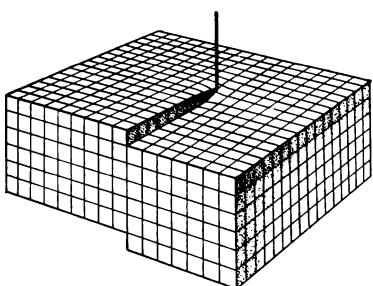


Fig. 32.6

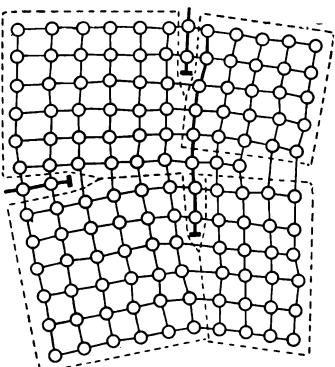


Fig. 32.7

A *screw dislocation* is formed when a lattice line of one of the blocks seems to slip one lattice constant above or below its normal direction. Maximum distortion is along the axis of the dislocation

(Fig. 32.6). At other parts of the crystal, especially those distant from the axis, long-range order is unviolated.

3. When there is only a small number of dislocations, the monocrystal formed consists of large blocks within each of which there is ideal long-range order. Sometimes the blocks are found to be turned slightly with respect to one another as illustrated in Fig. 32.7.

4. Violations of long-range order may also occur when a lattice point is occupied by a foreign atom (Fig. 32.8) or is vacant (Fig. 32.9) or a foreign atom occupies the spaces (interstices) between the lattice

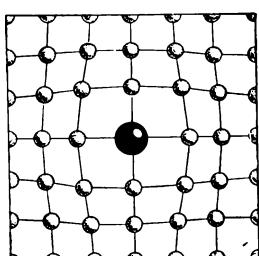


Fig. 32.8

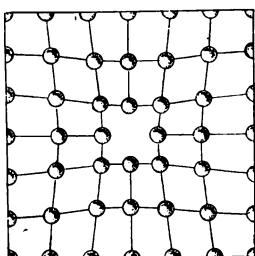


Fig. 32.9

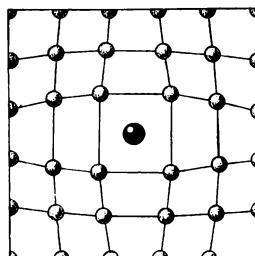


Fig. 32.10

points and lines (Fig. 32.10). Such defects may occur either at the place where dislocations are developed (and where the dislocations are often the cause of these defects) or inside a block.

The presence of defects in a crystal strongly affects its properties. For example, only 0.1 per cent impurities in the crystal lattice of germanium or silicon has no practical effect on the structure of the crystal but changes its electrical resistance very substantially, a thousandfold. The cause of this phenomenon will be discussed in Chapter 77.

In the subsequent sections we shall deal with the influence of defects of the crystal on its thermal and mechanical properties.

32.5. MOTION OF DEFECTS AND DIFFUSION

1. Experiments show that when two solids are in close contact for a long time, they diffuse into each other. It has proved especially convenient to observe diffusion phenomena in solids by means of "labelled", or "tagged", atoms which are radioactive isotopes of the given substance. But it is not difficult to show that diffusion is almost impossible in an ideal crystal. In such a crystal all the lattice points are occupied by the "home" atoms. Therefore, the intrusion of a "foreign" atom into a lattice point must be accompanied by a considerable amount of work done to extract the atom already

at the point. The available energy is not usually sufficient for this work. The intrusion of a foreign atom into the interstices between the lattice points also requires considerable energy because it is necessary to "shove" aside the neighbours and to overcome the substantial repulsive forces this causes.

2. It is an entirely different matter in a real crystal. The presence of defects of the dislocation type and lattice vacancies makes it much easier for foreign atoms to intrude into the crystal. The atoms may either occupy a vacancy, or get into the region of an edge or screw dislocation. Here the energy of thermal motion turns out to be sufficient for the atom to enter the region of a defect.

The same is true of defects of the substitutional type (see Fig. 32.8). Here, at a sufficiently high temperature, a process is quite probable in which the foreign atom and an atom of the base substance change places. This leads to movement of the foreign atom through the crystal.

As we see, the process of diffusion in a crystal consists in the motion of defects in the lattice. Naturally, the greater the number of defects in the lattice, the more probable diffusion will be. In particular, diffusion progresses at a much higher rate in polycrystals, all other conditions being equal, than in monocrystals.

3. It was found that diffusion is accelerated when the temperature of the crystal is raised. This is due to the more intensive vibration of the molecules at higher temperatures. The average distances between them are increased and, upon a favourable displacement of adjacent atoms or molecules in opposite directions, a third atom can "squeeze" through between them and occupy a new position in the lattice. This explains the possibility, in principle, of diffusion in an ideal crystal, though the probability is very low. All the more will a high temperature facilitate diffusion in a crystal having defects.

32.6. DISLOCATION MOTION AND THE DEFORMATION OF THE CRYSTAL

1. It has been established that any, even elastic, and all the more, plastic deformations of a crystal are associated with the displacement (slip) of blocks in the crystal with respect to one another. Here too the movement of dislocations was found to play a most important part.

Let us look into the mechanism of shear deformation in a perfect crystal and in one having an edge dislocation.

2. Assume that we want to initiate a shear displacement of one unit cell along plane AB of an ideal crystal. We will have to shift all the atoms simultaneously, i.e. set atom 1 opposite atom 2', atom 2 opposite 3', etc. (Fig. 32.11). This will require a considerable

effort, equal to the product of the cohesive force between two atoms by the total number of contacting atoms over the whole area of shear.

Now assume that plane AB coincides with the interface between two blocks, which contains the core of an edge dislocation (Fig. 32.12). Here layers 3, 4 and 5 will practically not be shifted at all by the

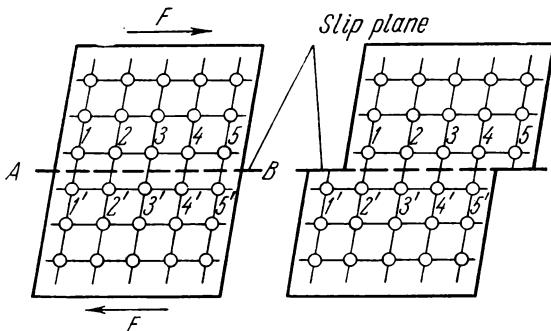


Fig. 32.11

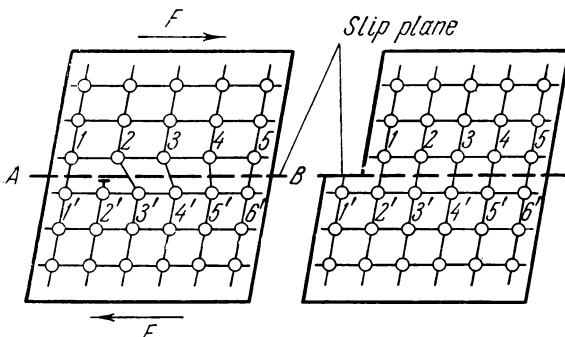


Fig. 32.12

force; layer 2 will be shifted one half of the distance between lattice points and occupy its natural position opposite point $3'$. Only layer 1 will have to be shifted a full lattice constant, and then only after layer 2 has already been displaced. The movement of layer 1 will be impeded only by its attraction to layer $1'$ but not by the repulsion of layer 2. Calculations indicate that the force required here for shear is only about one hundredth of that required when there are no dislocations.

3. We have seen that in deformation of a crystal order is improved to some extent in the arrangement of the planes of atoms, and the number of edge dislocations are reduced. Consequently, in the course

of its deformation the strength of a crystal should increase. Experience shows that the strength of materials actually is increased by such deformations as forging and rolling, but not a hundredfold as would follow from theory. It was found that screw dislocations play an important role here and that they do not always disappear in deformation.

4. The strength of crystals is drastically reduced by cracks and fissures occurring on their surfaces, evidently at the point of emergence of dislocations. Such fissures reduce the tensile strength. For instance, the rupture stress of rock salt crystals is reduced from the theoretical value of the order of 2×10^9 Pa to 4.5×10^6 Pa, i.e. to about 1/450 of the theoretical stress; for a monocrystal of zinc, from 2×10^{10} Pa to 5×10^7 Pa, etc. This was confirmed by experiments conducted by Academician A. F. Ioffe (1880-1960). He showed that if the surface of a crystal is continuously dissolved, thereby removing fissures and other surface defects, the strength of the specimen increases. Thus, in a tensile test of a crystal of rock salt in hot water he observed rupture taking place at stresses of about $(1.5 \text{ to } 1.6) \times 10^9$ Pa, which is quite close to the theoretical value.

Chapter 33

CLOSE PACKING OF PARTICLES

33.1. TYPES OF CRYSTALLINE BONDS

1. Up to this point we have concerned ourselves mainly with the geometric features of crystals, without paying any attention to either the nature of the particles forming the crystal or the character of the forces holding these particles at definite distances from one another. But an analysis of the physical nature of these forces is exactly what is required to enable us to understand both the geometry of crystal lattices and the difference in the properties of various types of crystals. Crystals can be arbitrarily classified into four types on the basis of the kind of forces that act between the particles making up the crystal: *ionic*, *atomic*, *molecular* and *metallic crystals*. Typical properties of these types of crystals will be described below, but first we must make the following remarks.

As mentioned previously, a condition for the stable state of any system of particles is minimum energy of the system under the given conditions (see Secs. 19.5 and 19.6). This is fully applicable to systems of molecules, atoms or ions that form a crystal. The particles are located at distances from one another for which the energy is minimal, i.e. at distances where the force of attraction equals the force of repulsion.

2. *Ionic crystals.* These include most of the inorganic compounds, for example, salts. The space lattices of these crystals consist of ions of alternating signs, positive and negative.

The mechanism of the formation of an ionic crystal can be explained by the example of common salt NaCl. An atom of sodium readily gives up one electron which becomes attached to an atom of chlorine. This forms two ions with unlike electric charges.

The ions thus formed have a spherically symmetrical electric field so that the Coulomb forces of attraction between them are of equal value in any direction. This enables us to deal with the structure of an ionic crystal as a system of spheres of a definite radius closely packed into a certain space lattice. The centres of the spheres are located at a distance from each other for which the interaction energy of the ions is minimal and the attraction force is counterbalanced by the forces of repulsion.

3. *Atomic crystals.* The space lattice of this type of crystal is formed by the close packing of atoms, mostly of identical atoms. No ions are formed in the interaction of identical atoms.

Typical examples of such crystals are the diamond and graphite, two different states of carbon, and also certain inorganic compounds (zinc sulphide ZnS, beryllium oxide BeO, etc.).

The atoms arranged in the space lattice are bound by *covalent* (otherwise, *homopolar*) forces of quantum mechanical origin. These forces have a sharply defined spatial directivity. Hence, the interaction between atoms depends upon their orientation with respect to each other. This is why the geometry of these crystals cannot be understood from an analysis of the packing of spheres.

4. *Molecular crystals.* These include many inorganic and most organic compounds in which the molecules retain their individuality in the crystallization process. Here molecules are located at the lattice points. Examples are bromine Br₂, iodine I₂, dry ice—solid carbon dioxide—CO₂, and also naphthalene, paraffins, fatty acids, etc. A typical feature of crystals of this kind is that the distances between particles within the molecule are considerably less than the intermolecular distances in the crystal. This is why each molecule can be pictured as a body bounded by parts of spherical surfaces, and the crystal itself as a close packing of molecules built up of truncated spheres (Fig. 33.1).

Between the molecules there are van der Waals forces resembling the interaction forces between induced dipoles (see Sec. 31.3). Owing to the weakness of these forces, molecular crystals are readily destroyed by the thermal motion of the molecules; they melt at very low temperatures and are readily evaporated. At standard atmospheric pressure crystals of iodine, for instance, evaporate (sublimate) at 39.6 °C above zero, and crystals of dry ice, at 78.5 °C below zero.

5. *Metallic crystals.* As its name implies, the given type of bond is a method of packing the atoms of metals in a crystal lattice. In the crystallization (freezing) of metals, their atoms draw closer together and large forces of quantum mechanical origin are developed between them. Owing to these forces, the electrons on the outer orbits (valence electrons) are separated from their atoms. However, in contrast to ionic bonds, here the electrons cannot become attached to other atoms to form positive ions because all the atoms of a metal

are identical. Therefore, the *valence electrons are collectivized*: they no longer belong to a single atom, as in the ionic bond, or to a pair of neighbouring atoms, as in the covalent bond, but to the whole crystal.

The metallic bond is accomplished by the attraction between the ionic lattice, formed by the positively charged ions, and the negatively charged electron cloud, which is called the *electron gas*.

We should note here that the availability of free electrons capable of being moved about within the crystal by an electric field provide for good electric conductivity of metals.

In the general case the packing of the ions in a metallic crystal cannot be reduced to the packing of spheres as for ionic or molecular bonds. There are cases, however, when it proves quite expedient to deal with metallic crystals as a system of most closely packed spheres of the same radius.

33.2. CLOSEST PACKING OF IDENTICAL SPHERES

1. Let us see how spheres of the same radius can be packed in the closest possible way. On a plane this will be in the manner shown in Fig. 33.2. Here each sphere is surrounded by six neighbours, this being the so-called *six-fold symmetry*. If we begin to pack the spheres in space we will have to turn and displace the same kind of flat lattice so that the spheres of the second layer enter the holes formed by spheres of the lower layer. But the spheres fill only half of the holes. They cannot get into the others. Figure 33.3 illustrates the packing of two layers of identical spheres, one above the other. The centres of the upper spheres are arranged above the holes marked with black dots in Fig. 33.2.

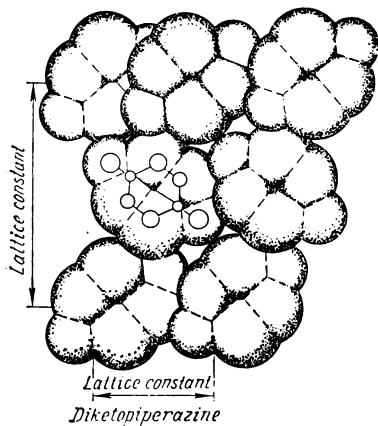


Fig. 33.4

In a similar way we can place a third layer above the first two. But here there may be two possible versions of the packing arrangement.

2. The centres of the third layer of spheres may be positioned above the centres of the spheres in the first layer. In this case the third layer is a repetition of the first. If we now continue by arranging

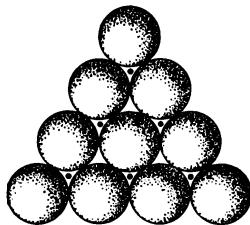


Fig. 33.2

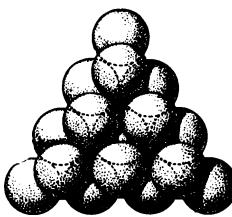


Fig. 33.3

the fourth layer above the second, the fifth above the third, the sixth above the fourth, etc., we will obtain a two-layer closest packing of identical spheres, the so-called *hexagonal close packing*. A model of such packing is shown in Fig. 33.4. The symbolic representation for such repeating layers is *ABABAB...*. The unit

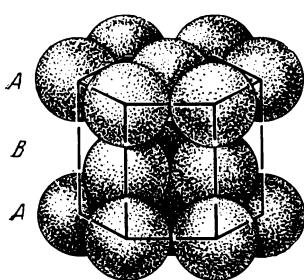


Fig. 33.4

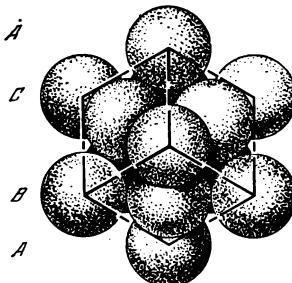


Fig. 33.5

cell is a right prism having a regular hexagon for its base. The height of the prism is greater than the sides of the base. Note that not all of the atoms are located at the lattice points.

Hexagonal close packing is typical of crystals of beryllium, magnesium, scandium, zinc, cadmium and many other metals and also inorganic compounds.

3. In the second version of packing the spheres of the third layer are arranged with the centres over the holes in the first layer that were not marked by black dots (Fig. 33.2). The fourth layer is a repetition of the first, the fifth of the second, the sixth of the third,

etc. The symbolic representation is *ABCABCABC...*. Packing of this type is illustrated in Fig. 33.5. The unit cell is a cube. The atoms are located at the corners of the cube (lattice points) and at the centres of the faces. For this reason the lattice formed by such packing is said to be *face-centred cubic*. Such metals as nickel, platinum, copper, gold, silver, aluminium, lead and, at high temperatures, iron have a lattice of this kind. Cobalt, for example, is found in the form of either a hexagonal or a face-centred cubic lattice. Such *polymorphism* is typical of many metals and other substances.

More complex multiple-layer combinations of the closest packing of spheres are also possible but we shall not discuss them here.

33.3. CLOSEST PACKING OF SPHERES OF DIFFERENT RADII

We shall confine ourselves to the discussion of a single example, the crystal of common salt. As mentioned earlier, this is an ionic crystal. X-ray structural analysis enabled the radii of the ions to be determined. They are 1.81 Å for chlorine and 0.98 Å for sodium. The large ions of chlorine form a close three-layer packing in which the large free spaces are filled by the smaller-sized ions of sodium (Fig. 33.6). It was found that each ion of sodium is surrounded by six ions of chlorine and, on the contrary, each ion of chlorine by six ions of sodium. With such packing the ions form a *face-centred cubic lattice*. Its unit cell is a cube with ions of one sign at the corners (lattice points) and in the centres of the faces, and ions of the other sign at the middle of the edges.

It is evident from the drawing that there are no molecules of NaCl in the crystal. Here the crystal constitutes a single giant molecule.

33.4. LATTICES WHICH CANNOT BE REPRESENTED AS THE PACKING OF SPHERES

1. At high temperatures the atoms of iron are packed in the closest three-layer structure, the face-centred cube. It was found, however, that at room temperatures the iron atoms are differently packed. X-ray structural analysis showed that here we have a cubic lattice with the atoms at the corners of the cube and one atom at its centre. This is called a *body-centred cubic lattice* (Fig. 33.7). This type of crystal lattice is found in all alkali metals, i.e. lithium, sodium, potassium, rubidium and cesium. It is also found in barium, chromium, molybdenum and tungsten.

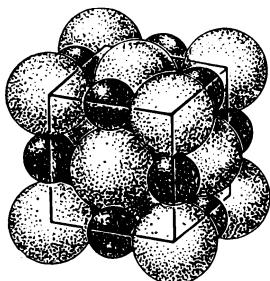


Fig. 33.6

A structure of this type cannot be represented as the close packing of spheres because spheres of the same radius cannot be closely packed into a cube with one sphere of the same radius at its centre. Evidently, the electron cloud of the positively charged ion, formed when the atom loses its valence electron, is not spherically symmetrical in this case. This should not surprise us. On the contrary, it is surprising that such complicated systems as atoms behave in many cases like spheres closely packed in the space lattice.

2. The physical properties of iron are different for the different types of packing of its atoms. With the closely packed face-centred

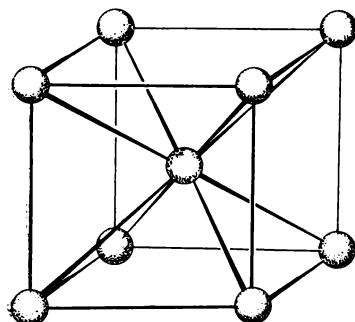


Fig. 33.7

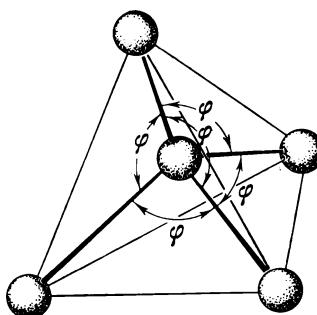


Fig. 33.8

lattice, when each atom has 12 neighbours, iron is hard and elastic. Iron with a body-centred lattice, when each atom has only 8 neighbours, is soft and plastic.

3. Tetravalent atoms (having four valence electrons), for instance carbon, can be bonded together by covalent forces, which have a pronounced spatial directivity. The atoms of carbon can form two types of space lattices.

In the first of these, typical of the *diamond*, four atoms are located at the vertices of a regular tetrahedron with a fifth atom located at its centre. The covalent forces are directed from the central atom to the vertices of the tetrahedron (these forces are shown by heavy lines in Fig. 33.8). As can be readily calculated, the angles between the bonds equal $\varphi = 109^\circ 29'$. According to data obtained by X-ray structural analysis, the length of the bonds is 1.54 Å. It is obvious that such a type of packing of the atoms cannot be obtained by close packing of identical spheres.

Crystals of silicon, germanium and white tin have a structure similar to that of the diamond. Their unit cells are cubes in which the atoms are located, not only at the corners, but at the centres of the faces and inside the cube (Fig. 33.9).

The second type of packing of carbon atoms, characteristic of *graphite*, is shown in Fig. 33.10. Here again each atom is bound to four neighbours but the forces and directions of the bonds differ. An atom has a strong bond with three other atoms that lie together with it in the same layer at angles of 120° , and a weaker bond with a fourth atom lying in the adjacent layer. This last bond makes

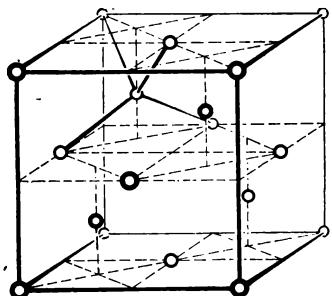


Fig. 33.9

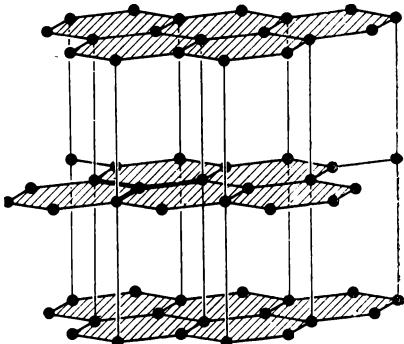


Fig. 33.10

an angle of 90° with the plane of the layer. The unit cells are right hexagonal prisms with edges 1.42 \AA and 3.35 \AA in length. Since the distance between the layers is 2.36 times more than between the atoms lying in one layer, the bonds between layers are weak.

4. The difference in structure of the diamond and graphite lattices explains the striking difference in their physical properties. Diamond is one of the hardest substances in nature; it can cut glass, as well as any hard rocks like granite, basalt and others. Graphite is very soft; it writes on paper because its layers slide easily with respect to one another. The diamond is an insulator, it has no free-electrons; graphite is a conductor of electric current. The diamond is transparent; graphite is opaque. The density of the diamond is 3500 kg/m^3 ; that of graphite is 2100 kg/m^3 .

Researches have shown that at pressures of the order of 60000 atm and a temperature above 1500°C , the atoms in the crystal lattice of graphite are brought closer together and rearranged so that this lattice may be transformed into the lattice of the diamond. This problem was solved in practice under the leadership of Academician L. F. Vereshchagin (b. 1909) at the Institute of High-Pressure Physics (USSR Academy of Sciences). "Diamond" presses developed in this institute have enabled synthetic diamonds to be obtained with a hardness about the same as that of natural ones.

33.5. THE STRUCTURE OF ICE

X-ray structural analysis has shown that crystals of ice have an exceptionally loose structure, which is illustrated in Fig. 33.11. Here the large spheres with holes conventionally represent atoms of oxygen, and the small necks between them are the hydrogen atoms. Each atom of oxygen is surrounded by four atoms of hydrogen, the latter serving as the binding links between two neighbouring oxygen atoms. The chemical formula for a molecule of water (H_2O) is retained, of course, but it is impossible to distinguish a separate molecule in the crystal.

The structure of ice, naturally, is not a closely packed one; the lattice has large "holes". When ice is melted and the crystals break down into separate molecules the substance becomes more dense. This is why the density of water at 0 °C and standard pressure is greater than that of ice under the same conditions:

$$\rho_{\text{water}} = 999.9 \text{ kg/m}^3 \text{ and } \rho_{\text{ice}} = 916.8 \text{ kg/m}^3$$

33.6. POLYMERS

1. High polymer substances (simply, *polymers*) consist of very long molecules in the form of chains. Each chain comprises regularly repeated groups of atoms bonded by covalent forces. The same forces bind together the groups of atoms into the chain. For example,

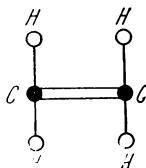


Fig. 33.12

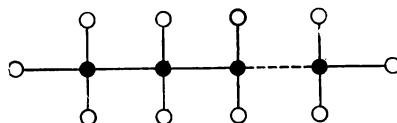


Fig. 33.13

a molecule of ethylene gas (C_2H_4) has the structure shown in Fig. 33.12. At high temperature and high pressure the polymer *Polyethylene* is produced. Its molecular chain may consist of thousands of monomers (Fig. 33.13).

The packing of such long molecules into a dense structure is a quite complicated problem. There are two possibilities: either the

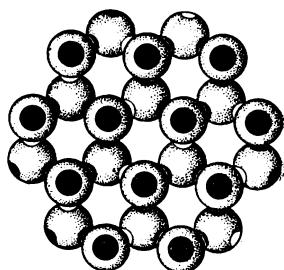


Fig. 33.11

chains are wound up into tight coils (*globules*) or they are extended in length and stacked in a *bundle*. The globules or bundles are then closely packed similar to the close packing of spheres.

2. Globules are of a size from 100 to 300 Å and can therefore be observed by an electron microscope. An electron micrograph of a globular crystal, the virus of the mosaic disease of the tobacco plant, is illustrated in Fig. 33.14. The crystals of many proteins are of similar shape.

We can see that the globules are packed according to the principles of the close packing of spheres. But X-ray structural analysis

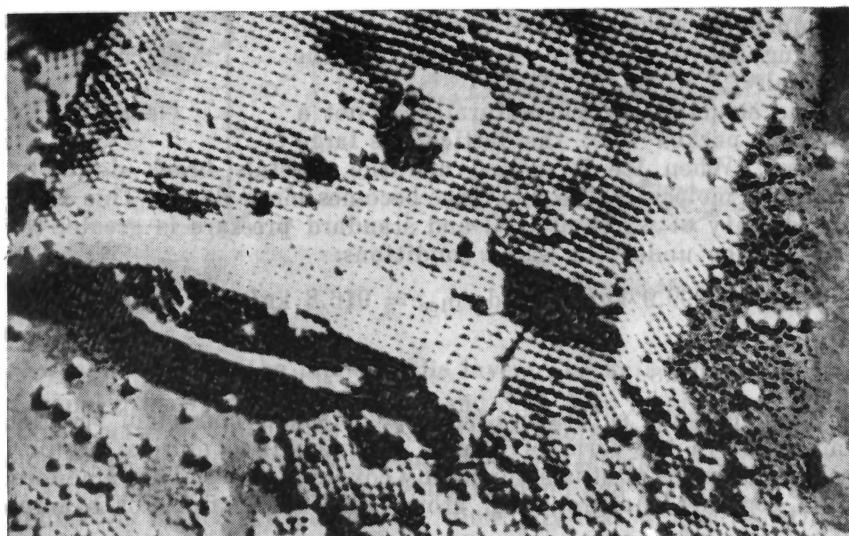


Fig. 33.14

yields a somewhat "blurred" picture. The fact is that in contrast to the crystals discussed above, where all the particles forming the crystal are identical, the particles of a globular crystal possess a certain individuality. The globules differ both in the length of the molecular chain, which varies for different complexes, and in the convolutions of the chain. Consequently, even though a globular crystal is characterized by long-range order, such order is impaired within each globule.

3. Bundle polymers sometimes have long-range order within the bundle. But this is a quite peculiar kind of order. If the molecular chain is represented by an elongated ellipsoid, then these ellipsoids

within the bundle can be arranged in three different ways (Fig. 33.15). In the first bundle long-range order is maintained both for the orientation of the centres of the chains as well as for the orientation of their direction in space. This is a crystal. In the second bundle the centres of the chains are in a definite order, but the chains themselves have random orientation. This is called gas-crystal packing. Finally, the third bundle has no long-range order whatsoever. It is said to be a liquid bundle.

The bundles can be easily folded and rolled up. This explains the high elasticity of rubber. Evidently, the bundles of rubber molecules are rolled up and the coiled molecules unwind when the rubber is stretched. When the external load is removed, the molecules coil up again, returning to the state of minimum energy.

4. A rigid system of the network type is often formed in the packing of macromolecules. This has given their name to *network polymers*.

A continuous network, extending over the whole polymer, makes the material hard, heat resistant and incapable of flowing or melting. Such thermosetting plastics are used to make articles that can withstand relatively high temperatures and high mechanical loads.

It is impossible here to describe in detail all the properties of polymers. A vast number of materials belong to the class of polymers. They range from natural materials (wood, rubber, proteins and nucleic acids) to synthetic ones (nylon, Plexiglas, etc.).

Extensive research is being conducted today on giant molecules, their properties, and methods of their production and application. New discoveries that are being made in this field have gone beyond the problems of pure chemistry and now concern themselves with biophysics and biology, for example, theories of heredity.

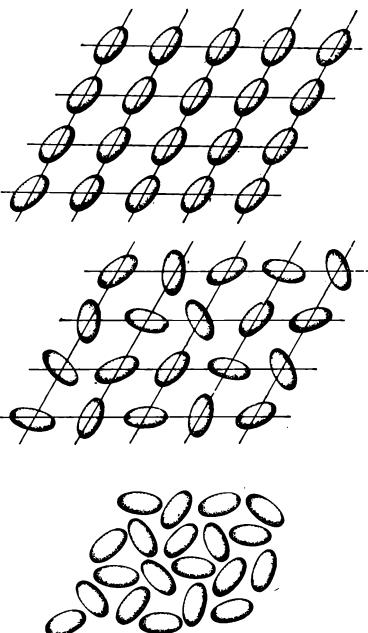


Fig. 33.15

Chapter 34

SHORT-RANGE ORDER

34.1. FEATURES OF THE LIQUID STATE

1. In the preceding chapters we discussed two limiting cases of the spatial packing of particles: those found in gases and in crystals. An ideal gas is characterized by a high degree of rarefaction and complete disorder in the spatial distribution of the particles. An ideal crystal, on the contrary, has a more or less dense packing of its particles and complete order in their arrangement in a certain space lattice (long-range order).

A liquid occupies a certain intermediate position between a crystal and a gas. Its molecules (or other particles) are packed quite densely, especially near the melting point, so that the density of a liquid is only 9 or 10 per cent less than the corresponding solid and in the case of molten metals, only 3 per cent less. The density of water is even more than that of ice by 9 per cent.

At the same time liquids are fully isotropic and it is impossible to find any anisotropic grains in them as are observed in polycrystals. This, naturally, means that long-range order is absent in liquids.

2. Up till the thirties it was thought that the structure of liquids has a complete lack of order and that it resembles that of a highly compressed gas. This point of view was upheld by the fact that the same laws could be applied, in many cases, in both hydro- and gas dynamics (see Ch. 30). These analogies, however, turned out to be very superficial, being based on a single property, the fluidity of liquids and gases. X-ray structural analysis of liquids (and amorphous solids) indicated that a definite order exists in the arrangement of particles in a liquid, especially when its temperature is near the freezing point. But this order is maintained only over short distances of about two or three layers of molecules. This permits us to speak of *short-range order* in the particle packing in liquids. It was found that many properties of liquids could be explained on this basis. This will be done, to some extent, below. At the same time a rigorous quantitative theory of the liquid state has not yet been fully developed.

3. A prominent role in the development of the new approach to the structure of liquids was played by the works of Ya. I. Frenkel (1894-1952). He not only suggested that liquids and crystals have properties closely resembling each other (at temperatures far from the critical point, see Sec. 35.5), but found that he could apply these ideas to derive various quantitative relationships for the processes

of diffusion, viscous flow and heat conduction, and also for an analysis of the freezing process and the properties of alloys. The general content of his ideas on liquids will be set forth in the following.

34.2. THE STRUCTURE OF A LIQUID AND ITS PROPERTIES

1. Several years ago J. D. Bernal (1901-1971) advanced the idea that in liquids we are dealing with so-called five-fold symmetry. This means that each particle in a plane model has five nearest neighbours (Fig. 34.1) and about eleven in spatial packing. This leads to many interesting consequences.

It can be readily seen that only short-range order is possible with such packing. Long-range order is unfeasible because a plane

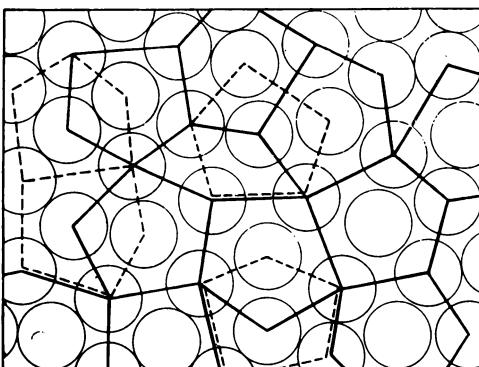


Fig. 34.1

can be covered with a network of regular triangles, squares or hexagons, but not pentagons. It is likewise impossible to fill space with polyhedrons having faces that are regular pentagons.

2. It follows that the pentagons that characterize the packing of particles in liquids have sides of somewhat different length. In crystals the particles are all located at distances d from one another, and the forces of attraction between them are counterbalanced by the forces of repulsion (see Sec. 31.4). But in liquids the distances between molecules on certain sides of the pentagon are somewhat greater than d and on other sides they are somewhat less. Only on an average do the distances equal d .

The energy of interaction between particles depends upon the distance between them (see Sec. 31.5). It is minimal at $r = d$, and increases when $r > d$ or $r < d$. Consequently, *for the same temperature, the internal energy of a crystal is less than that of its melt.*

This explains why in melting a crystal we must add to it the heat $Q_f = m\lambda$. Here m is the mass of the substance and λ is the *heat of fusion*, i.e. the energy which is necessary to transmit to unit mass of a substance at the melting point to change it from the crystalline to the liquid state.

A comparison of the internal energy of a crystal and its melt explains why, as a rule, liquids freeze (crystallize) at low temperatures. This is a manifestation of the general law: a system of particles tends to a state with minimum energy.

3. Polyhedrons with five-fold symmetry cannot form a regular crystal lattice when closely packed. They form certain compact spatial structures whose density may even be higher than that of a crystal. Bernal called these structures *pseudonuclei*, i.e. false nuclei. Between these nuclei there must necessarily be large gaps, or *holes*, as they were called by Frenkel. The presence of holes compensates for the surplus density of the nuclei, so that the overall density of liquids is usually less than that of crystals.

4. Pseudonuclei are insufficiently stable structures. Owing to the large number of holes, the particles can easily pass from one nucleus to another, forming polyhedrons of various shapes. This is why a liquid, in contrast to a crystal, has no single definite structure. It has a large number of equivalent, similar structures that correspond to the same energy of the system. Since the entropy of a system is determined by the number of different structures corresponding to the given state (see Sec. 28.7), it becomes evident that *at the same temperature, the entropy of a crystal is considerably less than that of its melt*.

34.3. MEAN RESIDENCE LIFETIME

1. The structure of a liquid changes continuously. This is due to the large number of holes in the liquid packing of the particles, which facilitate particle movement. A particle of liquid vibrates about an instantaneous equilibrium position for a length of time, which Frenkel so descriptively called the *residence lifetime*. Then the particle jumps into a hole and joins another pseudonucleus. These processes occur at random; various particles go over from one state to another in different time intervals, but for an immense number of particles we can introduce the concept of a *mean residence lifetime* τ as the length of time during which the structural configuration of the pseudonuclei, characterizing short-range order, remains unchanged.

The mean residence lifetime is inversely proportional to the probability that a particle will move to the place occupied by a hole and a hole to that occupied by a particle. The probability of such *migration* can be calculated as follows.

2. For a particle to break away from a pseudonucleus, it must possess kinetic energy in excess of the work that must be done to overcome the molecular forces of attraction. This energy is called the *activation energy*; let us denote it by ϵ . The higher the activation energy, the lower the probability of the migration of a particle.

On the other hand, this probability increases with the temperature of the liquid. Temperature characterizes the *average* kinetic energy of the molecules. But, owing to Maxwell's law of velocity distribution of particles, there is always a certain number of particles having a kinetic energy above the average value. As the temperature is raised, the percentage of such fast molecules increases as, consequently, does the percentage of particles having an energy that exceeds the activation energy. Thereby, the probability of migration is also increased.

3. We shall denote the ratio of the activation energy to the average kinetic energy of a particle by x (see Sec. 26.11). Thus

$$x = \frac{\epsilon_0}{\epsilon_{kin}} \approx \frac{\epsilon_0}{kT}$$

(since $\epsilon_{kin} \approx kT$). The probability of migration w will be some function of this ratio:

$$w = f\left(\frac{\epsilon_0}{kT}\right) = f(x) \quad (34.1)$$

An argument similar to the one in Sec. 26.11 will lead us to the conclusion that the probability of a particle moving to the position occupied by a hole is expressed by an exponential function of the form

$$f(x) = a^{-\alpha x}$$

The minus sign of the exponent has the following origin. As mentioned before, the probability of migration should decrease with an increase in the activation energy or with a reduction in the temperature. Consequently, $w = f(x)$ should be a decreasing function. And for the base $a > 1$ an exponential function decreases if the exponent is negative.

4. Thus, the probability of migration of a particle is

$$w = e^{-\frac{\epsilon_0}{kT}} \quad (34.2)$$

At low temperatures the magnitude of the exponent will be very large. It will be negative and the probability of migration will approach zero ($a^{-\infty} = 0$). At high temperatures, the exponent tends to zero and the probability of migration of a particle to unity ($a^0 = 1$).

5. The mean residence lifetime of a particle is inversely proportional to the probability of its migration. Thus, $\tau \propto 1/w$. From

equation (34.2) it follows that

$$\tau = \tau_0 e^{-\frac{e_0}{kT}} \quad (34.3)$$

where τ_0 is the mean residence lifetime at $e_0 = 0$ (or upon an unlimited increase in temperature, $T \rightarrow \infty$). Curves showing the dependence of the probability of migration and the mean residence lifetime on the temperature are shown in Fig. 34.2. It is evident that the mean residence lifetime decreases rapidly as the temperature is raised, while the probability that a particle will move to the place occupied by a hole increases.

34.4. DIFFUSION IN LIQUIDS

1. The presence of holes between the pseudonuclei enables us to explain qualitatively the mechanisms of diffusion and internal friction phenomena in liquids and to find the dependence of the coefficient of diffusion and the viscosity on the temperature.

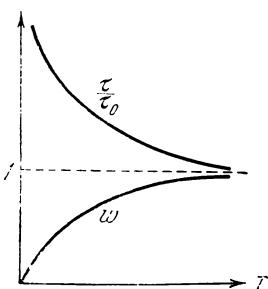


Fig. 34.2

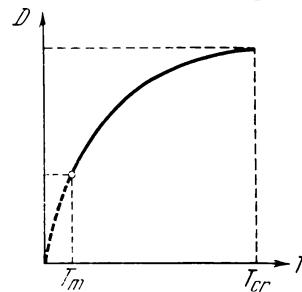


Fig. 34.3

The mechanism of diffusion in gases was discussed in Sec. 25.5. The formula for the coefficient of diffusion was derived there: $D = \frac{1}{3} \lambda v$, where λ is the mean free path of the molecules and v is the average velocity of random translational motion of the molecules (thermal motion).

2. Diffusion is much slower in liquids than in gases. But the coefficient of diffusion increases rapidly with the temperature. Experiments conducted with tagged atoms lead to the conclusion that the coefficient of diffusion in liquids (as well as in crystals) varies with the temperature according to the following law:

$$D = A e^{-\frac{e_0}{kT}} \quad (34.4)$$

where A = certain constant characteristic of the given substance

k = Boltzmann's constant

T = absolute temperature

e_0 = activation energy.

The value of ϵ_0 ranges from about 1 to 3 eV for liquid metals. A graph illustrating the dependence of the coefficient of diffusion on the temperature is given in Fig. 34.3.

3. On the basis of Frenkel's ideas the mechanism of diffusion in liquids can be explained as follows. Let us denote the mean free path of particles in a liquid by δ (this, in essence, is the distance between pseudonuclei). Then the average velocity of the particles in the liquid is $u \cong \delta/\tau$, where τ is the mean residence lifetime. By analogy with equation (25.10) we have

$$D = \frac{1}{3} \delta u = \frac{\delta^2}{3\tau} = \frac{\delta^2}{3\tau_0} e^{-\frac{\epsilon_0}{kT}} \quad (34.5)$$

This result complies with the experimental data [see equation (34.4)].

34.5. THE VISCOSITY OF LIQUIDS

1. Let us compare the dimensions of density $[\rho] = ML^{-3}$, the coefficient of diffusion $[D] = L^2 T^{-1}$ and viscosity $[\eta] = ML^{-1} T^{-1}$. We see that

$$\frac{[\eta]}{[\rho]} = [D]$$

from which

$$\eta = K\rho D \quad (34.6)$$

where K is a certain dimensionless factor.

Relationship (34.6) holds well for gases. For this case it was experimentally found that $K \cong 1$ and

$$\eta \cong \rho D \cong \frac{1}{3} \lambda v \rho \quad (34.7)$$

The viscosity of gases increases with the temperature. Indeed, if we substitute the values $\rho = m_0 n$, $\lambda = 1/n\sigma$ and $v = \sqrt{3kT/m_0}$ into equation (34.7), we obtain

$$\eta = \frac{1}{\sigma} \sqrt{\frac{m_0 k T}{3}} \quad (34.7')$$

Thus, the viscosity of gases is proportional to the square root of the absolute temperature.

2. As for liquids, it was found that equation (34.7') does not hold. It is not valid even qualitatively; in contrast to gases, the viscosity not only does not increase with the temperature, it drops sharply.

Frenkel was the first to point out this fact. He showed that the viscosity of liquids should be inversely proportional to the coefficient of diffusion. As proof we shall make use of equation (11.12), which

expresses the dependence of the friction drag on the viscosity, relative velocity and the characteristic dimensions of a body moving in a liquid.

When there is a velocity gradient (see Sec. 11.7) in the liquid, the pseudonuclei slide with respect to one another. The consequent interaction consists in a friction drag to which a hole is subject as it goes from one pseudonucleus to another. Therefore, the velocity v in formula (11.12) can be replaced by the average velocity u of the holes, and the characteristic dimension L by the hole size δ . Then the friction drag is

$$F = B\eta u \delta \quad (34.8)$$

where B is a numerical factor. But, according to equation (34.5), $u\delta = 3D$. Then we have

$$F = 3B\eta D$$

This last equation can also be written as

$$\eta = f/D \quad (34.9)$$

where $f = F/3B$ is a certain constant with the dimensions of a force. This shows that the viscosity of a liquid is inversely proportional to the coefficient of diffusion. Substituting into equation (34.9)

the value of the coefficient of diffusion
from equation (34.4) we obtain

$$\eta = \eta_0 e^{\frac{e_0}{kT}} \quad (34.10)$$

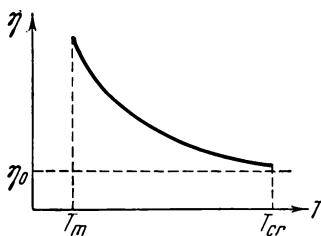


Fig. 34.4

where η_0 is the value of the coefficient of diffusion at $e_0 = 0$, i.e. in the absence of any activation energy. Thus, as we see from equation (34.10), the viscosity of a liquid really does drop sharply as the temperature is raised. A graph of this

relationship is shown in Fig. 34.4. The coefficient of diffusion vs temperature curve agrees well with experimental data.

3. It should be noted that the equations derived above for the dependence of the coefficient of diffusion (34.5) and the viscosity (34.10) on the temperature are valid only for a definite temperature interval—from the melting point to the so-called *critical temperature* T_{cr} above which a liquid cannot exist (see Sec. 35.5). This can also be seen in Figs. 34.3 and 34.4.

34.6. AMORPHOUS BODIES

1. Bodies are found in nature that are near to solids in mechanical properties but whose molecules have liquid packing, i.e. short-range order, as revealed by X-ray structural analysis. These substances

include, for instance, glass, various resins, pitch, congealed molten sugar (sugar candy) and many others. A feature of these substances is that they have no definite melting point. As their temperature is raised, these bodies gradually become soft, their viscosity drops and they begin to behave like ordinary viscous liquids.

The lack of a definite melting point and of long-range order in their packing leads us to the conclusion that amorphous bodies are not solids; they are supercooled and, thereby, very viscous liquids.

2. To understand the reason for the formation of amorphous bodies we can make use of the idea that liquids contain compact pseudonuclei between which there are holes. The internal energy of well packed pseudonuclei may turn out to be less than that of a crystal. Only because of the holes, possessing surplus energy, will the total internal energy of a liquid be found to exceed that of a crystal.

To form a crystal a part of the molecules must diffuse into the holes to build up the regular structure, long-range order. But at a low temperature the probability of diffusion is very low because the mean residence lifetime tends to infinity (see Fig. 34.2). Therefore, the formation of an amorphous body is found to be more probable in the rapid cooling of certain melts than the formation of an ordered crystal lattice.

3. We should note that in the amorphous state a body has surplus energy in comparison with the stable crystalline state. This is why the amorphous state is said to be *metastable*, i.e. not quite stable.

In the course of time amorphous bodies crystallize. For instance, crystals of sugar are formed in sugar candy after a certain time has passed. This explains why jam may become sugared when it is kept for a long time. In exactly the same way, glass "ages", i.e. polycrystalline grains are formed in it. When this happens, the glass loses its transparency and becomes brittle.

34.7. ENERGY OF THE SURFACE LAYER AND THE SURFACE TENSION OF LIQUIDS

1. Particles in the surface layer of a liquid have surplus energy when compared to the particles within the liquid. Inside a liquid each particle is surrounded on an average with Z nearest neighbours ($Z \approx 11$) while in the surface layer it has only one half as many (Fig. 34.5). At temperatures near the melting point the energy of interaction between two nearest molecules is $\epsilon \approx U_0$, where U_0 is the depth of the potential well of molecular interaction (see Fig. 31.7). Hence, the energy of a particle inside the liquid $\epsilon_{ins} \approx \approx Ze \approx ZU_0$, while the energy of the same particle at the surface of the liquid is $\epsilon_{sur} \approx \frac{1}{2} Ze \approx \frac{1}{2} ZU_0$. Since U_0 is negative, $\epsilon_{sur} > > \epsilon_{ins}$. Thus, the surplus energy possessed by a particle in the sur-

face layer with respect to the same particle within the liquid is expressed by

$$\Delta\epsilon_{sur} = \epsilon_{sur} - \epsilon_{ins} \cong -\frac{1}{2} Z U_0 = \frac{1}{2} Z |U_0|$$

Owing to the surplus energy, the average distance between molecules increases (Secs. 31.5 and 31.6). This means that the surface layer is looser than the interior parts of the liquid.

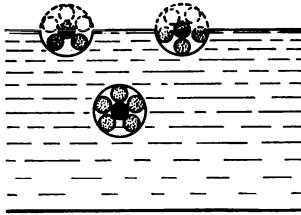


Fig. 34.5

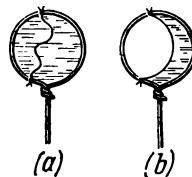


Fig. 34.6

2. Suppose that unit area of the surface layer contains n' particles. A surface layer of area A will possess a surplus energy of

$$\mathcal{E}_{sur} = \Delta\epsilon_{sur} n' A \cong \frac{1}{2} Z n' A |U_0| \quad (34.11)$$

The energy per unit area is

$$\alpha = \frac{\mathcal{E}_{sur}}{A} \cong \frac{1}{2} Z n' |U_0| \quad (34.12)$$

This quantity is called the *surface energy* of the liquid.

3. Any system of bodies tends to a state with minimal energy. For a liquid this means that in the absence of external forces it will assume a shape with the minimum surface for the given volume. This is why a drop of liquid becomes strictly spherical in the state of weightlessness whose conditions were discussed in Sec. 7.3. The tendency of a liquid to reduce its surface leads to the development of forces of *surface tension* at the boundary between a surface film and a solid. This can be demonstrated by a simple experiment. Let us tie a piece of thread to a wire loop as shown in Fig. 34.6a and dip the loop into a soap solution. The thread is located at random in the soap film that is formed. Now if we rupture the film at one side of the thread, the film on the other side will contract so that its surface is minimal (Fig. 34.6b). The thread is held in the stretched state by the forces of surface tension, directed into the film along tangents to its surface.

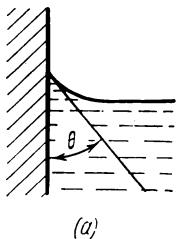
4. The *surface tension* σ is the ratio of the force F with which the film acts on the contour bounding it to the length l of this contour. Thus

$$\sigma = F/l \quad (34.13)$$

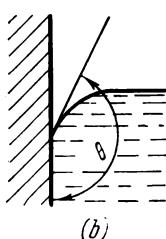
34.8. PRESSURE CAUSED BY THE CURVED SURFACE OF A LIQUID

1. In the contact of a liquid with a solid the liquid may either wet or fail to wet the surface of the solid. Mercury, for instance, readily wets the surface of clean metals but does not wet clean glass; water wets clean glass but does not wet greasy surfaces; etc.

The free surface of a liquid is curved at the boundary with a solid. The shape of the *meniscus* that is formed depends upon how the liquid wets the surface of the solid. Wetting liquids form a concave



(a)



(b)

Fig. 34.7

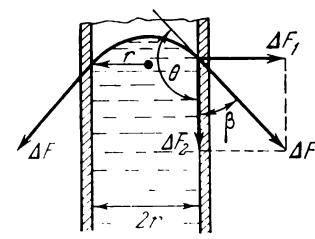


Fig. 34.8

meniscus; nonwetting liquids, a convex one. The angle between the wetted surface of the solid and the tangent to the meniscus at its point of intersection with the solid is called the *contact*, or *wetting*, angle θ .

If the liquid wets the surface of the solid, the contact angle is acute (Fig. 34.7a): $0 \leq \theta < \pi/2$; if not, it is obtuse (Fig. 34.7b): $\pi/2 < \theta \leq \pi$. The magnitude of the contact angle depends upon the chemical compositions of the solid, the liquid and the surrounding gas, upon the purity of these substances and their temperature. For ideal wetting $\theta = 0$; for ideal nonwetting $\theta = \pi$.

2. The curved surface exerts surplus pressure on the liquid. In calculating this pressure we shall confine ourselves to the case when the liquid is in a tube with the inside radius r (Fig. 34.8). The curved surface of the liquid forms a spherical segment whose surface makes the contact angle θ with the inner surface of the tube.

A portion of the contact boundary of the liquid with the solid tube wall of length Δl is subject to a surface tension force $\Delta F = \sigma \Delta l$. This force is directed along a tangent to the surface of the liquid. Let us resolve this force into two components: ΔF_1 directed into the wall and $\Delta F_2 = \Delta F \cos \beta = -\sigma \Delta l \cos \theta$ directed into the liquid. Component ΔF_2 causes a pressure, which is surplus with respect to the atmospheric pressure.

3. To find this surplus pressure we divide the force component $F_2 = -\sigma \times 2\pi r \cos \theta$ by the cross-sectional area $A = \pi r^2$ of a circle.

We have

$$p = \frac{F_2}{A} = -\frac{2\sigma \cos \theta}{r} \quad (34.14)$$

We see that the surplus pressure is negative if the contact angle is acute: at $0 \leq \theta < \pi/2$ we have $\cos \theta > 0$. If the contact angle is obtuse, i.e. $\pi/2 < \theta \leq \pi$, then $\cos \theta < 0$ and the surplus pressure is positive.

34.9. CAPILLARY PHENOMENA

1. Tubes with fine bores, a millimetre or less in diameter, are called *capillary tubes* (from the Latin *capillus* meaning a hair). If such a tube is immersed in a liquid which wets it, the liquid in the tube rises to a level higher than the free surface of the liquid in the wide

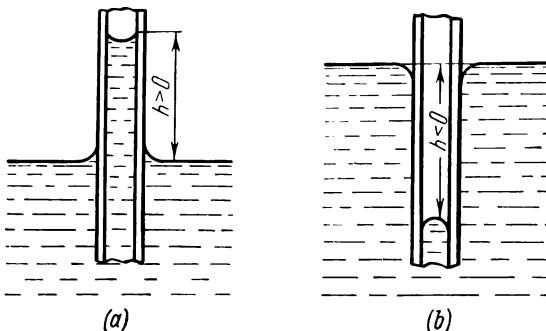


Fig. 34.9

vessel (Fig. 34.9a). A nonwetting liquid, on the opposite, will be depressed in the tube to a lower level than in the vessel (Fig. 34.9b).

The height of the capillary rise (or depression) is determined from the fact that the liquid in the tube will be in equilibrium if the sum of the hydrostatic pressure ρgh and the pressure under the curved surface $p = -\frac{2\sigma \cos \theta}{r}$ equals zero.

Thus

$$\rho gh - \frac{2\sigma \cos \theta}{r} = 0$$

from which

$$h = \frac{2\sigma \cos \theta}{r\rho g} \quad (34.15)$$

We can see that for a wetting liquid, where the contact angle is acute ($\cos \theta > 0$), the rise is positive; for a nonwetting liquid, the

contact angle is obtuse ($\cos \theta < 0$) and the rise is negative. The level of a nonwetting liquid will be lower in the tube than in the wide vessel.

2. Capillary phenomena are widely found in nature. They explain why certain bodies are hygroscopic, i.e. capable of absorbing moisture. Cotton wool, textiles, soils and concrete are hygroscopic owing to capillary action. This phenomenon is taken into account in construction practice. A layer of tar paper, asphalt or other waterproof material is provided between the foundation of a building and its walls to prevent the moisture from penetrating along the walls into the residential quarters.

The subsoil moisture rises in the capillaries of the soil to the surface, where it evaporates. To preserve the moisture in the soil the capillaries must be broken up. This is done by plowing and harrowing the fields.

34.10. ADSORPTION. THE REBINDER EFFECT

1. All the arguments and calculations that lead to the conclusion that a liquid has surplus energy in its surface layer (see Sec. 34.7) are equally valid for amorphous and crystalline solids. There is, however, one essential difference that arises from the anisotropy of a crystal. There are different numbers of particles (atoms, molecules or ions) on the various faces of a crystal, and the distances between these particles are also different. Hence, the surface energy and surface tension vary on the different faces of a monocrystal. There is naturally no anisotropy of the surface tension in polycrystals and amorphous bodies.

In the absence of external forces the shape of a liquid is fully determined by the forces of surface tension, and a free liquid becomes spherical. Owing to their negligible fluidity, this does not happen to crystals. This by no means implies, however, that surface tension forces play no role in crystals. On the contrary. It was found that they essentially influence such mechanical properties of crystals as hardness and strength.

2. Certain substances, said to be *surface-active*, readily concentrate on the surface of another substance, forming an extremely thin layer, usually a monomolecular one, i.e. a layer one molecule thick. This phenomenon is called *adsorption*. Butyl alcohol, for example, is readily adsorbed on the surface of water; mercury and tin on the surface of zinc; and the alkali metals (cesium, potassium and sodium) on the surface of mercury. It was found that adsorption is accompanied by a drastic reduction in surface energy which, in its turn, reduces the surface tension.

If 0.48 g of cesium are dissolved in a litre of mercury, the surface tension of the mercury is reduced from 0.470 N/m to 0.228 N/m,

i.e. by more than one half. But the concentration of cesium in the mercury is only 0.48×10^{-3} kg/13.6 kg, i.e. only 3.5×10^{-3} percent! In exactly the same way, the surface tension of water is reduced by one half when it contains a 4 per cent concentration of butyl alcohol.

3. It was also found that the adsorption of surface-active liquids on the surfaces of solids sharply changes the mechanical properties of the latter. Their strength, in particular, is substantially reduced. This is demonstrated by the following experiment.

At room temperature zinc is highly ductile, and a piece of sheet zinc, 2 or 3 mm thick, can be easily bent to a right angle without breaking it.

If we clean the surface of the zinc of oxides before bending it, and apply a drop of mercury of a mass not exceeding even a few milligrams, the character of the deformation will definitely change. A fine crack will originate under the drop of mercury, and the piece of zinc will fall apart into two pieces (Fig. 34.10).

Similar phenomena had been observed for a long time. Many cases were known in engineering practice when the strength of certain structures were essentially reduced upon changing their environment from air to some other medium.

Sometimes even the brand of lubricant influences the strength of a machine part. But no theory explaining this phenomenon had been developed.

4. In 1928 Academician P. A. Rebiner (1898-1972) advanced the idea that the change in the strength of solid bodies interacting with surface-active substances is caused by the reduction of the surface tension. In general terms the mechanism of this phenomenon can be described as follows.

Defects of a crystal, such as dislocations, grain boundaries and microfissures, possess surplus surface energy. Since the surface tension is reduced in the adsorption of surface-active substances, the surface of these defects will correspondingly increase so that the total energy of the surface layer, equal to the product of the surface energy by the surface area [see equation (34.12)], remains constant. This increase in the surface of the defects naturally reduces the strength of the body.

5. The fact that the deformation and failure of solid bodies are facilitated when their surface energy is reduced by the adsorption

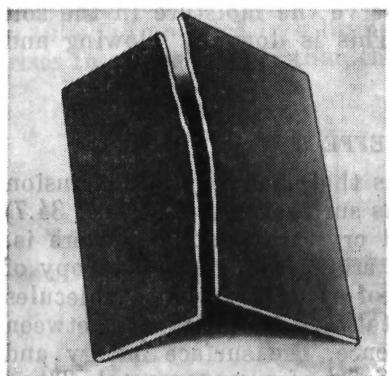


Fig. 34.10

of surface-active substances is called the *Rebinder effect*. This effect has many practical applications.

It was found that in the machining of metals the addition of surface-active substances to the cutting fluids (coolants) enables the cutting speeds to be substantially increased (sometimes by 30 or even 40 per cent). Another example is the adsorption method of strength reduction in drilling wells. Oil wells are often several kilometres deep and pass through extremely hard rock. To facilitate the operation of the boring tools, solutions of surface-active substances are pumped into the hole. This makes well drilling appreciably easier.

In connection with the application of liquid metals as heat-transfer agents in nuclear power reactors, the problem has been posed today of studying the strength of metals in the presence of melts of other metals. There are similar problems in rocket engineering. A proper choice of pairs of metals, one in which the liquid phase is adsorbed to the least possible extent on the solid phase, enables any reduction in the strength of the machine components to be avoided.

Intensive research was recently begun with the aim of prolonging the service life of catalysts. Catalysts participate only indirectly in chemical reactions; they do not enter into the reaction. It would seem that they last forever. This is not so. They wear at a high rate and are consumed in huge amounts in the chemical industry. It is impossible, in principle, to eliminate adsorption on catalysts; otherwise they would cease being catalysts. Therefore, other methods of increasing their strength are required. One of them is to eliminate internal stresses in the granules of these materials. This is accomplished by the vibratory compaction of catalysts from powders. A second way is to select the proper temperature conditions for their heat treatment.

The Rebinder effect is an extremely general phenomenon that characterizes physicochemical surface effects in solids and disperse media. A further development of surface phenomena theory and research in this field should solve the problem of producing materials with specified properties and, in particular, high-strength materials.

Chapter 35

VAPOURS

35.1. VAPORIZATION

1. As we know from experience, liquids are capable of evaporating, i.e. of going over to the gaseous state. The rate of vaporization is evaluated by the number of molecules passing from unit area of the-

surface into a vapour in unit time. This rate differs for various liquids. Ether poured into a saucer will evaporate in several minutes at room temperature; under the same conditions water will remain for several hours and mercury for several years. The vaporization of liquids takes place at any temperature but its rate increases rapidly as the temperature is raised. This feature is made use of in practice. If you wish to dry something rapidly, you put it next to a heated body, for instance, a stove or a heating radiator.

2. Not only liquids, but solids are also vaporized. Naphthalene and camphor, for example, evaporate at room temperature and standard pressure, omitting the liquid state. Crystals of bromine or iodine evaporate in exactly the same way, especially if they are heated. Ice also evaporates. If wet washing is hung outside in freezing weather, the water freezes and then the ice evaporates, leaving the clothes dry.

The evaporation of crystalline bodies is called *sublimation*, or *volatilization*.

The rate of sublimation also depends upon the temperature and chemical nature of the substance. Molecular crystals, like bromine and iodine, are readily sublimated because their forces of molecular attraction, holding the molecules in the crystal lattice, are quite weak. Ice also sublimates easily because it has a very loose crystal lattice. Closely packed ionic, covalent and metallic crystals, on the contrary, sublimate very poorly.

3. The mechanism of vaporization (and sublimation) can be described as follows. A molecule (or atom) at the surface of a liquid or crystal is pulled inward by the forces of molecular interaction and is, thereby, held at the surface. For a particle to fly away from a substance its kinetic energy must be more than the *energy of evaporation* u_0 , which is equal to the work that must be done to overcome the forces of molecular attraction and to withdraw the molecule to a distance from the surface at which molecular forces are no longer manifested. Therefore, the condition for evaporation (or sublimation) is $\epsilon_k \geq u_0$ or

$$\frac{m_0 v^2}{2} \geq u_0 \quad (35.1)$$

It follows that not any molecules can leave the surface layer, but only the faster ones whose velocity satisfies condition (35.1). This explains why a liquid is cooled in evaporation. We know that the temperature of a body is determined by the average kinetic energy of motion of its molecules. Naturally, if the faster molecules fly off, the average energy of the remaining molecules is reduced and this means a drop in temperature.

4. The probability of evaporation can be found by means of an argument similar to the one used in Sec. 34.3 to determine the pro-

bility of the migration of a particle into a hole. Substituting u_0 for ε_0 , we obtain

$$w_{evap} = e^{-\frac{u_0}{kT}} \quad (35.2)$$

The rate of evaporation is proportional to the probability of evaporation (or vaporization) because the more probable that molecules will fly off the surface of a body into a vapour, the greater the number of molecules leaving unit area of the surface and being transformed into vapour in unit time. Thus, the rate of vaporization $G \propto w_{evap}$ or

$$G = G_0 e^{-\frac{u_0}{kT}} \quad (35.3)$$

In this equation G_0 is proportional to the number of molecules capable of flying off the surface layer, i.e. molecules whose velocity satisfies condition (35.1). The number of such molecules naturally increases with the temperature. It can be shown that $G_0 = BT^\alpha$, where B is a constant depending upon the chemical composition of the substance and $\alpha \approx 1/2$. Substituting into equation (35.3), we obtain the final expression for the rate of vaporization:

$$G = BT^\alpha e^{-\frac{u_0}{kT}} \quad (35.4)$$

35.2. SATURATED VAPOUR

1. If a sufficiently large amount of liquid is put into a closed vessel, a part of the liquid will vaporize and the rest will remain unchanged for an unlimited period of time. But the process of vaporization cannot cease; molecules whose energy satisfies condition (35.1) will continue to leave the surface of the liquid. Consequently, together with the process of vaporization, there is a compensating process continuously taking place: the condensation of vapour into liquid. The rate of condensation is determined by the number of molecules going over from the vapour into the liquid through unit surface area in unit time.

Thus, in a closed vessel *dynamic equilibrium* is established after a certain length of time between the processes of vaporization and condensation. The rate of condensation becomes equal to the rate of vaporization and, beginning with this instant, the amount of liquid remains constant.

Vapour which is in a state of dynamic equilibrium with its liquid (or crystal) is said to be *saturated vapour*.

2. Let us find the parameters on which the concentration of molecules of saturated vapour depends. Here we must begin with the fact that the rate of vaporization does not depend upon the con-

centration of the molecules of vapour above the liquid, but that the rate of condensation is proportional to this concentration. The more molecules per unit volume of the vapour above the liquid, the higher the probability that a molecule will come near to and be absorbed by the surface layer of liquid. Thus, the rate of condensation is

$$g = \beta n \quad (35.5)$$

where β = a certain coefficient

n = concentration of the molecules of vapour.

In dynamic equilibrium the rate of condensation equals the rate of vaporization. Equating the right-hand sides of equations (35.4) and (35.5), we obtain the equation for the concentration of molecules of the saturated vapour:

$$n_{sat} = AT^\alpha e^{-\frac{u_0}{kT}} \quad (35.6)$$

3. We can find the density of saturated vapour if we multiply the concentration by the mass of a molecule. Thus

$$\rho_{sat} = m_0 n_{sat} = Am_0 T^\alpha e^{-\frac{u_0}{kT}} \quad (35.7)$$

A graph of this relationship is shown in Fig. 35.1. We see that the density of saturated vapour increases sharply with the temperature.

35.3. THE PRESSURE OF SATURATED VAPOUR

1. Substituting equation (35.6) for the molecule concentration in formula (26.10), we obtain the expression for the pressure of saturated vapour:

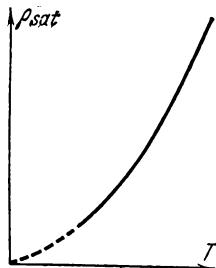


Fig. 35.1

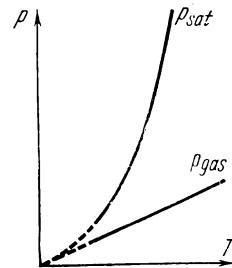


Fig. 35.2

rated vapour:

$$p_{sat} = n_{sat} kT = AkT^{\alpha+1} e^{-\frac{u_0}{kT}} \quad (35.8)$$

The graph of this relationship is given in Fig. 35.2. Also plotted here for comparison is the graph of the dependence of the pressure

of an ideal gas on the temperature (at constant volume). The pressure of an ideal gas increases linearly with the temperature; that of saturated vapour increases much faster. The reason is that the increase in pressure of a gas is due only to the increase in the kinetic energy of the molecules. Not only the kinetic energy of the molecules increases in a saturated vapour but their concentration as well.

2. It should be pointed out that equation (35.8) is not fully rigorous because in its derivation we used equation (26.10), which is for the pressure of an ideal gas. In the latter the forces of molecular interaction are neglected due to the large distances between the molecules. But the molecule concentration of saturated vapour may be quite large and the distances between molecules small enough so that the molecular forces are appreciable. Nevertheless, the properties of saturated vapour are, in the main, correctly described by

Table 35.1

PRESSURE AND DENSITY OF SATURATED STEAM

t °C	p , mm-Hg	ρ , g/m³	t °C	p , mm-Hg	ρ , g/m³	t °C	p , mm-Hg	ρ , g/m³
-30	0.28	0.33	0	4.58	4.84	30	31.82	30.3
-29	0.31	0.37	1	4.93	5.22	31	33.70	32.1
-28	0.35	0.41	2	5.29	5.60	32	35.66	33.9
-27	0.38	0.46	3	5.69	5.98	33	37.73	35.7
-26	0.43	0.51	4	6.10	6.40	34	39.90	37.6
-25	0.47	0.55	5	6.54	6.84	35	42.18	39.6
-24	0.52	0.60	6	7.01	7.3	36	44.56	41.8
-23	0.58	0.66	7	7.51	7.8	37	47.07	44.0
-22	0.64	0.73	8	8.05	8.3	38	49.69	46.3
-21	0.70	0.80	9	8.61	8.8	39	52.44	48.7
-20	0.77	0.88	10	9.21	9.4	40	55.32	51.2
-19	0.85	0.96	11	9.84	10.0	45	71.88	65.4
-18	0.94	1.05	12	10.52	10.7	50	92.5	83.0
-17	1.03	1.15	13	11.23	11.4	55	118.0	104.3
-16	1.13	1.27	14	11.99	12.1	60	149.4	130
-15	1.24	1.38	15	12.79	12.8	65	187.5	161
-14	1.36	1.51	16	13.63	13.6	70	233.7	198
-13	1.49	1.65	17	14.53	14.5	75	289.1	242
-12	1.63	1.80	18	15.48	15.4	80	355.1	293
-11	1.78	1.96	19	16.48	16.3	85	433.6	354
-10	1.95	2.14	20	17.54	17.3	90	525.8	424
-9	2.13	2.33	21	18.65	18.3	95	633.9	505
-8	2.32	2.54	22	19.83	19.4	100	760.0	598
-7	2.53	2.76	23	21.07	20.6	120	1489	1091
-6	2.76	2.99	24	22.38	21.8	140	2710	1890
-5	3.01	3.24	25	23.76	23.0	160	4636	3084
-4	3.28	3.51	26	25.21	24.4	180	7521	4782
-3	3.57	3.81	27	26.74	25.8	200	11664	7099
-2	3.88	4.13	28	28.35	27.2			
-1	4.22	4.47	29	30.04	28.7			

expression (35.8). This is evidenced by Table 35.1 and by Fig. 35.3. The latter shows a curve of the dependence of the pressure of saturated steam (water vapour) on its temperature, which was plotted from the data listed in the table. It is evident that the character of this dependence coincides, in general, with the theoretical expression (35.8) plotted in Fig. 35.2.

3. An analysis of equation (35.8) shows that the pressure of saturated vapour depends only on its chemical composition and temperature but not on the free volume of the vessel above the liquid.

This conclusion is confirmed by experiments and can be easily explained as follows.

Assume that we increase the free space above the liquid by raising the piston in the device illustrated in

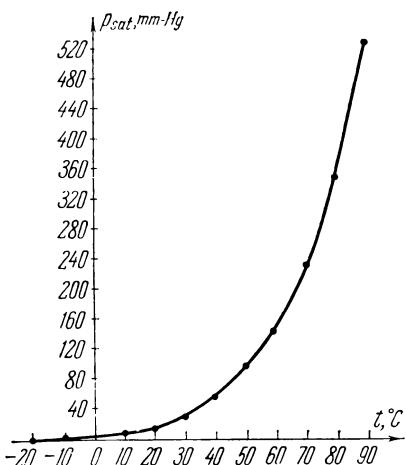


Fig. 35.3

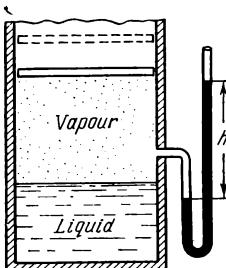


Fig. 35.4

Fig. 35.4. First the manometer will show a certain drop in pressure, which shows that concentration of the molecules of vapour has been reduced. But this will decrease the rate of condensation [see equation (35.5)] and the state of dynamic equilibrium will be violated. Since the rate of vaporization, depending only on the temperature, remains constant, some quantity of the liquid will be vaporized additionally and the state of dynamic equilibrium between the vapour and liquid will be restored. After this, the manometer will indicate that the pressure of the saturated vapour has returned to its initial value.

A similar process will be observed when we lower the piston. As the vapour is compressed, its concentration increases and the rate of condensation exceeds the rate of vaporization. The violated dynamic equilibrium is restored when a part of the vapour is additionally condensed. This restores the initial concentration of the molecules and the vapour pressure, which depend only on the temperature.

35.4. VAPOUR ISOTHERMS

1. In the device illustrated in Fig. 35.4 we shall continuously withdraw the piston. More and more of the liquid will vaporize and at a certain volume V_1 there will be no more liquid in the vessel. The pressure of the saturated vapour, however, will remain constant up to volume V_1 .

If we continue to withdraw the piston so that the volume of the vapour exceeds V_1 , the pressure will begin to decrease. Indeed, if the liquid has disappeared from the vessel, the reduction in concentration of molecules as the vapour expands is no longer compensated by vaporization of the liquid. Such vapour is said to be *unsaturated*, or *superheated*. At constant temperature its pressure drops with an increase in volume approximately according to the Boyle-Mariotte law.

2. Next we shall conduct the reverse process: lowering the piston we shall compress the superheated vapour. Its pressure will increase and at $V = V_1$ will equal the pressure of saturated vapour. If we continue to reduce the volume, a part of the vapour will condense and the pressure will remain equal to that of saturated vapour.

Finally, at a certain volume $V_2 < V_1$ all the vapour will be condensed and the piston will have advanced right up to the surface of the liquid. Any further reduction in volume will require extremely high pressures because the particles in a liquid are packed sufficiently close and its compressibility is low.

3. The process described above is shown graphically in Fig. 35.5. Portion 0-1 of the curve corresponds to superheated vapour, portion 1-2 to saturated vapour and portion 2-3 to the liquid. The whole curve is known as a *vapour isotherm*, or an Andrews curve, in honour of the Irish chemist Thomas Andrews (1813-1885) who first investigated isotherms of this kind in 1863.

It should be noted that the expansion and compression of the vapour should be sufficiently slow. The vaporization of a liquid is accompanied by cooling (see Sec. 35.1), and vapour condensation by heating of the liquid. A constant temperature is maintained by heat exchange of the device with the environment. And since heat exchange takes place quite slowly, the piston should be moved just as slowly.

4. If we compress the vapour at a higher temperature $T_1 > T$, we shall find that the state of saturated vapour is obtained at a new volume $V'_1 < V_1$ (Fig. 35.6). The explanation is quite simple: the pressure of the saturated vapour increases rapidly with the temperature (see Sec. 35.3), and the volume of the vapour must be substantially reduced for the pressure of the superheated vapour to reach that of the saturated vapour.

At a higher temperature the saturated vapour will be completely condensed and the piston will reach the surface of the liquid at

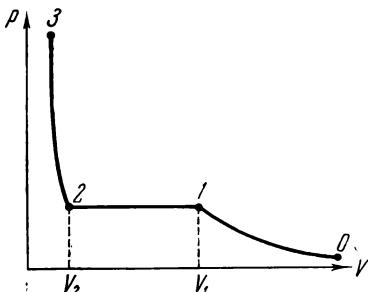


Fig. 35.5

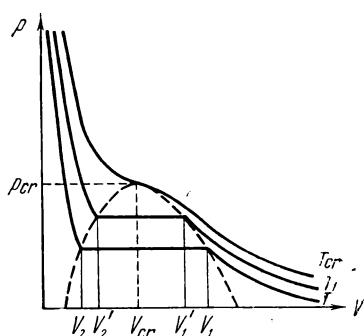


Fig. 35.6

a volume $V'_2 > V_2$. This is due to the thermal expansion of the liquid, which occupies a larger volume at a higher temperature.

35.5. THE CRITICAL STATE OF A SUBSTANCE

1. As we increase the temperature in the preceding experiments, the region of saturated vapour is reduced. Therefore, we can make the assumption that at a certain temperature T_{cr} no saturated vapour will be formed at all and that at a definite volume V_{cr} and pressure p_{cr} the superheated vapour will be continuously transformed directly into the liquid state (see Fig. 35.6). This assumption is confirmed by experiment.

The temperature T_{cr} corresponding to such a state of a substance is called its *critical temperature*; the pressure p_{cr} , its *critical pressure* and the density $\rho_{cr} = m/V_{cr}$, its *critical density*. For water, for example, $t_{cr} = 374^\circ\text{C}$, $p_{cr} = 218$ atm and $\rho_{cr} = 329 \text{ kg/m}^3$.

2. We can reach the concept of a critical state of substances from somewhat different considerations as well. It has been shown that the density of saturated vapour increases very rapidly with the temperature (see Sec. 35.2). Owing to thermal expansion the density of a liquid is reduced as the temperature is raised. If we plot curves of the dependence of the density of vapour and liquid on the temperature (Fig. 35.7), we shall find that they intersect at a certain point. This enables us to propose the following definition: the *critical temperature is the one at which the density of the saturated vapour equals that of the liquid*.

The critical state of a substance can be demonstrated by an experiment proposed by the Russian physicist M. P. Avenarius (1835-1895). A small sealed glass flask (Fig. 35.8) contains a certain amount

of ether. At a low temperature a sharply defined boundary is seen between the liquid and the saturated vapour. If the flask is heated, the liquid level rises even though a part of the liquid is vaporized when it is heated. This indicates that there is a substantial reduction in the density of the liquid and increase in the density of the vapour. At the critical temperature the boundary between the liquid and vapour disappears and the whole volume of the substance in the

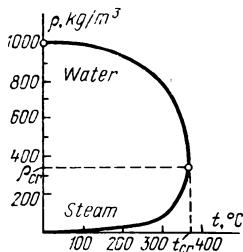


Fig. 35.7

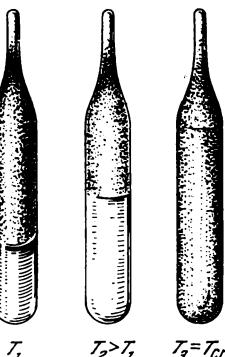


Fig. 35.8

flask becomes cloudy. This is caused by fluctuations in density throughout the volume at the critical temperature. Microscopic drops of the liquid continually appear and are immediately vaporized again. These drops scatter light and thereby impart the cloudiness to the contents of the flask.

3. At temperatures above the critical no liquid or, all the more, crystal can be formed even at the highest pressures. The reason is that the intensity of thermal motion of the molecules (or atoms) is so high that the molecular forces cannot maintain even short-range, let alone long-range, order though the molecules are relatively closely packed due to the high pressures.

This means that at above-critical temperatures a substance can exist only in its gaseous state, with completely unordered motion of its component particles, even at very high pressures.

4. At exceptionally high temperatures and the enormous pressures caused by the high gravitational forces in certain types of stars called *white dwarfs* matter is transformed into a special ultradense state. At this the electrons are torn away from their nuclei and form an electron gas similar to that found in metallic crystals. There is, however, an essential difference between these two states.

In metals the atoms lose only their outer-shell (valence) electrons, and the remaining ions form a crystal lattice. In the ultradense state atoms are stripped of *all* their electrons and the remaining

bare nuclei, of a size only about one ten thousandth of that of atoms, form an extremely close but entirely unordered structure. The density reached by such matter is enormous, of the order of 10^7 or 10^8 kg/m^3 . A glass of such matter (200 cm^3) would have a mass of about 20 tons (a glass of mercury has a mass of about 3 kg).

Nevertheless, this state of matter, called an ultradense plasma, is nearer in structure and properties to a gas than to a crystal or liquid. But the properties of this gas drastically differ from those of an ideal gas.

35.6. AIR HUMIDITY

1. The *absolute humidity* f of the air is evaluated on the basis of the mass of water vapour in 1 m^3 of air under the given conditions, i.e. the *density of the water vapour*. For convenience in calculations the density is expressed in grams per cubic metre rather than in SI units. Thus

$$f = \rho_{vap} (\text{g/m}^3) \quad (35.9)$$

In meteorology, the absolute humidity is usually evaluated, not by the density of water vapour, but by its pressure, expressed in millimetres of mercury column, i.e. $f = p$ (mm-Hg). We shall now show that the numerical values of these quantities differ only slightly at temperatures near to room temperature (see Table 35.1).

According to the equation of an ideal gas, $p = \rho RT/M$ [see equation (26.18)]. For water vapour $M = 18 \text{ kg/kmol}$. If the pressure is expressed in millimetres of mercury column ($1 \text{ mm-Hg} = 133.3 \text{ Pa}$, see Sec. 26.2) and the density in grams per cubic metre ($1 \text{ g/m}^3 = 10^{-3} \text{ kg/m}^3$), the state equation of the gaseous state can be written as

$$133.3 p = 0.463 \rho T$$

from which

$$p \text{ (mm-Hg)} = \rho T / 288 \text{ (g/m}^3\text{)}$$

But since room temperature $T \cong 300 \text{ K}$,

$$p \text{ (mm-Hg)} \cong \rho \text{ (g/m}^3\text{)}$$

2. The absolute humidity can be determined from the *dew point*. This is the temperature at which air becomes saturated by the water vapour in it. Thus, we can see from Table 35.1 that at an absolute humidity of the air equal to 5.60 g/m^3 , the air becomes saturated with water vapour at 2°C and moisture condenses, forming dew.

The dew point is determined by means of a *hygrometer* (from the Greek word *hygros* meaning wet or moist). A Lambert hygrometer is shown schematically in Fig. 35.9. Vessel A is filled with ether,

and a current of air is bubbled through it with a rubber bulb. As the ether is intensively evaporated, the temperature drops. When it reaches the dew point, condensation occurs on the walls of the vessel. Collar K is separated from the vessel by a washer of heat-insulating material and is, therefore, not cooled. The contrast between the shiny collar and the film of moisture on the vessel enables us to

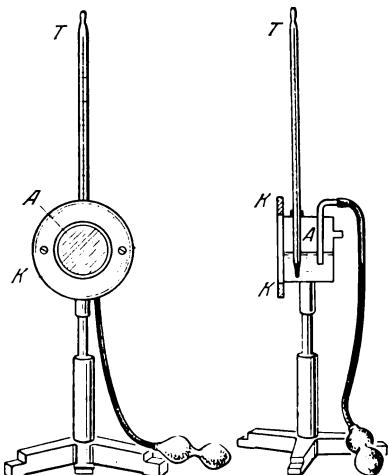


Fig. 35.9

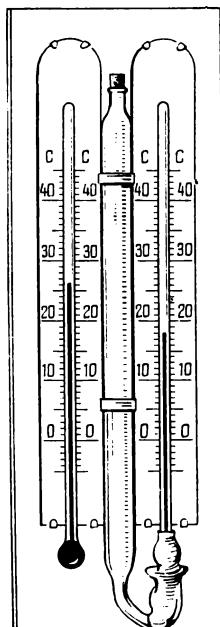


Fig. 35.10

sharply define the instant condensation begins. The dew point is determined by thermometer T , which has its bulb of mercury immersed in the ether.

3. The concept of the absolute humidity, introduced above, is insufficient because we are interested not only in the absolute moisture content of the air but in the degree that the humidity approaches the saturation point. Hence, we introduce the concept of the relative humidity B . *The relative humidity is the ratio of the absolute humidity to the density (or in meteorology, to the pressure) of saturated vapour at a given temperature.* Thus

$$B = \frac{f}{\rho_{sat}} \cong \frac{f}{p_{sat}} \quad (35.10)$$

4. If we measure the dew point with a hygrometer and know the air temperature, we can find the relative humidity from Table 35.1.

For example, if the dew point equals 8 °C and the air temperature is 19 °C, then the relative humidity is

$$B = \frac{8.3}{16.3} \left(\text{or } \frac{8.05}{16.48} \right) \cong 50\%$$

However, the relative humidity is usually determined by other instruments: hair, membrane or semiconductor hygrometers and also *psychrometers*. Wet-and-dry-bulb-thermometer psychrometers are instruments consisting of two thermometers, the bulb of one being wrapped in a strip of cambric with its end immersed in a vessel of distilled water (Fig. 35.10). The cambric acts as a wick that keeps the thermometer bulb wet. The dry-bulb thermometer indicates the air temperature; the wet-bulb one, the temperature of evaporating water. We have already mentioned that the temperature of a liquid drops as it is vaporized (see Sec. 35.1). Evidently, the drier the air, i.e. the lower its relative humidity, the more intensively water will evaporate from the wet cambric and the more it will be cooled. And the opposite: if the air contains much water vapour and its relative humidity is high, the water will evaporate slowly from the cambric and it will be cooled only slightly. Thus, the difference in the readings of the dry-bulb and wet-bulb thermometers (the so-called *psychrometric difference*) depends upon the relative humidity of the air. Special tables are available that list the air humidity according to the psychrometric difference.

Chapter 36

PHASE TRANSITIONS

36.1. CHANGES IN THE STATE OF AGGREGATION

1. We know from experience that solids melt when they are sufficiently heated, and then vaporize. We also know that we can observe the reverse process if we cool a substance. There are also substances whose crystals do not melt upon being heated but are directly vaporized (sublimated). It was found that not only the temperature but the pressure on the surface of the substance plays a vital role in these processes.

The state of aggregation of matter depends on its temperature and on the pressure on its surface.

Transitions of substances from one state of aggregation to another, accompanied by a change in the type of packing of the particles (long-range order, short-range order or complete disorder) are called *first-order phase transitions*.

2. Academician L. D. Landau (1908-1968) showed that there are *second-order phase transitions* as well. Certain properties of substances change in these transitions but not the type of packing of their particles, as is evident from X-ray structural analysis. Some second-order phase transitions will be described in the discussion of the magnetic properties of substances (see Ch. 42).

36.2. LIQUID-GAS TRANSITION DIAGRAM

1. Let us consider in more detail the graph of function (35.8), which expresses the dependence of the pressure of saturated vapour on the temperature (Fig. 36.1). Here a point *A*, lying on the curve, corresponds to a *two-phase state*: saturated vapour above a liquid and in a state of dynamic equilibrium with it. Let us find the states corresponding to points *B*, *C*, *D* and *E*, which are not on the curve.

2. The transition from state *A* to state *B* takes place by isothermally increasing the pressure. But we already know from Sec. 35.4 that when saturated vapour is compressed isothermally, it is transformed into the liquid state. Consequently, point *B* corresponds to a liquid compressed to a pressure above that of saturated vapour at the given temperature. We can readily see that point *C* corresponds to the liquid state with a temperature lower than that of saturated vapour at the given pressure.

Hence, points lying above the curve correspond to the liquid state of the substance.

3. The transition from state *A* to state *D* is accomplished by lowering the pressure isothermally. As shown in Sec. 35.4, the isothermal expansion of saturated vapour leads to additional vaporization of the liquid and the transformation of saturated vapour into unsaturated vapour, i.e. into the gaseous state. Point *D* corresponds to a gas whose pressure is lower than that of saturated vapour at the given temperature. It is likewise evident that point *E* corresponds to a gaseous state of the substance with a temperature higher than that of saturated vapour at the given pressure, i.e. to superheated vapour.

We see that points lying under the curve correspond to the gaseous state of the substance.

36.3. CRYSTAL-GAS TRANSITION DIAGRAM

1. The pressure of the saturated vapour above a crystal that is being sublimated can be expressed by the same function (35.8) as the vapour pressure above a liquid, if the vaporization energy u_0 is

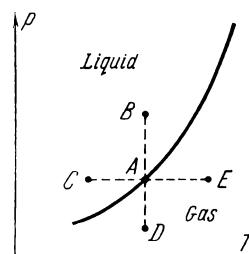


Fig. 36.1

replaced by the sublimation energy w_0 and constant A by a new constant B that characterizes the given substance. A plot of this function will yield a curve resembling the one in Fig. 36.1 but

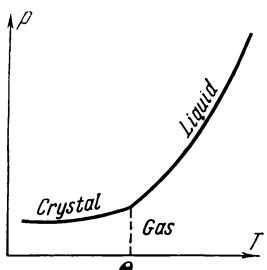
inclined somewhat differently. We find that at a certain point, corresponding to temperature Θ , the two curves intersect (Fig. 36.2).

2. If the temperature of the crystal is raised, it is transformed into a gas. If the pressure of the gas is increased, it is transformed into a crystal since the packing of the particles in crystals is much closer than in gases.

Thus, the crystal-gas phase transition diagram is quite similar to the liquid-gas diagram. Here too, points lying below and

to the right of the curve (less pressure and higher temperature) correspond to the gaseous state of the substance; points lying above and to the left of the curve (higher pressure and lower temperature), to the crystalline state.

Fig. 36.2



36.4. CRYSTAL-LIQUID TRANSITION DIAGRAM

1. The two cases discussed above—the transformations of a substance from the gaseous to the liquid and to the solid states, and vice versa—have much in common. In both cases the rise in temperature promotes the transition to the gaseous state because it leads to an increase in disordered molecular motion, which is typical of a gas.

In exactly the same way an increase in pressure promotes a transition to a more condensed state, liquid or crystalline, because the molecules are brought closer together and the resulting forces of attraction will help to obtain some kind of order in their arrangement: short-range at higher temperatures and long-range at lower temperatures. As a result the diagrams of these transitions are very much alike (Fig. 36.2). The curves separating the regions with different states of aggregation are inclined to the axis of abscissas at an acute angle.

2. The liquid-crystal transition diagram is a somewhat more complex matter.

Here again the rise in temperature promotes the transition from a more ordered to a less ordered state, i.e. from a crystal to a liquid. Therefore, on the diagram the region of the liquid state is to the right of the crystalline state. With regard to the pressure, there may be two different cases.

3. In most substances the particles are packed more closely in the crystalline state than in the liquid one. An increase in pressure

for such substances will consequently promote the transition from the looser liquid packing of the particles to the closer crystalline packing. This means that the points corresponding to the liquid state of the substance will be located to the right and below the phase transition curve (higher temperatures and less pressures), and the points corresponding to the crystalline state, to the left

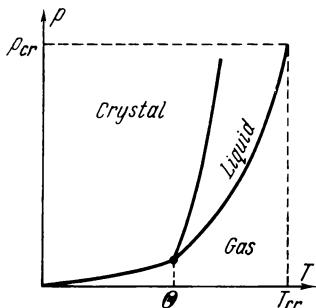


Fig. 36.3

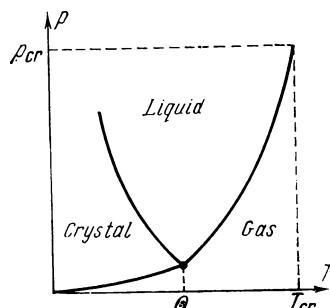


Fig. 36.4

and above the curve (lower temperatures and higher pressures). A diagram of the states and phase transitions of such a substance is illustrated in Fig. 36.3.

4. But substances are found in which the crystalline packing is looser than the liquid one. Here the density of the crystal is less than that of the liquid.

Common examples are water, as well as certain metals and alloys (bismuth and grey cast iron). An increase in pressure compacts their molecular packing and leads to the melting of these substances. Raising the temperature has the same effect. Here points corresponding to the liquid state are located to the right and above the phase transition curve (higher temperatures and pressures); points corresponding to the crystalline state, to the left and below the curve (lower temperatures and pressures).

A diagram of the states and phase transitions of such a substance is shown in Fig. 36.4. Here the melting curve is inclined to the axis of abscissas at an obtuse angle.

36.5. CRYSTAL-CRYSTAL TRANSITION DIAGRAM

1. Any substance has only one liquid and one gaseous state, but it may have several crystalline states. The reason lies in the different degrees of order of the molecular packing of the three states.

The gaseous state of a substance corresponds to 'complete molecular disorder. Such a distribution of molecules (or atoms) complies with an immense number of possible regroupings of the molecules

in space. However, the physical properties of the substance are the same for any of these regroupings. For this reason, they all correspond to a single gaseous state.

The liquid state is characterized by short-range order in the molecular packing (Ch. 34). But here again the spatial regrouping of the particles has no effect on the physical properties of the substance. That is why a substance has only one liquid state.

The crystalline state is an entirely different matter. Here different ways of packing the particles in space are quite feasible and the

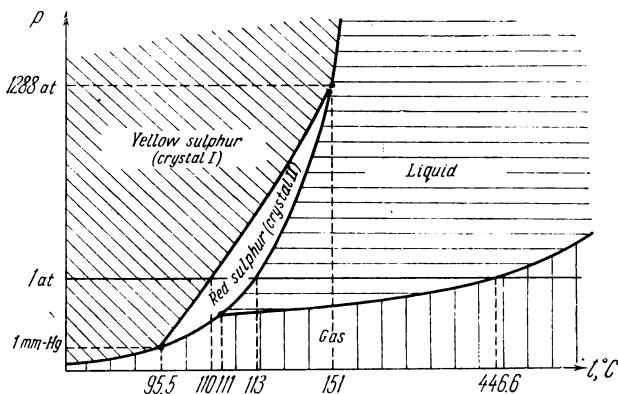


Fig. 36.5

substance will have different properties with each arrangement. Examples were discussed in detail in Chapter 33.

2. The order in which the particles in a crystal lattice are packed is determined by the conditions under which crystallization takes place, namely, the temperature of the substance and the external pressure. If these conditions are changed, the type of packing of the particle also changes.

Hence, processes of the recrystallization of substances, i.e. processes in which the crystal lattice is rearranged upon a change in the external conditions, are also examples of first-order phase transitions, and can be represented by phase diagrams.

3. Let us consider sulphur as an example. At a pressure of 1 mm-Hg and a temperature of 95.5 °C both yellow and red crystals are solidified out of sulphur vapour. At lower pressures and lower temperatures only yellow crystals are obtained; at higher temperatures and pressures, only red crystals.

A similar picture is observed in the freezing of a sulphur melt. At a temperature of 151 °C and pressure of 1288 at crystals of both kinds freeze out of the melt simultaneously; at pressures above 1288 at, only yellow crystals and at less pressures, only red ones.

The diagram of the states and phase transitions of sulphur is illustrated in Fig. 36.5. Note that the pressure and temperature scales on the coordinate axes are somewhat distorted to enable the large ranges to be shown.

36.6. THE TRIPLE POINT

1. All phase diagrams have characteristic points where the three phase transition curves intersect. In Figs. 36.3 and 36.4 these are points Θ , the junctions of the gas-liquid, gas-crystal and liquid-crystal transitions, i.e. the processes of boiling, sublimation and melting. In Fig. 36.5 such points are $\Theta_1 = 95.5^\circ\text{C}$, corresponding to the gas-crystal I, gas-crystal II and crystal I-crystal II transitions; $\Theta_2 = 151^\circ\text{C}$, corresponding to the crystal I-crystal II, crystal I-liquid and crystal II-liquid transitions; and $\Theta_3 = 111^\circ\text{C}$, corresponding to the gas-crystal II, gas-liquid and liquid-crystal II transitions. At these three temperatures and the corresponding definite pressures three phases can exist simultaneously, for instance, gas, liquid and crystal.

Any point on a phase diagram corresponding to conditions under which three phases (three different states of aggregation of the substance) can coexist in equilibrium is called a *triple point*.

2. Since a triple point is a quite definite constant for the given substance, it can serve as a standard (datum) point in establishing a temperature scale. In the International System (SI) of Units, the triple point of water has been adopted as the datum point for the absolute temperature scale. Namely, the triple point of water is assigned the temperature of exactly 273.16 K.

This leads us to the definition of the SI unit for measuring absolute temperature (see Sec. 26.7): the *kelvin* is $\frac{1}{237.16}$ of the thermodynamic temperature of the triple point of water.

36.7. CHANGES IN INTERNAL ENERGY IN FIRST-ORDER PHASE TRANSITIONS

1. First-order phase transitions are accompanied by a change in the internal energy of the system of particles that make up the body. Let us consider the processes of melting and vaporization from this standpoint.

In a crystalline packing the particles vibrate about equilibrium positions. The distances between these positions correspond to the minimum value of the interaction energy for the given temperature of the substance. When a crystal is melted, its space lattice is broken down and long-range order in the packing of the particles goes over to short-range order. As a rule the packing is loosened, the distance between the particles increases, and this leads to an increase

in the energy of interaction between them. However, if the liquid packing is more compact than the crystalline one (as for water), the distance between the molecules becomes less than that corresponding to minimum interaction energy, and this also leads to an increase in the internal energy.

Thus, the *melting (fusion) of a crystal is accompanied by an increase in the internal energy* due to the increased energy of interaction between the particles. Naturally, it is necessary, in this process, to add a certain amount of energy, the so-called *heat of melting, or fusion*.

2. With the aid of similar arguments, the student can find for himself that the *processes of sublimation and vaporization are also accompanied by an increase in the internal energy of the system*. For this purpose, the student should recall that, according to the conditions stated in Sec. 31.5, the interaction energy of gases is taken zero, and in a close packing it is negative. Consequently, in both sublimation of a crystal and in vaporization of a liquid a certain amount of energy must be added. It is called the *heat of phase transition* (*heat of sublimation and heat of vaporization*).

In phase transitions the energy of each molecule is changed. The greater the number of molecules making up the body, the more the energy required for the heat of phase transition. Denoting the change in energy of a single molecule in this process by ε_0 and the number of molecules by N , we have

$$Q = \varepsilon_0 N \quad (36.1)$$

Multiplying and dividing the right-hand side of the equation by m_0 , the mass of a molecule, we obtain

$$Q = \frac{\varepsilon_0}{m_0} m_0 N = \lambda m \quad (36.2)$$

where Q = total heat of phase transition

$m = m_0 N$ = mass of the substance

$\lambda = \varepsilon_0/m_0$ = heat of phase transition per unit mass.

3. The distance between particles changes only very slightly when a crystal is melted. In vaporization or sublimation the distance between the molecules becomes about ten times greater than in the liquid or solid states. Hence, the heat of vaporization should be considerably more than the heat of melting. Thus, the heat of melting for ice at standard pressure is $\lambda_{melt} = 80 \text{ kcal/kg} = 3.35 \times 10^5 \text{ J/kg}$, while the heat of vaporization at the same pressure is $\lambda_{vap} = 539 \text{ kcal/kg} = 2.26 \times 10^6 \text{ J/kg}$.

It follows from the law of conservation of energy that in the reverse processes (the condensation of vapour into the liquid or crystalline state, or the freezing of a melt) the internal energy is correspondingly reduced and the heat of phase transition is evolved. This conclusion can be experimentally confirmed.

A body can receive the heat of phase transition by heat exchange with its environment. In this case the phase transition proceeds at a constant temperature that depends upon the pressure the substance is subject to. The phase transition temperature can be found from the phase diagram of the substance. It is evident from Fig. 36.5, for instance, that at standard atmospheric pressure sulphur melts at 113°C and boils at 446.6°C .

4. The graphs given above show that the temperatures of boiling of a liquid and sublimation of a crystal increase with pressure because it always follows from $p_2 > p_1$ that $T_2 > T_1$ as well. As to the melting point, it increases with the pressure for bodies in which crystalline packing is more compact than liquid packing (see Fig. 36.3).

Bodies like ice, in which the crystalline packing is looser than the liquid one, behave differently. Here the melting point drops with an increase in pressure. Indeed, here it follows from $p_2 > p_1$ that $T_2 < T_1$ (see Fig. 36.4).

5. If insufficient energy is delivered in heat exchange, the heat of phase transition can be obtained only from a change in the energy of molecular motion. At this the temperature of the body is changed.

Such a process makes it possible, for instance, to store liquefied gases in a Dewar flask. Here heat exchange with the environment has been reduced to a minimum (Sec. 21.5). The liquid intensively evaporates and is thereby cooled. This is why a Dewar flask containing liquefied gas must not be corked. Sealing stops vaporization and then the liquid is soon heated to the temperature of the environment. Since this temperature is higher than the critical one, the liquid will be transformed at once to the gaseous state, causing an explosion.

Owing to intensive sublimation, solid carbon dioxide, dry ice, can be kept for some time at atmospheric pressure. Its temperature will be 75°C below zero even though the temperature of its environment is considerably higher.

36.8. METASTABLE STATES

1. It should not be supposed that a body immediately goes over to a new state of aggregation, in accordance with the phase diagram, as soon as the temperature or pressure is changed. Only two phase transitions are accomplished without delay when the temperature or pressure changes: the sublimation and melting of a crystal. All other phase transitions are often delayed. Though the temperature and pressure of a body correspond to the new state of aggregation, the phase transition may fail to take place.

If a body is in a state of aggregation that does not agree with its temperature and pressure according to the phase diagram, it is said to be in a *metastable state*.

One example of a metastable state was discussed in Sec. 34.6. We refer to the amorphous state of a substance which, in this case, is a *supercooled liquid*. A substance may also be a *superheated liquid* or a *supercooled vapour*, and a crystal lattice of a certain type may exist when one of a different type is more stable under the given conditions.

2. Let us consider the last process in more detail. Any rearrangement of the crystal lattice to comply with changes in the external conditions takes place in the presence of strong bonds between the particles which are closely packed in the space lattice. The molecules are in "cramped" conditions; it is very difficult for them to become rearranged into the new crystal lattice that is better suited to the changed conditions. Hence, the transition from one crystalline state to another is usually retarded. If, for example, we heat yellow sulphur to 110 °C at standard pressure, it should be transformed, according to the phase diagram (see Fig. 36.5), into the red type. Actually, this does not occur, and further heating to 113 °C will cause the sulphur to melt. When the molten sulphur is cooled below 113 °C, red crystals are formed. But upon further cooling, below 110 °C and down to room temperature, they are not transformed immediately into the yellow variety. It will take considerable time, of the order of a 24-hour day, for the red crystals to change into a yellow powder.

3. The process of recrystallization can be markedly accelerated if a small yellow crystal is brought into contact with the red crystals that are in the metastable state. It is found that it is much easier for the particles to fit onto the ready faces of the crystal lattice that is stable under the new conditions.

In the formation of a new crystal lattice from a metastable crystal an essential role is played by fluctuations. They form the microscopic seed of stable crystalline packing around which the growth of the new crystal begins.

Since such microscopic seeds may appear simultaneously at various places and are differently oriented, the formation of a polycrystalline structure is much more probable than of a monocrystal.

4. The availability of seeds, which serve as centres of crystallization, facilitates the transition of a supercooled liquid to the crystalline state. Molten hyposulfite, for instance, can be carefully cooled to room temperature, and it can remain in this supercooled state for days. But if a single small crystal is dropped into the melt, the liquid will almost instantly crystallize. We should note that in this process the temperature of the hyposulfate is raised because it evolves the heat of melting.

Superheated liquids and supercooled vapours will be dealt with in the following sections.

36.9. CONDENSATION. SUPERSATURATED VAPOUR

1. If the temperature of vapour is lowered at constant pressure below the dew point, the two-phase mixture, saturated vapour,

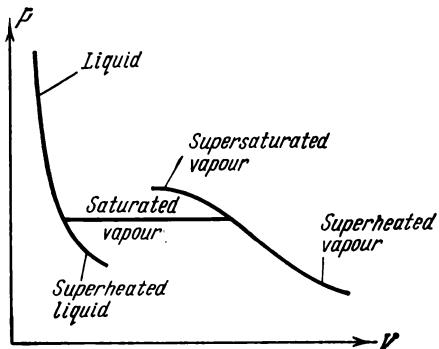


Fig. 36.6

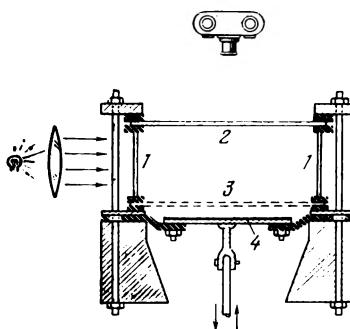


Fig. 36.7

should be formed. It was found, however, that drops of liquid can be formed only if the gas contains *condensation centres*, or *nuclei*.

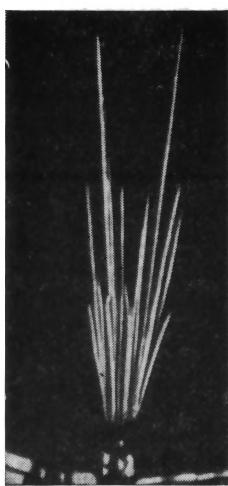


Fig. 36.8

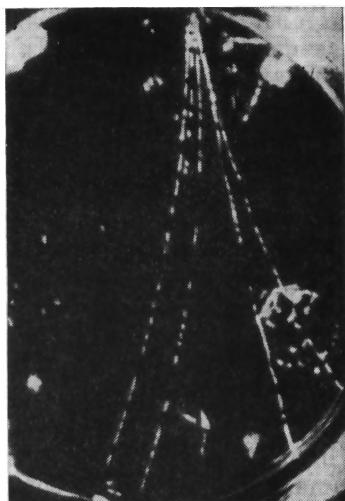


Fig. 36.9

These may be dust specks, ions, sharp points on the surface of solids, etc. If no condensation centres are available, no drops of liquid

are formed and we have a metastable state: *supersaturated (supercooled) vapour*.

Supersaturated vapour can also be obtained by isothermally compressing vapour to a pressure that exceeds that of saturated vapour at the given temperature. A graph showing the dependence of the pressure of supersaturated vapour on its volume is shown in Fig. 36.6.

2. The capability of ions to serve as condensation centres is applied in the *Wilson cloud chamber* (Fig. 36.7), invented in 1911 by C.T.R. Wilson (1869-1959). This valuable research instrument consists of cylindrical glass vessel 1. with glass cover 2 on top. Below, the vessel is enclosed by a layer of moist black velvet or broadcloth (on screen 3) on the surface of which the saturated vapour is formed. As diaphragm 4 is pulled quickly downward, the gas adiabatically expands and its temperature drops. Due to this cooling action the vapour becomes supercooled (supersaturated).

Charged particles, produced in radioactive decay, such as alpha or beta particles, "shoot" through the gas and form a chain of ions along their paths. Then droplets of liquid are formed on these ions which act as condensation centres. Thus, the flying particles leave traces (tracks), which are quite visible and can be photographed. Shown in Fig. 36.8 is a photograph of the tracks made by alpha particles (short heavy lines). The tracks of beta particles (long intermittent lines) are illustrated in Fig. 36.9.

36.10. BOILING. SUPERHEATED LIQUID

1. The evaporation of a liquid proceeds from its free surface at any temperature. The higher the temperature, the greater the rate of evaporation.

Boiling is the process of vigorous vapour formation, not only at the surface of a liquid but throughout its volume. It is necessary for this purpose to heat the liquid to a high temperature. An important part in boiling is played by the bubbles of gas in the liquid. Such bubbles are formed at the boundary between the liquid and a solid body. These bubbles are filled with saturated vapour of the liquid.

As the temperature of the liquid rises, the pressure of the vapour in a bubble increases as does its volume. The buoyant force, acting on the bubble according to Archimedes' principle, increases with its volume. Under definite conditions this force becomes greater than the cohesive forces between the bubbles and the solid walls. Then the bubble breaks away from the wall, rises to the surface and bursts, releasing the vapour it contains.

2. The pressure p_{sat} of the saturated vapour inside a bubble at the surface of the liquid is equal to the sum of the external pressure

p_{ext} on the liquid and the pressure under the curved surface:

$$p_{sat} = p_{ext} + \frac{2\sigma}{r} \quad (36.3)$$

where r = radius of the bubble

σ = surface tension (Sec. 34.8).

If the bubble is several millimetres in size or larger, the second term on the right-hand side of equation (36.3) can be neglected. For example, the surface tension for water at a temperature of 100 °C is $\sigma = 58.8 \times 10^{-3}$ N/m, and if $r = 1$ mm, then

$$\frac{2\sigma}{r} = \frac{2 \times 58.8 \times 10^{-3}}{10^{-3}} = 118 \text{ Pa}$$

But standard atmospheric pressure equals 1.01×10^5 Pa, which is about a thousand times greater. Hence, for sufficiently large bubbles, the condition for the bubble to rise to the surface and to burst can be written as

$$p_{sat} \geq p_{ext} \quad (36.4)$$

This is the *condition for boiling*.

In this way a liquid can boil if it contains bubbles of sufficiently large size (of the order of a millimetre for water). Boiling begins when the pressure of the saturated vapour reaches that of the external pressure on the surface of the liquid.

We have previously established that the pressure of saturated vapour is determined only by the temperature of the liquid. It follows from condition (36.4) that the boiling point depends on the external pressure; it increases with the external pressure. The boiling point of water at various pressures can be found in Table 35.1 (p. 371).

3. If the liquid contains no bubbles to promote the boiling process, it can be superheated without beginning to boil. In the corresponding metastable state the liquid is said to be *superheated*.

A superheated liquid can also be obtained by reducing the external pressure above the liquid to a value less than that of saturated vapour at the given temperature. A graph illustrating the dependence of the pressure of a superheated liquid on its volume is given in Fig. 36.6.

4. A superheated liquid is used in *bubble chambers*, invented in 1952 by D. A. Glaser (b. 1926) for making the tracks of relativistic particles visible. Relativistic particles are ones travelling at velocities approaching that of light. Such particles produce very few ions in gases, and their tracks are not visible in a Wilson cloud chamber. But they produce considerably more ions in liquids, where the molecules are closely packed, and their tracks are visible and can be registered.

The principle of a bubble chamber is similar to that of the cloud chamber. A relativistic particle, passing through a heated liquid, produces a chain of ions along its path. If the pressure above the

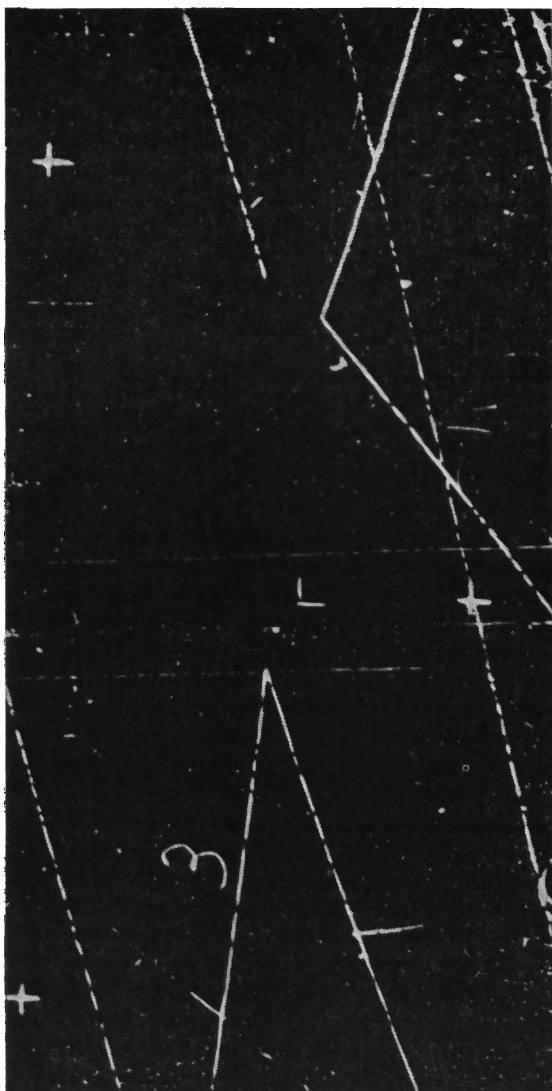


Fig. 36.10

liquid is sharply reduced, the liquid becomes superheated. The ions serve as centres for boiling, and the resulting bubbles form a track along the path of travel of the particle. This track is photographed (Fig. 36.10). Liquefied hydrogen, propane or xenon is used as the working fluid.

36.11. THE LIQUEFACTION OF GASES

1. For a gas to be transformed into a liquid its temperature must be lowered below the critical value. Such gases as chlorine ($t_{cr} = -144^\circ\text{C}$), ammonia ($t_{cr} = 132.4^\circ\text{C}$), carbon dioxide ($t_{cr} = 31.1^\circ\text{C}$), xenon ($t_{cr} = 16.7^\circ\text{C}$) and certain others, having a critical temperature of the order of room temperature, can be liquefied without much difficulty. Their temperature is lowered below the critical value, then they are compressed isothermally and the liquid is obtained. It is much harder to produce liquid oxygen ($t_{cr} = -118.8^\circ\text{C}$), nitrogen ($t_{cr} = -147.1^\circ\text{C}$), hydrogen ($t_{cr} = -239.9^\circ\text{C}$) and especially liquid helium ($t_{cr} = -267.9^\circ\text{C}$). Special equipment, required for this purpose, enables very low temperatures to be obtained, in some cases, as for helium, close to absolute zero.

Many methods are known for liquefying gases. We shall confine ourselves to only one of these: cooling gases by having them perform work by adiabatic expansion. Refrigerating machines operating by this principle are called *expansion engines*, or gas-expansion machines. Academician P. L. Kapitza (b. 1894) developed a highly efficient expansion turbine. This is a miniature turbine driven by precompressed gas.

2. Let us consider the principle of an installation for liquefying gases. It consists of a compressor, heat exchanger and expansion engine (Fig. 36.11).

The compressor is a piston-type machine for compressing the gas. Surplus heat, evolved when the gas is compressed, is removed by a water-type cooling system. The compressed gas is passed through the inner tube of the heat exchanger, which has an outer tube of larger diameter through which cooled gas flows in the opposite direction along the outside of the inner tube. After leaving the heat exchanger, the compressed gas passes through a Laval nozzle on its way to the vanes of the turbine rotor, which it drives. In this adiabatic expansion the gas does work at the expense of its internal energy. In this process the temperature of the gas drops so sharply that a part is liquefied and flows into a Dewar flask. The rest of the cooled, but not yet liquefied, gas is returned through the heat exchanger, where it cools a new portion of compressed gas, and back to the compressor where it is compressed again.

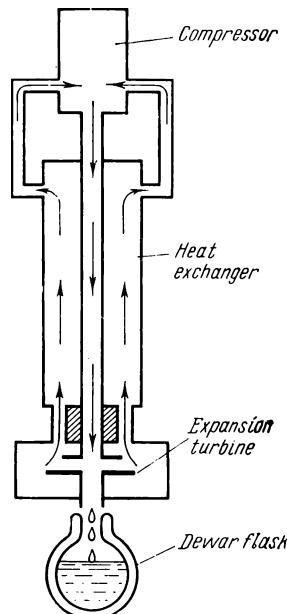


Fig. 36.11

3. The development of methods of liquefying gases and thereby reaching low temperatures has led to the founding of *cryogenics*. Liquefied gases are extensively employed today, both in physical research when low temperatures are required and in engineering. Thus, for instance, in the evaporation of liquid air, nitrogen evaporates first and pure oxygen remains. At the present time this is the cheapest method of obtaining pure oxygen, which is widely used in medicine, for gas welding, to speed up the blast furnace process and for smelting steel. When liquid oxygen is mixed with certain organic combustible substances (cotton wool, cotton or sawdust) explosives are obtained (oxyliquits), which find application in the mining industry. Their advantageous feature is that if no explosion occurs, the oxygen evaporates after a certain time and the blasting cartridge becomes entirely harmless.

Chapter 37

A FIELD OF FIXED CHARGES
IN A VAC U M

37.1. LINES OF FORCE

1. In Chapters 10 and 18 we dealt with the interaction of point electric charges and of dipoles in a vacuum. For convenience, let us repeat the main relationships and recall their meaning.

A field of fixed charges, an *electrostatic field*, is described by two physical quantities: the field strength E and the potential φ . According to definition, the field strength vector

$$\mathbf{E} = \mathbf{F}/q \quad (37.1)$$

where q = test charge

\mathbf{F} = force acting on the test charge at the given point in the field (see Sec. 10.5). The potential is

$$\varphi = U/q \quad (37.2)$$

where U is the potential energy of a test charge at the given point in the field (see Sec. 18.7).

2. If the field is set up by a point charge Q , then the field strength is

$$E = \frac{Q}{4\pi\epsilon_0 r^2} \quad (37.3)$$

and the potential (see Sec. 18.8) is

$$\varphi = \frac{Q}{4\pi\epsilon_0 r} \quad (37.4)$$

3. A convenient way of visualizing electric fields is by means of lines of force, a method proposed by Michael Faraday (1791-1867).

A *line of force* is a straight or curved line whose tangent at each point coincides with the direction of the field strength vector. A line

of force is attributed a definite direction, coinciding with that of the field strength vector.

If the source of the field is a positive point charge, the field strength vector has the same direction as the radius vector; if it is a negative charge, the field strength vector has the opposite direction (see Fig. 10.7). Hence, lines of force begin at positive charges and end at negative charges.

4. Lines of force should not be credited with more meaning than that of being an illustrative representation of a field. We must not assume that the field "consists" of lines of force (as a brush consists of bristles). In fact, we could draw a line of force through any point in a field, so that there would be an infinite number of them.

Neither should we identify lines of force with the paths of charges in a field. The point is that it is the velocity vector that is directed along a tangent to the path of a charge, and the direction of the velocity need not coincide with the direction of the force, and all the more with the direction of the field strength vector. It is highly objectionable, of course, to try to explain electrical phenomena as the "tension", "pressure" and "interaction" of lines of force as sometimes practised in earlier physics textbooks.

37.2. EQUIPOTENTIAL SURFACES

1. There is another way of depicting electrostatic fields: the method of equipotential surfaces. An *equipotential surface* is the locus of points, all of which have the same electrical potential.

If the source of the field is a point charge, then we see from equation (37.4) that the equipotential surfaces will comprise a family of concentric spheres with their centre at the point where the charge is located. A plane cross section of such a system of equipotential surfaces is illustrated in Fig. 37.1. The dotted lines are the lines of force.

As we see in the illustration, the direction of the line of force at each point in the field is normal to the equipotential surface passing through

this point. We shall show that this is a general property of lines of force, typical of fields set up by an arbitrary charge or system of charges.

2. For this proof we shall choose two nearby points on an equipotential surface, with the distance Δl between them. Let us cal-

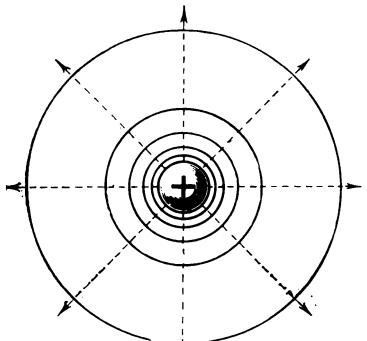


Fig. 37.1

culate the work done in moving charge q from one point to the other.

As we know (see Sec. 18.7), the work done in moving a charge between two points equals the product of the charge by the potential difference:

$$W = q (\varphi_1 - \varphi_2) \quad (37.5)$$

This work equals zero because the points lie on a single equipotential surface and $\varphi_1 = \varphi_2$. But this work can also be expressed (see Sec. 18.1) as

$$W = F \Delta l \cos \alpha = qE \Delta l \cos \alpha \quad (37.6)$$

Equating these two expressions for the work, we have

$$qE \Delta l \cos \alpha = 0$$

But neither the charge, nor the field strength, nor the distance between the points equals zero. Therefore, $\cos \alpha = 0$, and $\alpha = \pi/2$. Since these are two arbitrary points lying on an equipotential surface, it follows that the field strength vector is normal to this surface.

3. This conclusion is very significant. On the basis of this property we can draw a system of lines of force from a known family of equipotential surfaces and vice versa. We can, for example, draw a picture of a *uniform field*. A field is said to be uniform if the field strength vector at all points of the field has the same magnitude and direction, i.e. $E = \text{const}$. Then the lines of force have the same direction at all points of this field. And this is true only if they comprise a system of parallel straight lines. Hence, the equipotential surfaces will correspondingly be a system of parallel planes perpendicular to the lines of force (Fig. 37.2).

4. The lines of force illustrate how the field changes from point to point.

In the field of a point charge, for

instance, the lines of force are crowded closest together near the charge, i.e. where the magnitude of the field strength vector is large (Fig. 37.1). In a uniform field, on the contrary, where the field strength is the same at all points, the lines of force are equally spaced throughout the field (Fig. 37.2).

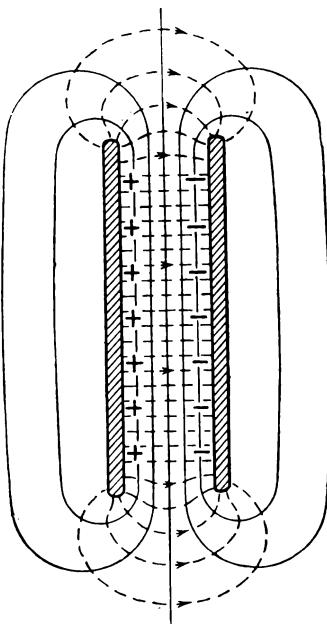


Fig. 37.2

37.3. THE RELATION BETWEEN FIELD STRENGTH AND POTENTIAL

1. Suppose that the field strength vector \mathbf{E} is constructed at a certain point of a field, and the equipotential surface M_1N_1 is drawn. All points of this surface have the same potential φ_1 (Fig. 37.3). At a short distance Δn from the first surface we draw a new equipotential surface M_2N_2 , all points on which have a potential φ_2 .

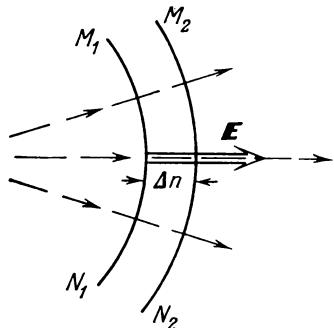


Fig. 37.3

The work done in moving a test charge q over the distance Δn can be calculated in two ways, as in the preceding section. We can write either

$$W = q(\varphi_1 - \varphi_2) \quad \text{or} \quad W = qE \Delta n$$

from which

$$E = \frac{\varphi_1 - \varphi_2}{\Delta n} = -\frac{\Delta\varphi}{\Delta n} \quad (37.7)$$

where $\Delta\varphi = \varphi_2 - \varphi_1$ is the change in potential in moving the test charge q over the distance Δn along a normal to the equipotential surface, i.e. along a line of force.

Equation (37.7) suggests the name of the SI unit of electric field strength—volts per metre (V/m).

2. The result obtained above is not quite accurate since the formula $\Delta W = F \Delta l \cos \alpha = F \Delta n$ can be used only to calculate an element of work, i.e. the work done by a force over an infinitesimal displacement. We obtain the precise expression when equation (37.7) is written at the limit under the condition that Δn tends to zero. Thus

$$E = -\lim_{\Delta n \rightarrow 0} \frac{\Delta\varphi}{\Delta n} \quad (37.8)$$

Therefore, the *electric field strength is characterized by the change in potential per unit length along the line of force*, i.e. the field strength vector always has the direction of the decrease in potential.

3. The strength of a uniform field is always constant. The force acting on a charge in such a field is, therefore, also constant, which allows us to employ formula (37.7) for any sufficiently long lengths of lines of force. Thus, in a uniform field

$$E = \frac{\varphi_1 - \varphi_2}{d} \quad (37.9)$$

where d is the distance between the two points.

37.4. A DIPOLE IN AN ELECTRIC FIELD

1. Let a dipole (Sec. 10.4) be located in a uniform electric field so that the dipole moment vector makes the angle α with the direction of the field strength vector (Fig. 37.4). In this case, the dipole is

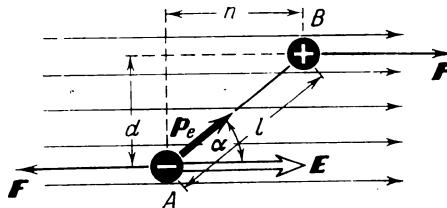


Fig. 37.4

subject to the action of a force couple. The moment of this couple equals the product of the force by its moment arm (Sec. 22.4): $M = Fd = qEl \sin \alpha$. But since $p_e = ql$, we have, finally,

$$M = p_e E \sin \alpha \quad (37.10)$$

Thus, the torque acting on a dipole in an electric field is equal to the product of the electric dipole moment by the field strength and by the sine of the angle between them.

The torque of the force turns the dipole to the position in which its electric moment becomes parallel to the field strength vector. In this position both the torque and the resultant electric force equal zero, and the dipole will, therefore, be in a state of stable equilibrium. This can be proved by calculating the energy of the dipole.

2. The energy of a dipole in an electric field is the sum of the energies of the two charges of which it consists. Thus

$$\mathcal{E} = U_A + U_B = q\varphi_B - q\varphi_A \quad (37.11)$$

After certain transformations we can write

$$\mathcal{E} = -ql \frac{\varphi_A - \varphi_B}{n} \times \frac{n}{l}$$

where $ql = p_e$ = dipole moment

$(\varphi_A - \varphi_B)/n = E$ = field strength

$n/l = \cos \alpha$ (see Fig. 37.4).

Then the energy of a dipole in an electric field is

$$\mathcal{E} = -p_e E \cos \alpha \quad (37.12)$$

If the angle $\alpha = 0$, the energy of the dipole is minimal: $\mathcal{E}_{\min} = -p_e E$, and the torque equals zero. This means that the dipole is in a state of stable equilibrium.

At $\alpha = \pi$, the energy of the dipole will be maximal: $E_{\max} = p_e E$ and the torque equals zero. Evidently, in this position the dipole is in a state of unstable equilibrium (Secs. 19.6 and 22.5).

3. This orienting effect of an electric field on a dipole is utilized to obtain illustrative pictures of electric fields. Models of electrodes are cut out of metal foil and glued on the bottom of a glass vessel. A viscous nonconducting fluid (turpentine, glycerine or castor oil) is poured into the vessel, and then small particles, for example, hair clippings, semolina (farina), grass seed, etc. are poured into

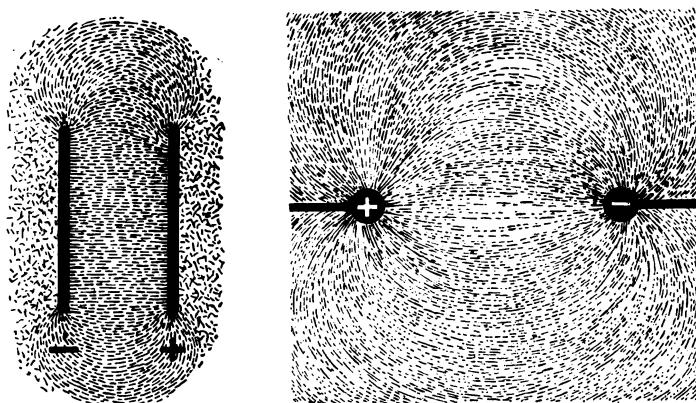


Fig. 37.5

the liquid. A large potential difference, of the order of several thousand volts, is set up between the electrodes. In the electric field the particles are polarized, i.e. they acquire the properties of miniature dipoles, extending along the lines of force of the field (Fig. 37.5). We see in these illustrations (drawn from photographs) that the field is quite nonuniform near charged spheres, but that a practically uniform field, distorted only at the edges, is set up between parallel plates.

4. Besides the torque, a dipole is also subject to the action of the resultant force when it is in a nonuniform field. The reason for this is that the positive and negative charges of the dipole are located at different points in the field, having different field strengths (Fig. 37.6). The resultant force is

$$F = F_A - F_B = qE_A - qE_B = ql \frac{E_A - E_B}{l} \quad (37.13)$$

But $ql = p_e$ is the dipole moment, and $(E_A - E_B)/l = \Delta E/\Delta l$ is the *field strength gradient*, i.e. the quantity showing the change

in field strength per unit length. Thus

$$F = p_e \frac{\Delta E}{\Delta l} \quad (37.14)$$

This force pulls the dipole into the region with a stronger field.

For the same reason a charged body attracts light objects: pieces of paper, specks of dust, pieces of foil, etc. By the action of the

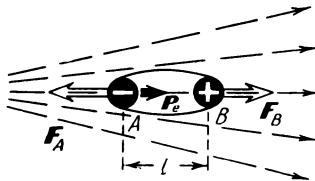


Fig. 37.6

field these objects first acquire a dipole moment and are then pulled to a region where the field strength is greater, i.e. closer to the electrified body.

37.5. THE PARALLEL-PLATE CAPACITOR

1. A parallel-plate capacitor is a system of two flat metal plates arranged parallel to each other. Each one has the area A . The distance d between the plates is much less than their linear dimensions. In this case, the plates can be assumed to be infinitely large and distortion of the field at their edges can be neglected.

Let the charge on one plate equal q , while that of the other plate is equal in magnitude but of opposite sign. The *surface density* σ of the charge is the ratio of the charge to the area of a plate:

$$\sigma = \frac{q}{A} \quad (37.15)$$

The SI unit of surface density is

$$[\sigma] = \text{C/m}^2 \text{ (coulomb per metre squared)}$$

2. Since the field between the plates of such a capacitor is uniform (Fig. 37.2) its field strength is the same at all points. Experiments show that the field strength is proportional to the surface density of the charge. Thus

$$\sigma = k \epsilon_0 E \quad (37.16)$$

where k is a dimensionless factor. It is shown in more rigorous courses that $k = 1$ in the International System of Units (SI). Therefore, the relation between the surface density of the charge and the field

strength of a parallel-plate capacitor is

$$\sigma = \epsilon_0 E \quad (37.17)$$

3. Let us find the expression for the potential difference between the capacitor plates. According to equation (37.9), we have

$$\varphi_1 - \varphi_2 = Ed \quad (37.18)$$

Substituting the values of E and σ , we obtain

$$\varphi_1 - \varphi_2 = \frac{\sigma d}{\epsilon_0} = \frac{qd}{\epsilon_0 A} \quad (37.19)$$

37.6. CAPACITANCE

1. From the example of the parallel-plate capacitor we see that the potential difference between its plates is proportional to their charge [equation (37.19)]. It was found that a similar result is obtained for a system of charged conductors of arbitrary shapes. This has led to the introduction of a quantity called the *capacitance* of a system. The capacitance of a system of conductors is a quantity equal in magnitude to the ratio of the charge to the potential difference. Thus

$$C = \frac{q}{\varphi_1 - \varphi_2} \quad (37.20)$$

According to equations (37.19) and (37.20), the capacitance of a parallel-plate capacitor is

$$C = \frac{\epsilon_0 A}{d} \quad (37.21)$$

2. The SI unit of capacitance is the *farad*, i.e. the capacitance of a conductor in which a change in charge of one coulomb produces a change in potential of one volt. Thus

$$1 \text{ F} = \frac{1 \text{ C}}{1 \text{ V}}$$

Capacitors employed in radio engineering have capacitances considerably less than a farad. They are therefore specified by submultiples of the farad: the *microfarad* ($1 \mu\text{F} = 10^{-6} \text{ F}$) and the *picofarad* ($1 \text{ pF} = 10^{-12} \text{ F}$).

3. Equation (37.21) suggests the name (or dimensionality) of the permittivity constant: the *farad per metre*. Thus

$$[\epsilon_0] = [Cd/A] = \text{F} \times \text{m/m}^2 = \text{F/m}$$

37.7. THE ENERGY OF THE ELECTRIC FIELD. ENERGY DENSITY

1. Let us calculate the energy of a capacitor with a capacitance of C and a potential difference between the plates $\Delta\varphi = \varphi_1 - \varphi_2$. The charges on its plates are equal in magnitude but of opposite sign: $q_1 = -q$ and $q_2 = q$. Substituting these values into equation (18.30), we obtain

$$U = \frac{1}{2} (q_2\varphi_1 + q_1\varphi_2) = \frac{q}{2} (\varphi_1 - \varphi_2) = \frac{q\Delta\varphi}{2} \quad (37.22)$$

But, since $C = q/\Delta\varphi$, we have

$$U = \frac{q\Delta\varphi}{2} = \frac{C\Delta\varphi^2}{2} = \frac{q^2}{2C} \quad (37.23)$$

2. Within the scope of electrostatics it is impossible to determine where the energy is concentrated: only on the charged bodies or at all points throughout the field. Variable fields, however, can exist independently of charges. It was found that variable fields propagate in the form of electromagnetic waves that carry energy. Thus, the energy is stored in the field itself.

This idea was first advanced by Maxwell. He also introduced the concept of the *energy density* $u = U/V$, which is a quantity equal to the ratio of the energy stored in a certain volume V to this volume. According to Maxwell, the energy density of a field (in a vacuum) is

$$u_0 = \frac{\epsilon_0 E^2}{2} \quad (37.24)$$

37.8. THE FORCE OF INTERACTION BETWEEN CAPACITOR PLATES

1. Making use of the equation for the energy of a charged capacitor, we can calculate the force of interaction between its plates. We shall assume that the capacitor is first charged and then disconnected from the source of energy, so that the charge on its plates remains constant. The energy of the capacitor will be expressed in terms of its capacitance and charge. Thus

$$U_1 = \frac{q^2}{2C_1} = \frac{q^2d_1}{2\epsilon_0 A}$$

Now we shall imagine that we have displaced one of the plates by the distance $\Delta l = d_2 - d_1$, thereby changing the energy of the capacitor to

$$U_2 = \frac{q^2d_2}{2\epsilon_0 A}$$

It follows from the definition of energy that the element of work equals the decrease in energy:

$$\Delta W = U_1 - U_2$$

which we can write as

$$F\Delta l = \frac{q^2d_1}{2\varepsilon_0 A} - \frac{q^2d_2}{2\varepsilon_0 A} = -\frac{q^2\Delta l}{2\varepsilon_0 A}$$

Then the force of interaction between the capacitor plates is

$$F = -\frac{q^2}{2\varepsilon_0 A} \quad (37.25)$$

The minus sign shows that it is a force of attraction (see Secs. 10.2 and 31.4).

2. Let us rewrite equation (37.25), recalling that $q = \sigma A$, where σ is the surface density of the charge. Then

$$F = -\frac{\sigma^2 A}{2\varepsilon_0}$$

Substituting equation (37.17) for σ , we have

$$F = -\frac{\varepsilon_0 E^2 A}{2} \quad (37.26)$$

3. Since the force per unit area is the pressure $p = F/A$, it follows from equation (37.26) that the negative pressure exerted by the electric field (in a vacuum) on a capacitor plate is

$$p_0 = -\frac{\varepsilon_0 E^2}{2} = -u_0 \quad (37.27)$$

Hence, the *pressure* exerted by an electric field on a charged surface is equal to the *energy density* of the electric field, taken with the opposite sign, at the point where an element of the given surface is located. We obtained this result for the field of a parallel-plate capacitor, but Maxwell proved that this result is valid for any non-uniform field as well.

Note that if the capacitor plates had like charges, the signs of the force F and the pressure p would be positive, corresponding to forces of repulsion.

37.9. A CONDUCTOR IN AN ELECTRIC FIELD

1. On the basis of their conductivity all bodies can be divided into two classes: *conductors* and *dielectrics* (also called *insulators*). In conductors electric charges are free to move throughout the volume of such bodies. In dielectrics, on the contrary, the charges remain fixed at the places where they were initially distributed.

Inside a conductor the field strength equals zero. If this were not so, an ordered motion of the charges would occur in the conductor without any expense of energy from an external source. But this contradicts the law of conservation of energy. Hence, the field strength within a conductor $E_{\text{inside}} = 0$.

It follows that all points of a conductor have the same potential and the *surface of a conductor is an equipotential surface*. Indeed, from

$$E_{\text{inside}} = \frac{\varphi_1 - \varphi_2}{d} = 0$$

it follows that $\varphi_1 = \varphi_2$.

It was shown above that lines of force are perpendicular to an equipotential surface. Therefore, near a conductor the lines of force are perpendicular to its surface.

2. If there is a cavity in a conductor or it is hollow, the field strength inside the cavity equals zero whatever the field outside the conductor.

The phenomenon known as *electrostatic shielding* is based on this principle: if an instrument is put inside a closed metallic hood, then changes in the external electric field will have no effect on the instrument. Commonly used for this purpose is a copper screen grid that is earthed, so that the potential of the grid is kept equal to the earth's potential.

3. If two conductors are brought into contact, the charge from one of them will flow over to the other until their potentials become equal. We shall now make use of this property to show that charges are located only on the external surface of a conductor.

Let us touch the outer surface of a charged body with a small uncharged sphere and then touch an electroroscope with the sphere. At this the leaves of the electroroscope diverge. This shows that a certain charge flows over to the sphere when it is brought into contact with the outer surface of the charged body. If we touch the inner surface of the charged body with the same sphere and then touch an electroroscope, we will find that the leaves do not diverge. This means that *there are no charges on the inner surface of a charged body*.

This principle can be used to completely transfer the charge from a sphere to a previously charged conductor. We do this by inserting the sphere into the cavity of the conductor and touching its *inner* surface. The charge from the sphere will then be completely transferred to the outer surface of the conductor, and the sphere will be completely discharged. This conclusion is confirmed by experiments.

4. Based on this property is the Van de Graaff electrostatic generator, shown schematically in Fig. 37.7. Here hollow metal sphere (or spherical shell) 1, from 1.5 to 2 m in diameter, is mounted at the top of insulating column 2 at a height of about 10 m above the

earth's surface. An endless rubber belt 3 is driven by motor 4. By means of brush 5 and electrode 6, charge from a small rectifier is "sprayed" onto the belt. This charge is carried by the belt into the shell where comb-shaped brushes 7, soldered to the inner surface, remove the charge from the belt and it flows to the inner surface of the spherical shell.

In principle, an unlimited potential can be obtained by this method. Actually the potential is restricted by the insulating properties of the column and the surrounding air. In practice a potential of several million volts can be obtained in an atmosphere of dry nitrogen at a pressure of 30 at.

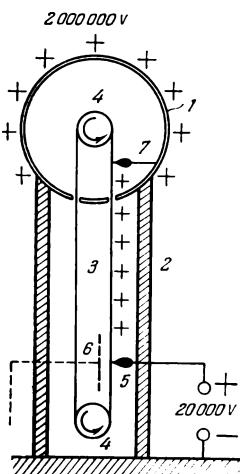


Fig. 37.7

some kind of elementary electric charges. This idea was precisely formulated in 1874 by G. J. Stoney (1826-1911) and in 1881 by H. von Helmholtz (1821-1894). In 1891 Stoney named this elementary electric charge the *electron*.

'At the turn of the century, J. J. Thomson (1856-1940), P. Lenard (1862-1947), W. Kaufmann (1871-1947) and others showed that cathode rays (Sec. 48.6), particles emitted by metals when they are heated (Sec. 47.1) and irradiated (Sec. 68.1), and beta particles emitted by certain radioactive elements, have similar properties. It was found that they are all negatively charged and that their charge-to-mass ratio (i.e. ratio of the charge to the rest mass) is the same and equal to about 10^{11} C/kg. It became clear that all these kinds of radiation were actually electrons obtained in different ways. This posed the problem of determining the charge and rest mass of the electron with the greatest possible accuracy.

2. The method employed by R. A. Millikan (1868-1953) was based on the study of the motion of charged microscopic particles in a uniform electric field. It is known as the *oil-drop experiment*. He began making measurements in 1906 and continued to 1914.

Millikan's oil-drop apparatus is illustrated schematically in Fig. 37.8. Here 1 is a parallel-plate capacitor with a small hole in

37.10. DETERMINING THE CHARGE OF THE ELECTRON

1. As far back as the thirties of last century Faraday came to the conclusion on the basis of an analysis of the laws of electrolysis, that he had discovered (see Secs. 46.3 and 46.4), that there must be what he called "atoms of electricity", i.e.

the top plate through which an occasional oil drop will fall. The drops are from a cloud produced by atomizer 2. The drops are irradiated by X-rays and are observed through low-powered microscope 3.

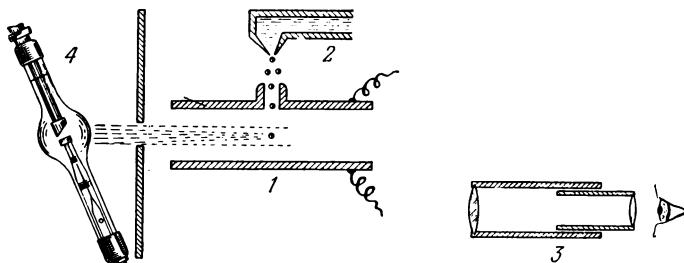


Fig. 37.8

scope 3. X-ray tube 4 is switched on periodically; its rays ionize the drop, which acquire a charge.

As shown in Sec. 11.9, a spherical drop falls through air with constant terminal velocity

$$v = \frac{2r^2g(\rho - \rho_0)}{9\eta} \quad (37.28)$$

where ρ = density of the oil

ρ_0 = density of the air

η = viscosity of the air

r = radius of the drop.

The velocity of the drop is determined by the time it falls between two crosshairs in the microscope field of view. The other quantities are known. Hence, we can determine the radius r by formula (37.28).

3. Then we set up an electric field E between the capacitor plates. If the drop carries the charge q , it will, in addition to the force of gravity, buoyant force and the force of resistance (see Sec. 11.9), be subject to the electric force $F_e = qE$. The equation of motion of the drop will be

$$\frac{4}{3}\pi r^3 \rho g - \frac{4}{3}\pi r^3 \rho_0 g - 6\pi r \eta u + qE = 0 \quad (37.29)$$

where u is the velocity of the drop in the field. Substituting equation (37.28) for the radius of the drop, we have

$$q = \frac{6\pi\eta(u-v)}{E} \sqrt{\frac{9\eta v}{2g(\rho - \rho_0)}} \quad (37.30)$$

4. The experiments registered the change in the charge of a drop when it was subjected to brief X-ray irradiation. From equation

(37.30) we can write

$$q_2 - q_1 = \frac{6\pi\eta(u_2 - u_1)}{E} \sqrt{\frac{9\eta v}{2g(p - p_0)}} \quad (37.31)$$

This expression served as the working formula.

Numerous difficulties were encountered in Millikan's investigations, especially in measuring the viscosity of the air. It was found that for microscopic drops the viscosity of the air depends on the size of the drops. A correction was introduced to take this dependence into account. It proved necessary to vary the conditions of the experiment in wide ranges in order to obtain reliable results. The potential difference was varied from 1700 to 5000 V; in the chamber with the capacitor the pressure was varied from standard atmospheric pressure to 45 mm-Hg; and drops were used with a radius from 0.446 to 5.856 microns.

All the experiments led to the following conclusion: the *change in the charge of a drop is always a multiple of the elementary charge*. In other words, $\Delta q = Ne$, where N is an integer and the elementary charge is $e = (1.592 \pm 0.0017) \times 10^{-19}$ C. This is the charge of the electron. The most recent and probable value of the charge of the electron is

$$e = (1.6021917 \pm 0.0000070) \times 10^{-19}$$
 C

Chapter 38

DIELECTRICS

38.1. AN ELECTRIC FIELD WITH A DIELECTRIC

1. Let us charge a capacitor to a certain potential difference φ_0 and then disconnect it from the current source. If we fill the space between the plates with a dielectric, we shall find that the potential difference is reduced to $\varphi < \varphi_0$ though the charge on the plates did not change.

Consequently, the field strength in the dielectric $E = \varphi/d$ is less than the strength of the field $E_0 = \varphi_0/d$ set up by the same charges in a vacuum. The ratio of the field strength E_0 in a vacuum to the field strength E in a dielectric is called the *dielectric constant*, or relative permittivity, of the substance (dielectric):

$$\epsilon = \frac{E_0}{E} \quad (38.1)$$

We see from the definition that the dielectric constant is a dimensionless quantity (an abstract number).

2. How can we explain the reduction of the field in a dielectric

in comparison to that in a vacuum? The field strength in a vacuum is fully determined by the charges on the capacitor plates (see equation (37.17)):

$$E_0 = \frac{\sigma}{\epsilon_0} \quad (38.2)$$

where σ is the surface density of free electric charges. The decrease in the field strength in the dielectric can be explained if we assume that the electric field will induce a *polarization charge* with a surface density σ_{pol} in the dielectric (Fig. 38.1). Then

$$E = \frac{\sigma - \sigma_{pol}}{\epsilon_0} \quad (38.3)$$

The appearance of a polarization charge in a dielectric due to an external electric field is referred to as the *polarization* of the dielectric.

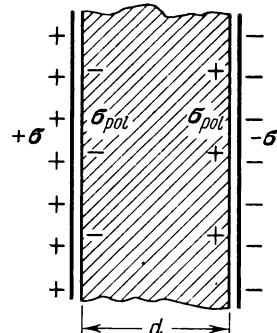


Fig. 38.1

38.2. THE POLARIZATION VECTOR

1. The essence of polarization phenomena is that the particles making up the dielectric (atoms or molecules) acquire dipole moments that tend to align with the lines of force of the external electric field. The mechanism of polarization in some dielectrics will be considered below. Here we shall introduce a quantity that characterizes the degree of polarization of the dielectric, the *polarization vector*.

The polarization vector \mathbf{P} is the dipole moment of unit volume of the polarized dielectric. It is calculated by finding the *vector sum* of the dipole moments of all the molecules or atoms contained in unit volume of the substance. For simplicity let us assume that the dielectric is homogeneous, and, therefore, all of its molecules (or atoms) have the same dipole moment \mathbf{p}_e . Then the polarization vector

$$\mathbf{P} = \frac{N\mathbf{p}_e}{V} = n\mathbf{p}_e \quad (38.4)$$

where n is the concentration of molecules (or atoms).

2. The dimensions of the magnitude of the polarization vector coincide with those of the surface density of a charge:

$$[P] = [Np_e/V] = [\sigma]$$

We shall now show that the magnitude of the polarization vector is equal to the surface density of the polarization charge. For proof we shall refer again to Fig. 38.1. A polarized dielectric can be treated

as a dipole carrying the polarization charge $q_{pol} = \sigma_{pol}A$. The moment of this dipole is equal to $q_{pol}d = \sigma_{pol}Ad = \sigma_{pol}V$. Dividing the dipole moment of the whole dielectric by its volume we obtain, according to the definition of the magnitude of the polarization vector,

$$P = \frac{q_{pol}d}{V} = \frac{\sigma_{pol}AD}{V} = \sigma_{pol} \quad (38.5)$$

Substituting the value $\sigma_{pol} = P$ into equation (38.3), we have
 $\epsilon_0 E = \sigma - P$ (38.6)

3. We introduce an auxiliary quantity D , called the *displacement vector*, or simply the *electric displacement*. Its magnitude is equal to the surface density of free charges: $D = \sigma$. Then we have

$$\epsilon_0 E = D - P \quad (38.7)$$

Since the directions of the field strength and polarization vectors coincide in a homogeneous and isotropic dielectric, we see that the displacement vector should have the same direction too. Consequently, equation (38.7) can be written in vector form:

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \quad (38.8)$$

38.3. ELECTRIC SUSCEPTIBILITY

1. In most dielectrics, except certain types of crystals (called ferroelectrics), the polarization vector is proportional to the field strength:

$$\mathbf{P} = \chi_e \epsilon_0 \mathbf{E} \quad (38.9)$$

The quantity χ_e is called the *electric susceptibility* of the dielectric. It will be left as a problem for the student to show, by analyzing the dimensions, that the electric susceptibility is a dimensionless quantity.

2. Substituting into equation (38.6) the expressions for the polarization vector and the surface density of free electric charges $\sigma = \epsilon_0 E_0 = \epsilon \epsilon_0 E$, we have

$$\epsilon_0 E = \epsilon \epsilon_0 E - \chi_e \epsilon_0 E$$

This suggests that the dielectric constant and electric susceptibility of a substance are in the following relationship:

$$\epsilon = 1 + \chi_e \quad (38.10)$$

We leave it as an exercise for the student to show that the displacement and field strength vectors are related as follows:

$$\mathbf{D} = \epsilon \epsilon_0 \mathbf{E} \quad (38.11)$$

38.4. THE FIELD ENERGY IN A DIELECTRIC

1. In a parallel-plate capacitor with a dielectric, the field strength and the potential difference are

$$E = \frac{E_0}{\epsilon} = \frac{\sigma}{\epsilon\epsilon_0} = \frac{q}{\epsilon\epsilon_0 A} \quad (38.12)$$

and

$$\varphi_1 - \varphi_2 = Ed = \frac{qd}{\epsilon\epsilon_0 A} \quad (38.13)$$

Hence the capacitance of a parallel-plate capacitor with a dielectric between its plates is

$$C = \frac{q}{\varphi_1 - \varphi_2} = \frac{\epsilon\epsilon_0 A}{d} \quad (38.14)$$

2. Equation (37.23) for the energy of a charged capacitor is valid for this case as well. The expression for the energy density becomes

$$u = \frac{U}{V} = \frac{q\varphi}{2V} = \frac{C\varphi^2}{2V} = \frac{\epsilon\epsilon_0 A\varphi^2}{2d \times Ad} = \frac{\epsilon\epsilon_0 \varphi^2}{2d^2}$$

But $\varphi/d = E$, then

$$u = \frac{\epsilon\epsilon_0 E^2}{2} \quad (38.15)$$

Hence, in the presence of a dielectric the energy density of the electric field increases ϵ -fold in comparison to that in a vacuum, but only if the field strength vector remains constant. If the displacement vector remains constant, then the energy density is reduced ϵ -fold. Indeed, in this case we should substitute the value $E = \frac{D}{\epsilon\epsilon_0}$ from equation (38.11) into equation (38.15). Then

$$u = \frac{D^2}{2\epsilon\epsilon_0} \quad (38.16)$$

3. Let us find the change in the force of interaction between the capacitor plates if the space between them is filled with a gaseous or liquid dielectric. By the same arguments we employed in Sec. 37.8, we come to the conclusion that the pressure p of the electric forces is equal to the energy density of the electric field, but taken with the opposite sign: $p = -u$ [see equation (37.27)].

If the capacitor plates are disconnected from the current source, the charge on them will remain constant. Consequently, the displacement vector will also remain constant because it equals, by definition, the surface density of the charge on the plates ($D = \sigma$). The force of interaction is

$$F = pA = -uA = -\frac{D^2 A}{2\epsilon\epsilon_0} = -\frac{q^2}{2\epsilon\epsilon_0 A} = \frac{F_0}{\epsilon} \quad (38.17)$$

Thus, in this case the force of interaction between the plates with a dielectric is ϵ -fold less than in a vacuum.

If, however, the plates remain connected to the current source, the potential difference between the plates is not changed, nor, consequently, is the field strength. Here the force of interaction between the plates will be ϵ -fold larger than in a vacuum.

Thus

$$F = pA = -uA = -\frac{\epsilon\epsilon_0 E^2}{2} A = -\frac{\epsilon_0^2 A}{2\epsilon_0} = \epsilon F_0 \quad (38.17')$$

4. As we have seen, the force of interaction between charged bodies depends in a quite complex way on whether there is a dielectric between them. If there is, and the displacement vector does not change, the interaction force will be ϵ -fold less than that in a vacuum; but if it is the field strength vector that remains constant, the interaction force will be ϵ -fold larger than that in a vacuum.

In all other cases a detailed analysis will be required to determine how the presence of a dielectric changes the force of interaction between charged bodies. We cannot give a general formula here like Coulomb's law of the interaction of charges in a vacuum.

38.5. DEFORMATION POLARIZABILITY

1. In introducing the concept of the polarization vector we assumed that the molecules or atoms have a dipole moment of a direction aligned with the field strength (Sec. 38.2). Let us investigate the mechanism of the polarization of dielectrics, i.e. the reason the molecules or atoms have or acquire dipole moments that tend to align with the field. For the sake of simplicity, we shall restrict ourselves to a discussion of the mechanism of polarization of gases, in which we can neglect the action of neighbouring molecules in comparison to the effect of the external field.

There turns out to be two possible mechanisms of polarization, either one or the other predominating in various substances in accordance with the structure of their molecules.

2. The atoms or molecules of many substances have a symmetrical structure and therefore have no dipole moment. When an external electric field is applied to such an atom, its electron cloud is distorted (displaced), its symmetry is violated, and it acquires a dipole moment. Since polarizability of this kind is caused by the deformation of the atom's electron cloud, it is called *deformation*, or *electronic, polarizability*.

The simplest model of such an atom is illustrated in Fig. 38.2. Located at the centre of the atom is its positively charged nucleus, about which the electron rotates. To explain the symmetrical spherical structure of this atom, it is reasonable to assume that the orbit

of the electron continuously rotates so that the electron travels quite rapidly along the sphere at the centre of which the nucleus is located. Since the centre of the positively charged nucleus and the centre of the negatively charged electron cloud coincide, the atom has no dipole moment.

3. Forces of the external electric field act on the nucleus and electron in opposite directions. As a result the electron orbit is displaced with respect to the nucleus by the distance l . At this the atom acquires the dipole moment $p_e = el$. As has been previously shown (see Sec. 31.3), the dipole moment of a deformed atom is

$$p_e = el = 4\pi r^3 \epsilon_0 E = \alpha \epsilon_0 E \quad (38.18)$$

where $\alpha \cong 4\pi r^3$ is called the polarizability of the molecule. The magnitude of the polarization vector is

$$P = np_e = n\alpha \epsilon_0 E \quad (38.19)$$

Comparing this expression with equations (38.9) and (38.10), we can write the expression for the electric susceptibility and the dielectric constant of a substance (a gas) with nonpolar molecules:

$$\chi_e = n\alpha \text{ and } \epsilon = 1 + \chi_e = 1 + n\alpha \quad (38.20)$$

4. Let us evaluate this result, taking the radius of the atom to be about $r \cong 10^{-10}$ m and the concentration of the gas molecules, $n \cong \cong 2.7 \times 10^{25}$ m⁻³. This yields the value $\chi_e \cong 4\pi r^3 n \cong 3 \times 10^{-4}$ for the susceptibility. The order of magnitude turns out to be correct: under standard conditions, the electric susceptibility is 2.6×10^{-4} for hydrogen, 0.5×10^{-4} for helium, 5.2×10^{-4} for oxygen and 5.4×10^{-4} for argon. Even though the model of the atom used in our reasoning was quite primitive, the agreement of the calculated and experimental data is very good.

38.6. ORIENTATIONAL POLARIZABILITY

1. In many substances the molecules have a dipole moment, due to their asymmetric structure, even when there is no external field. Examples are the molecules of water (dipole moment $p_e = 6.36 \times 10^{-30}$ C-m), ammonia (4.88×10^{-30} C-m), acetone (9.30×10^{-30} C-m), ethyl alcohol (5.59×10^{-30} C-m) and benzene (0.27×10^{-30} C-m). Molecules of this type are said to be *polar*.

In the absence of an external field the dipole moments of polar molecules are oriented entirely at random. Consequently, the polarization vector, equal to the *vector sum* of the dipole moments in unit

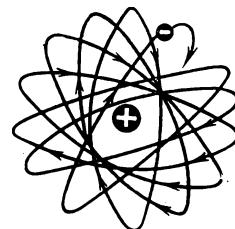


Fig. 38.2

volume of the dielectric, should equal zero if there is no external field. If a substance with polar molecules is placed into an electric field, the molecules should be distorted, for one thing, and their dipole moments should change and, for another, the molecules should turn, due to the moment of force that acts on any dipole in an electric field (Sec. 37.4), so that their dipole moments are aligned along the lines of force of the field. The first effect, caused by the deformations of the molecules in a field, was dealt with in the preceding section. Moreover, it is usually much weaker than the second effect which is called *orientational polarizability*.

The degree of orientation of polar molecules in an electric field is determined both by the field strength and the temperature of the substance. The action of the field turns the molecular dipoles so that they are aligned with the field. Thermal motion violates this order, disarranging the moments of the molecular dipoles so that they point in all possible directions. The situation here is quite similar to the distribution of gas molecules in a gravitational field, in which the force field seems to "press" the molecules to the earth's surface and the thermal motion scatters them in all possible directions (Secs. 26.10 and 26.11). As in the action of a gravitational field, the probability that a molecular dipole has the energy \mathcal{E} in an electric field is determined according to formula (26.25) by the expression $w = e^{-\mathcal{E}/kT}$, where $\mathcal{E} = -p_e E \cos \alpha$ is the energy of the dipole [equation (37.12)], T is the absolute temperature of the substance and k is Boltzmann's constant.

2. Let us carry out a simplified calculation. When there is no external field, the molecular dipoles are arranged entirely at random, and we can assume that the dipole moments are uniformly distributed along all the coordinate axes. Then one third of the dipoles are arranged along any axis: $N_x = N_y = N_z = N/3$. On an average, one half of the dipoles are oriented in the direction of the axes and the other half in the opposite direction.

We put the dielectric in an electric field whose strength vector is along the x -axis. Then the number of dipoles N_1 directed along the field will increase and the number N_2 directed opposite the field will decrease. The magnitude of the polarization vector is

$$P = \frac{N_1 p_e - N_2 p_e}{V} \quad (38.21)$$

According to formula (26.26), we have

$$N_1 = A e^{-\frac{\mathcal{E}_1}{kT}} \text{ and } N_2 = A e^{-\frac{\mathcal{E}_2}{kT}} \quad (38.22)$$

where A is a constant. For a dipole oriented along the field, $\alpha = 0$ and its energy is $\mathcal{E}_1 = -p_e E$; for one oriented opposite the field, $\alpha = \pi$ and its energy is $\mathcal{E}_2 = p_e E$.

3. Dipole moments are very small ($p_e \cong 10^{-30}$ C-m). Therefore, even in very strong fields ($E \cong 10^6$ V/m) the energy of a molecular dipole is $\mathcal{E} \leq 10^{-24}$ J. This is considerably less than the energy of thermal motion at room temperature:

$$kT = 1.38 \times 10^{-23} \times 300 \cong 4 \times 10^{-21} \text{ J}$$

Thus the ratio $x = -\mathcal{E}/kT$ is a very small value: $|x| < 10^{-3}$. For such small values of x we can write: $e^x \cong 1 + x$. Then we obtain the following expressions for the numbers of dipoles oriented along with and against the field:

$$\left. \begin{aligned} N_1 &\cong A(1+x_1) = A\left(1 - \frac{\mathcal{E}_1}{kT}\right) = A\left(1 + \frac{p_e E}{kT}\right) \\ N_2 &\cong A(1+x_2) = A\left(1 - \frac{\mathcal{E}_2}{kT}\right) = A\left(1 - \frac{p_e E}{kT}\right) \end{aligned} \right\} \quad (38.23)$$

Therefore, the equation: $N_x = N_1 + N_2 = 2A$ is valid for weak fields. But since $N_x = N/3$, $2A = N/3$. Thus

$$A = N/6 \quad (38.24)$$

4. The results obtained for the numbers of molecules oriented along with the field N_1 and opposite the field N_2 enable us to calculate the magnitude of the polarization vector. Substituting equations (38.23) and (38.24) into equation (38.21), we obtain

$$P = \frac{(N_1 - N_2) p_e}{V} = \frac{2A p_e^2 E}{V k T} = \frac{N p_e^2 E}{3V k T} = \frac{n p_e^2 E}{3kT} \quad (38.25)$$

It follows that the electric susceptibility of a gas made up of polar molecules is of the form

$$\chi_e = \frac{n p_e^2}{3\epsilon_0 k T} \quad (38.26)$$

Notwithstanding the assumptions made to simplify calculations, the result obtained is correct.

The fact that with a constant concentration of the molecules (and therefore constant density of the gas), the electric susceptibility in orientational polarizability varies inversely proportional to the absolute temperature is called the *Debye law*:

$$\chi_e = \frac{\text{const}}{T} \quad (38.27)$$

It has been reliably confirmed experimentally for many gases and vapours.

This law, naturally, does not hold for low temperatures, where we may find that $p_e E \cong kT$. But in this case all of our arguments are incorrect.

Chapter 39

DIRECT CURRENT

39.1. NONELECTROSTATIC FIELDS. VOLTAGE AND EMF

1. An electric current is the ordered motion of electric charges. If such motion occurs in a conductor, the electric current is said to be a conduction current.

Many factors initiate an ordered motion of charges. Primarily, these may be electric (Coulomb) forces by whose action positive charges start to move along the lines of force of the field, and negative charges against the lines. We shall call a field of such forces a Coulomb field and denote its strength by E_{Coul} .

Besides, electric charges may be subject to nonelectrostatic forces, for instance, magnetic ones (see Ch. 41). The action of these forces is similar to that of a certain electric field. We shall call them *extraneous forces*, and the field of such forces, an extraneous field with the strength E^* .

Finally, the ordered motion of electric charges may be established without any action of external forces. It may be due to diffusion or to chemical reactions in the current source. The work done to maintain the ordered motion of electric charges is at the expense of the internal energy of the current source. Though there is no direct action of any forces on the free charges, the phenomenon proceeds as if the charges were subject to some extraneous field. Therefore, for the sake of generality we shall refer to an effective extraneous field E^* in this case as well.

2. The energy characteristic of a Coulomb field is the potential. We can introduce quantities that could serve as the energy characteristic of an extraneous field. According to definition (Sec. 18.7), the potential difference is equal to the ratio of the work done by Coulomb forces in moving a charge to the magnitude of the charge:

$$\Delta\varphi = \varphi_1 - \varphi_2 = W_{\text{Coul}}/q \quad (39.1)$$

If, in addition to Coulomb forces, work is done to move the charge by extraneous forces, the total work is

$$W = W_{\text{Coul}} + W_{\text{extr}} \quad (39.2)$$

where W_{Coul} = work done by the Coulomb forces

W_{extr} = work done by nonelectric sources of energy, for example, magnetic forces, or at the expense of the change in internal energy in the diffusion of charged particles, or the change of the internal energy of substances participating in a chemical reaction, or at the expense of light energy (photoelectric effect, see Sec. 68.1).

Dividing the left and right sides of equation (39.2) by the magnitude of the charge being moved, we obtain

$$\frac{W}{q} = \frac{W_{\text{Coul}}}{q} + \frac{W_{\text{extr}}}{q} \quad (39.3)$$

3. The *voltage* over a segment of a circuit is the quantity equal to the ratio of the total work done in moving a charge along this segment to the magnitude of the charge. Thus

$$u = W/q \quad (39.4)$$

The *electromotive force* (abbreviation *emf*) over a given segment of a circuit is a quantity equal to the ratio of the work done by non-electric forces in moving a charge to the magnitude of the charge

$$\mathcal{E} = W_{\text{extr}}/q \quad (39.5)$$

The definition of these quantities indicates that in SI units both the voltage and the emf are measured in volts (V).

4. Combining equations (39.3), (39.4) and (39.5) we have

$$u = \varphi_1 - \varphi_2 + \mathcal{E} \quad (39.6)$$

Thus, the voltage over a segment of a circuit is equal to the sum of the potential difference and of the electromotive force.

The following special cases are possible:

(a) if a segment of a circuit is uniform and not subject to extraneous forces, the emf equals zero and the voltage equals the potential difference:

$$u_{\text{unif}} = \varphi_1 - \varphi_2 \quad (39.7)$$

(b) In a closed circuit the voltage equals the algebraic sum of the emf's. Thus, suppose the closed circuit consists of three segments, then

$$u_1 = \varphi_1 - \varphi_2 + \mathcal{E}_1, \quad u_2 = \varphi_2 - \varphi_3 + \mathcal{E}_2 \quad \text{and} \quad u_3 = \varphi_3 - \varphi_1 + \mathcal{E}_3$$

The work done in moving a charge along a closed circuit equals the sum of the work done in each segment. Hence

$$u_{\text{closed}} = u_1 + u_2 + u_3 = \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3 \quad (39.8)$$

5. As shown in Sec. 37.3, the strength of a Coulomb field is $E_{\text{Coul}} = (\varphi_1 - \varphi_2)/l$. Dividing equation (39.6) by the length l of the conductor, we obtain

$$\frac{u}{l} = \frac{\varphi_1 - \varphi_2}{l} + \frac{\mathcal{E}}{l} \quad (39.9)$$

By analogy we shall call the quantity u/l the *average strength of the composite field* and denote it by E . Thus

$$E = u/l \quad (39.10)$$

Accordingly, we shall call the quantity \mathcal{E}/l the *average strength of the extraneous field* and denote it by E^* . Thus

$$E^* = \mathcal{E}/l \quad (39.11)$$

On the basis of these definitions we can rewrite equation (39.9) as follows:

$$E = E_{\text{Coul}} + E^* \quad (39.12)$$

i.e. the strength of the composite field is the sum of the strengths of Coulomb and extraneous fields.

39.2. CURRENT AND CURRENT DENSITY

1. The quantitative characteristic of the current is the quantity equal to the ratio of the charge Δq passing through a cross section of a conductor during the time Δt to this time interval. Thus

$$i = \frac{\Delta q}{\Delta t} \quad (39.13)$$

This was previously called the current strength but is now simply called the *current*.

We obtain the exact value of the current under the condition that the time interval Δt is infinitely small:

$$i = \lim_{\Delta t \rightarrow 0} \frac{\Delta q}{\Delta t} \quad (39.14)$$

2. If the current and its direction do not change with time it is called a *direct* (more exactly, *constant*) *current*. For a direct current

$$i = q/t \quad (39.15)$$

It should be noted that in a simple (unbranched) circuit of direct current, the current is the same through all cross sections of the conductor. Indeed, the charge Δq_1 flowing through a cross section A_1 of the conductor during the time Δt should be equal to the charge Δq_2 flowing through another cross section A_2 during the same interval of time. Otherwise the charge would accumulate between the two cross sections changing the electric field. This in turn would change the current which we have taken to be constant. But it follows from $\Delta q_1 = \Delta q_2$ that $\Delta q_1/\Delta t = \Delta q_2/\Delta t$ or $i_1 = i_2$ which is what we started out to prove.

3. The SI unit of current is the *ampere* (A). An ampere is the constant electric current corresponding to a flow of one coulomb of charge per second through a cross section of a conductor:

$$1 \text{ A} = 1 \text{ C}/1 \text{ s}$$

4. It is sometimes more convenient to use the quantity called the *current density*. The average current density is equal to the ratio

of the current to the cross-sectional area of the conductor. Thus

$$j = i/A \quad (39.16)$$

The SI unit of current density is the A/m². Also used in practice is the A/mm²: 1 A/mm² = 10⁶ A/m².

5. Let us express the current and current density in terms of the rate of the ordered motion of free charges in a conductor. Suppose the concentration of free charges is $n = \Delta N/\Delta V$, where ΔV is a part of the volume of the conductor. The electric charge of one free current carrier equals e and the average velocity of its ordered motion equals \bar{v} . Then the charge $\Delta q = e\Delta N = en\Delta V = enA\bar{v}\Delta t$ is carried through cross section A during Δt . The current is

$$i = \frac{\Delta q}{\Delta t} = enA\bar{v} \quad (39.17)$$

The current density is

$$j = i/A = en\bar{v} \quad (39.18)$$

We shall use this equation further on for solving various problems.

6. Let us estimate the average velocity of the ordered motion of conduction electrons in metals. The concentration of free electrons coincides approximately with the concentration of atoms: $n = 10^{28}$ to 10^{29} m⁻³. The highest current density in metal conductors is about 10 A/mm² = 10⁷ A/m². Then

$$\bar{v} = \frac{j}{en} \cong \frac{10^7}{1.6 \times 10^{-19} \times 10^{28}} \cong 6 \times 10^{-3} \text{ m/s} = 6 \text{ mm/s}$$

Experiments confirm the order of magnitude of this value. Such a low velocity for the ordered motion of conduction electrons would seem to contradict the fact that in switching on an electric circuit a steady current is established practically instantaneously. However, this is no paradox. The fact is that the electric field set up by the current source appears practically simultaneously throughout the circuit. Therefore, the velocity v of ordered motion, equal to about 6 mm/s, is added simultaneously to the velocity vectors of thermal electronic motion in any segment of the conductor.

39.3. OHM'S LAW FOR A UNIFORM SEGMENT OF A CIRCUIT :

1. The student already knows of one of the principal laws of direct currents, Ohm's law for a uniform segment of a circuit. In essence this law states the following. Suppose that we have a uniform conductor; it can be a piece of homogeneous metal of constant cross section, all points of which are at the same temperature. If we maintain a constant potential difference $\Delta\varphi = \varphi_1 - \varphi_2$ between the ends of this conductor we will have the current i which is proportional to the

potential difference. Thus

$$i = G \Delta\varphi = \frac{\Delta\varphi}{R} \quad (39.19)$$

Hence, in a uniform segment of a circuit the current is proportional to the potential difference at its ends.

2. Factor G is called the *conductance* of the segment. The reciprocal

$$R = \frac{1}{G} = \frac{\Delta\varphi}{i} \quad (39.20)$$

is called the *electrical resistance* of the segment (or simply the *resistance*).

39.4. RESISTANCE

1. The student knows that the resistance of a metallic conductor is proportional to its length l and inversely proportional to its cross-sectional area A . Thus

$$R = \frac{\rho l}{A} \quad (39.21)$$

A conductor whose resistance is determined by equation (39.21) is commonly called a *resistor*.

The quantity ρ is called the *resistivity* of the substance of which the conductor is made. The reciprocal

$$\gamma = 1/\rho \quad (39.22)$$

is called the *electric conductivity*.

2. The SI unit of resistance is the *ohm*. The ohm is defined as the resistance offered by a conductor with a current of one ampere if a potential difference of one volt is maintained at its ends. Thus

$$1 \text{ ohm} = 1 \text{ V}/1 \text{ A}$$

The SI unit of resistivity is the ohm-m:

$$[\rho] = [RA/l] \cong \text{ohm} \times \text{m}^2/\text{m} = \text{ohm-m}$$

The unit frequently used in practice is only one millionth of the SI unit:

$$1 \text{ ohm} \times \text{mm}^2/\text{m} = 10^{-6} \text{ ohm-m}$$

3. For quite a wide temperature range, far from absolute zero, the resistivity of metallic conductors is a linear function of the temperature t (on the Celsius scale). Thus

$$\rho = \rho_0 (1 + \alpha t) \cong \rho_0 \alpha T \quad (39.23)$$

(as a matter of fact, $1 + \alpha t \cong 1 + \frac{t}{273} = \frac{T}{273} \cong \alpha T$ since $\alpha \cong \cong (4 \text{ to } 5) \times 10^{-3} \text{ K}^{-1} \cong \frac{1}{273} \text{ K}^{-1}$).

39.5. OHM'S LAW IN DIFFERENTIAL FORM

1. Let us carry out some transformations in Ohm's law for a segment of a circuit. For this purpose in formula (39.19) we shall express the current in terms of the current density according to equation (39.16), the resistance in terms of properties of the conductor according to equation (39.21), and the potential difference in terms of the field strength [see equation (37.9)]. We then obtain $jA = = ELA/\rho l$ or, after cancelling, $j = E/\rho$. Expressing the resistivity in terms of the conductivity according to equation (39.22), we have

$$j = \gamma E \quad (39.24)$$

Thus, *the current density is proportional to the field strength at the given point of the conductor*. This is called Ohm's law in the differential form.

2. Substituting into equation (39.24) expression (39.18) for the current density, we have $env = \gamma E$, from which

$$\bar{v} = \frac{\gamma E}{en} \quad (39.25)$$

We find that the *average velocity of ordered motion of free charges in a conductor* (for example, free electrons in metals) *is proportional to the strength of the electric field in the conductor*.

3. The ratio of the average velocity of ordered motion of a free charge to the field strength in the conductor is called the *mobility of the charge*:

$$b = \frac{\bar{v}}{E} = \frac{\gamma}{en} \quad (39.26)$$

39.6. OHM'S LAW FOR A NONUNIFORM SEGMENT OF A CIRCUIT AND FOR A CLOSED CIRCUIT

1. The term E in equation (39.24) is actually the strength of a Coulomb field because this expression was derived from Ohm's law for a uniform segment of a circuit where we have only a Coulomb field. It makes no difference, however, what kind of field acts on the charge—Coulomb, extraneous or composite field. Consequently, if an extraneous field acts in the given (nonuniform) segment of the circuit in addition to the Coulomb field, then the *velocity of motion of the charges and, accordingly, the current density will be proportional to the strength of the composite field*.

Substituting into equation (39.24) the value of the strength of the composite field [equation (39.10)] and of the current density [equation (39.16)], we have $i/A = \gamma u/l$, from which

$$u = \frac{il}{\gamma A} = \frac{i\rho l}{A} = iR \quad (39.27)$$

2. Hence, for a nonuniform segment of a circuit, Ohm's law is stated as follows: the current is proportional to the voltage over the segment and inversely proportional to its resistance. Thus

$$i = u/R \quad (39.28)$$

Substituting the value of the voltage from equation (39.6), we can write Ohm's law in yet another form:

$$i = \frac{\varphi_1 - \varphi_2 + \mathcal{E}}{R} \quad (39.29)$$

Here we should take into account the sign of the emf. If the direction of the strength vector of the extraneous force field coincides with that of the Coulomb field, then the emf and the potential difference are of like sign; otherwise they are of opposite sign.

3. In a closed circuit the potential difference equals zero since a Coulomb field is conservative. Therefore, for a closed circuit Ohm's law will be

$$i = \mathcal{E}/R \quad (39.30)$$

Here R is the total resistance of the circuit including the internal resistance of the current source*, and \mathcal{E} is the algebraic sum of the emf's. Ohm's law for a closed circuit is often written as

$$i = \frac{\mathcal{E}}{R+r} \quad (39.31)$$

where R is the resistance of all the resistors forming the circuit (resistance of the load) and r is the internal resistance of the current source.

39.7. THE JOULE-LENZ LAW

1. In resistors there is an irreversible conversion of the energy of the electric current into the internal energy of the conductor. We shall clear up the mechanism of this phenomenon in Sec. 44.6. Here we shall deal only with certain quantitative relationships.

If conductors are stationary and no chemical reactions occur in them as a result of the current, then their temperature will change. The change in the internal energy of a conductor (quantity of heat) Q is equal to the work W done by the composite field in moving charges. Thus

$$Q = W = qu = iut \quad (39.32)$$

* Usually called the *seat of emf*.

Making use of Ohm's law (39.28), we obtain two equivalent expressions

$$Q = i^2 R t = u^2 t / R \quad (39.33)$$

This is the Joule-Lenz law. Note that the quantity of heat Q is expressed in the SI unit, the joule (J).

2. Naturally, either of the formulas (39.33) can be used to calculate the thermal effect of a current. If we wish to compare two resistors on the basis of the thermal processes that take place in them, we must first find out whether they have the same current or whether the same voltage is applied to them.

If the current is the same in the resistors then, according to the formula

$$Q = i^2 R t$$

the internal energy increases a greater amount in the resistor with the higher resistance. This is the case when we have a circuit with resistors connected in series.

If the voltage is the same over the resistors, then, according to the formula

$$Q = u^2 t / R$$

the resistor with the lower resistance is more intensively heated. This effect is observed, for instance, in the parallel connection of resistors.

39.8. CHARGING AND DISCHARGING CAPACITORS

1. Let us consider a circuit consisting of a resistor, a capacitor and a current source (Fig. 39.1). Assume that at the initial instant of time the capacitor is not charged. Then the potential difference between its plates equals zero.

Current begins to flow when we close the circuit. Initially the current is $i_0 = \mathcal{E}/R$ because the capacitor is uncharged. As the capacitor becomes charged, a potential difference is set up between its plates and the Coulomb field will resist the motion of the free charges thereby reducing the current in the circuit. According to Ohm's law

$$i = \frac{\mathcal{E} - \Delta\varphi}{R}$$

When the potential difference between the plates becomes equal to the emf of the current source, there will no longer be a current. A graph of the variation of the potential and current with time is shown in Fig. 39.2.

2. An exact calculation of the phenomena in a circuit is impossible by elementary methods. An application of the law of conservation of energy, however, will enable us to estimate approximately the time interval Δt during which the capacitor becomes charged.

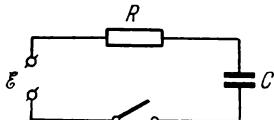


Fig. 39.1

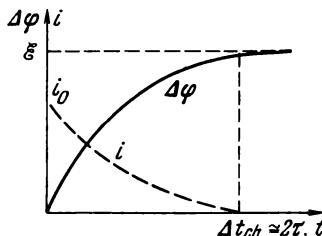


Fig. 39.2

In fact, the work done by the current source during this interval equals the change in the internal energy of the resistor plus the change in the energy of the capacitor. Thus

$$W_{ext} = Q + \frac{1}{2} C \varphi^2$$

The final potential of the capacitor is equal to the emf of the source, and the work done by the latter to change the internal energy of the resistor can be expressed in terms of the average current:

$$i_{av} \mathcal{E} \Delta t = i_{av}^2 R \Delta t + \frac{1}{2} C \mathcal{E}^2$$

We can assume that $i_{av} \cong i_0/2 = \frac{\mathcal{E}}{2R}$. Then we have

$$\frac{\mathcal{E}^2}{2R} \Delta t \cong \frac{\mathcal{E}^2 R \Delta t}{4R^2} + \frac{C \mathcal{E}^2}{2}$$

or

$$\Delta t_{ch} \cong 2RC = 2\tau \quad (39.34)$$

The product $\tau = RC$ is called the *capacitive time constant* of a circuit consisting of a resistor and capacitor. It is also known as the *relaxation time*. This quantity characterizes the time required for an equilibrium state to be established in the circuit.

3. It will be left as an exercise for the student to determine the approximate time required for a charged capacitor to discharge through a resistor. This is done by equating the energy of the charged capacitor to the change in the internal energy of the resistor. To simplify matters we shall assume that $i_{av} \cong \frac{\varphi_0}{2R}$. Then

$$\Delta t_{dis} \cong 2RC \cong 2\tau$$

Chapter 40

A MAGNETIC FIELD IN A VACUUM

40.1. THE INTERACTION OF CURRENTS. MAGNETIC FORCES

1. Let us stretch two tinfoil bands parallel to each other and connect them to a current source. If the currents are in the same direction along these two conductors, the bands will attract each other. If the currents are in opposite directions, the bands will repel each other. Let us try to find the explanation of this phenomenon.

Metals have free electrons (electron gas) distributed inside their ion lattice. The total charge of the positively charged lattice and the negatively charged electron gas equals zero. The charges are distributed uniformly in the conductor so that the system has no dipole moment. As a result the total electric field of the ion lattice and electron gas outside the conductor equals zero, and there is no electric field around the conductor. This is why the conductors do not interact when there is no current in them.

However, if there is a current, i.e. ordered motion of free charges, a force of interaction appears between the conductors. We, therefore, conclude that the *forces of interaction between moving electric charges differ from the forces of interaction between fixed charges*.

The additional forces of interaction existing between moving charges have been named, for historical reasons, *magnetic forces*. The fact is that similar forces act on a conductor with a current when it is near a permanent magnet.

2. A fact to be noted is that it is impossible to explain the nature of magnetic forces on the basis of Newtonian mechanics. Indeed, in Newtonian mechanics forces of interaction are associated only with accelerations. According to this principle the forces of interaction between charges moving uniformly should not differ from those between fixed charges. But this is not confirmed by experiments.

We can also approach this question in other ways. In Newtonian mechanics force is a function of the distance between interacting bodies and of their relative velocities. But the distance between conduction electrons does not change when a current is established. Their relative velocity with respect to one another remains equal to zero. It would seem that no forces could appear under these circumstances.

However, experiments demonstrate that forces appear in the motion of electrons; the forces are not there when the current is absent. This is another manifestation of the restrictions of Newtonian mechanics whose concepts are incapable of explaining magnetic interaction.

3. In the following sections we shall show that magnetic interaction is correctly interpreted by the concepts of relativity theory. The laws derived on this basis completely agree with experiment.

40.2. THE TRANSFORMATION LAW FOR THE SIDEWAYS MOMENTUM AND THE SIDEWAYS FORCE

1. The relativistic transformations for the coordinates and time were derived in Chapters 12 and 13. These included the Lorentz transformations (see Sec. 12.7) and the expressions for the momentum and force (see Secs. 13.1 and 13.2). Let us find the transformation for the projections of the momentum and force on the z -axis, i.e. for the *sideways momentum* and *sideways force*.

Suppose some body (or, more exactly, a mass point) with rest mass m_0 is moving at the velocity $u_z = \Delta z / \Delta t$ along the z -axis. Then the projection of its momentum on this axis is

$$p_z = mu_z = \frac{m_0 u_z}{\sqrt{1 - u_z^2/c^2}} \quad (40.1)$$

The projection of the force on this axis is

$$F_z = \frac{\Delta p_z}{\Delta t} \quad (40.2)$$

Next we shall determine the components of the momentum p'_z and force F'_z in another frame of reference that moves uniformly along the axis of abscissas at the velocity v (Fig. 40.1).

2. Since we are dealing with the motion of the body only along the z -axis, the abscissa of the body does not change ($x = \text{const}$). According to the Lorentz transformations, we have

$$z' = z \text{ and } \Delta t' = \frac{\Delta t}{\sqrt{1 - v^2/c^2}} \quad (40.3)$$

and the projection of the velocity of the body on the z -axis in the new frame is

$$u'_z = \frac{\Delta z'}{\Delta t'} = \frac{\Delta z}{\Delta t} \sqrt{1 - v^2/c^2} = u_z \sqrt{1 - v^2/c^2} \quad (40.4)$$

The mass in the new frame of reference is

$$m' = \frac{m_0}{\sqrt{1 - v^2/c^2}} \quad (40.5)$$

We should note that different velocities appear in equations (40.3) and (40.5). The point is that we consider the clock to be always at rest with respect to the reference frame. Hence, the velocity v in the time transformation law is that of one reference frame with respect to the other. The body moves with respect to both reference

frames: with the velocity u_z with respect to the unprimed frame and the velocity $V = \sqrt{v^2 + (u'_z)^2}$ with respect to the primed one (Fig. 40.1). This is the velocity in the mass transformation law.

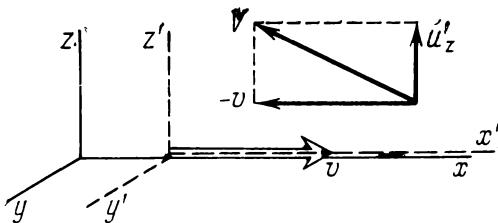


Fig. 40.1

Then we have

$$\begin{aligned} 1 - \frac{V^2}{c^2} &= 1 - \frac{v^2}{c^2} - \frac{(u'_z)^2}{c^2} = 1 - \frac{v^2}{c^2} - \frac{u_z^2}{c^2} \left(1 - \frac{v^2}{c^2}\right) \\ &= \left(1 - \frac{v^2}{c^2}\right) \left(1 - \frac{u_z^2}{c^2}\right) \end{aligned}$$

Substituting into equation (40.5) and combining with equation (40.1), we obtain

$$m' = \frac{m_0}{\sqrt{1-v^2/c^2}} \sqrt{1-u_z^2/c^2} = \frac{m}{\sqrt{1-v^2/c^2}} \quad (40.6)$$

Next we shall determine the sideways momentum in the new frame of reference. Using equations (40.4) and (40.6), we have

$$p'_z = m'u'_z = \frac{m}{\sqrt{1-v^2/c^2}} u_z \sqrt{1-v^2/c^2} = mu_z = p_z \quad (40.7)$$

Thus, in a transition from one inertial reference frame to another the *sideways momentum does not change*.

3. The sideways component of the force in the new reference frame is

$$F'_z = \frac{\Delta p'_z}{\Delta t'} = \frac{\Delta p_z}{\Delta t} \sqrt{1-v^2/c^2} = F_z \sqrt{1-v^2/c^2} \quad (40.8)$$

Hence, in contrast to Newtonian mechanics, the *sideways force in relativity theory does depend upon the velocity*. This explains the origin of magnetic forces.

40.3. INTERACTION BETWEEN MOVING CHARGES

1. Suppose that two point charges Q and q are located at the distance r from each other and are fixed with respect to the reference frame $x_0y_0z_0$ (Fig. 40.2a). Between the charges there will be the Coulomb

force

$$F_{0z} = \frac{qQ}{4\pi\epsilon_0 r^2} \quad (40.9)$$

Let us find the forces acting between the same charges in reference frame xyz , which moves to the right along the axis of abscissas with

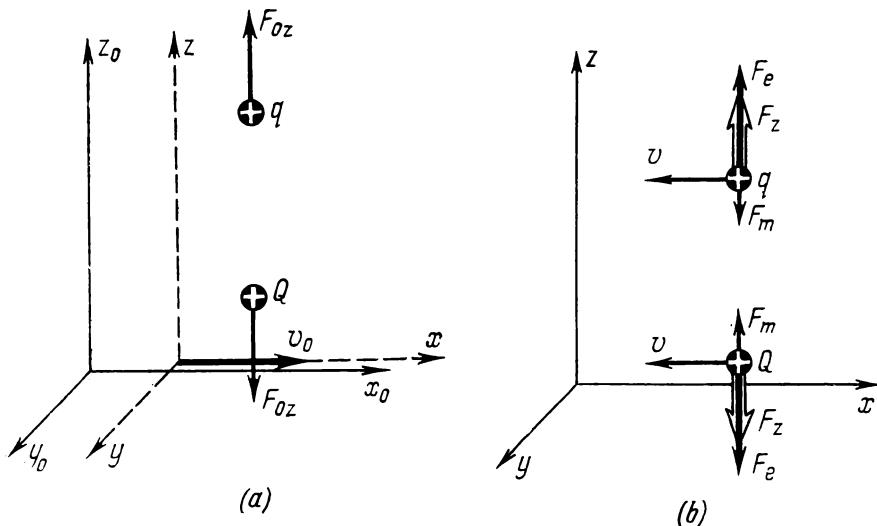


Fig. 40.2

the velocity v_0 (Fig. 40.2b). According to equation (40.8), we have

$$F_z = \frac{qQ}{4\pi\epsilon_0 r^2} \sqrt{1 - v_0^2/c^2} \quad (40.10)$$

2. With respect to frame xyz the charges Q and q travel at the velocity $v = -v_0$ parallel to each other. It was found that the force of interaction between the charges is less in this reference frame than in frame $x_0y_0z_0$ with respect to which they are at rest.

After multiplying the numerator and denominator of equation (40.10) by $\sqrt{1 - v^2/c^2}$ and factoring the numerator into two terms, we obtain

$$F_z = \frac{qQ}{4\pi\epsilon_0 r^2 \sqrt{1 - v^2/c^2}} - \frac{qQv^2}{4\pi\epsilon_0 r^2 c^2 \sqrt{1 - v^2/c^2}} \quad (40.11)$$

3. The first term in this expression can be regarded as the *electric component of the sideways force*:

$$F_e = \frac{qQ}{4\pi\epsilon_0 r^2 \sqrt{1 - v^2/c^2}} = qE \quad (40.12)$$

where E is the strength of the transverse electric field in reference frame xyz . Thus

$$E = \frac{F_e}{q} = \frac{Q}{4\pi\epsilon_0 r^2 \sqrt{1-v^2/c^2}} \quad (40.13)$$

The second term in equation (40.11) can be regarded as the *magnetic component of the sideways force*:

$$F_m = \frac{qQv^2}{4\pi\epsilon_0 r^2 c^2 \sqrt{1-v^2/c^2}} \quad (40.14)$$

4. Dividing equation (40.14) by (40.12), we have the ratio of the magnetic force to the electric one:

$$\frac{F_m}{F_e} = \frac{v^2}{c^2} \quad (40.15)$$

As we can see, the magnetic component of the sideways force is considerably less than the electric component. Therefore in calculating the force of interaction between free charges we can neglect the magnetic forces as a first approximation, carrying out calculations to the formulas of electrostatics. However, when charges move along a conductor the electric forces are compensated for (Sec. 40.1) and only the magnetic force remains. This is the explanation of the magnetic interaction of conductors carrying a current, as well as other magnetic interactions in electromagnets, electric motors, etc.

40.4. MAGNETIC INDUCTION. LINES OF INDUCTION

1. Expression (40.14) can be rewritten as follows:

$$F_m = qv \frac{Qv}{4\pi\epsilon_0 r^2 c^2 \sqrt{1-v^2/c^2}} = qvB \quad (40.16)$$

where the quantity

$$B = \frac{Qv}{4\pi\epsilon_0 r^2 c^2 \sqrt{1-v^2/c^2}} = \frac{\mu_0 Qv}{4\pi r^2 \sqrt{1-v^2/c^2}} \quad (40.17)$$

is called the *magnetic induction* (magnetic field vector) and $\mu_0 = \frac{1}{\epsilon_0 c^2}$ is the *permeability constant*. Analogically to the concept of an electric field used to describe electric interaction, we now introduce the idea of a magnetic field to describe magnetic interaction. Thus, we shall assume that a *magnetic field is set up around a moving charge Q and that this field, in its turn, acts on moving charge q* .

The magnetic induction is the force characteristic of a magnetic field in the same way as the field strength vector is the one of an electric field.

2. The induction of a magnetic field is equal to the ratio of the magnetic force to the product of the charge by its velocity, on the condition that the charge is moving perpendicular to the vector of magnetic induction. Thus

$$B = \frac{F_m}{qv_{\perp}} \quad (40.18)$$

The SI unit of magnetic induction is the *tesla* (T). This is the magnetic induction of a field which acts on a charge of 1 C, moving at a velocity of 1 m/s perpendicular to the magnetic induction, with a force of 1 N. Thus

$$1 \text{ T} = \frac{\text{N-s}}{\text{C-m}} = \frac{\text{N}}{\text{A-m}} = \frac{\text{J-s}}{\text{C-m}^2} = \frac{\text{V-s}}{\text{m}^2}$$

The cgs unit for magnetic induction is the *gauss* (Gs): $1 \text{ Gs} = 10^{-4} \text{ T}$.

3. The direction of the magnetic induction is determined as follows. Let positive charge Q be moving with the velocity v ; radius vector r is perpendicular to vector v

and is directed from the charge to a point in the field. Then vector B is perpendicular to the plane in which vectors v and r lie and all three form a system that complies with the *corkscrew rule*. Thus, if we screw in a corkscrew in the direction of vector v (Fig. 40.3) and turn the handle in the direction along radius vector r , then the direction of its rotation coincides with the direction of vector B .

4. We introduced the idea of lines of force (Sec. 37.1) to help us visualize electric fields. For the same purpose we shall use *lines of induction* for representing magnetic fields. A line of induction is a curve whose tangent at each point coincides with the magnetic induction of the field.

It is evident on the basis of symmetry that the lines of induction of a charge moving in a straight line are concentric circles.

40.5. THE MAGNETIC FIELD OF A CURRENT-CARRYING CONDUCTOR

1. If a conductor carries a current, a magnetic field in its, so to say, pure form, i.e. with no electric component, is set up around the conductor. The reason why such a field is set up can be understood on the basis of the conduction mechanism.

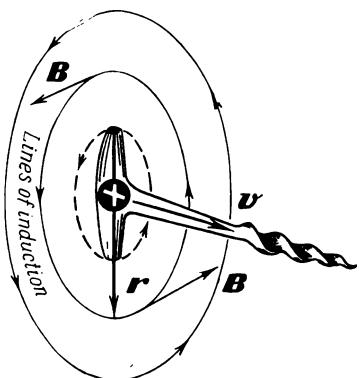


Fig. 40.3

Current in a metallic conductor is the ordered motion of free electrons. It was shown above [see equation (40.13)] that the strength of a transverse electric field of a moving charge is somewhat higher than the field of a fixed charge. But the velocity of conduction electrons is negligibly small in comparison with the velocity of light (see Sec. 39.2) so that the strength of the electric field of the conduction electrons practically compensates for the strength of the ion lattice electric field even when there is a current, in the same way as

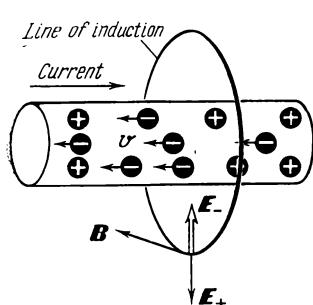


Fig. 40.4

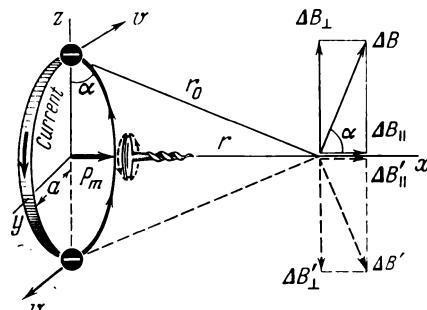


Fig. 40.5

when there is no current. Only the induction of the magnetic field set up by the conduction current is uncompensated. This is why we observe a magnetic field here in its pure form without an electric component (Fig. 40.4).

It will be left as an exercise for the student to show that in electrolytes, where current is the ordered motion of positive and negative ions in opposite directions, magnetic fields are set up by the same mechanism: the electric components of the field compensate each other and the magnetic components are added together.

2. Let us suppose that a conductor in the form of a ring of radius a has a current i (Fig. 40.5). The centre of the ring is at the origin of coordinates and the conductor lies in plane yz with the axis of abscissas as its axis of symmetry. We shall calculate the magnetic induction at an arbitrary point on this axis.

The magnetic induction of a moving electron can be found by formula (40.17). Putting $Q = e$ (charge of the electron) and taking into account the fact that the velocity of a conduction electron is much less than that of light in a vacuum, we have

$$\Delta B = \frac{\mu_0 ev}{4\pi r_0^2} \quad (40.19)$$

3. The induction vector ΔB is perpendicular to radius r_0 . Let us resolve the induction vector into one along the axis of symmetry, $\Delta B_{||} = \Delta B \cos \alpha$, and one perpendicular to the axis, $\Delta B_{\perp} = \Delta B \sin \alpha$ (Fig. 40.5). Each electron will have an electron located symmetrical-

ly at the opposite end of the diameter they are on and moving in the opposite direction. We see from considerations of symmetry that

$$\Delta B'_{||} = \Delta B_{||} \text{ and } \Delta B'_{\perp} = -\Delta B_{\perp}$$

Next we add together the induction vectors of the fields set up by all the conduction electrons. It is clear that the *sum of all the transverse projections of the induction vector equals zero*. Only the *longitudinal component of the induction vector* remains; it is directed along the axis of symmetry:

$$B_{||} = N\Delta B_{||} = N\Delta B \cos \alpha \quad (40.20)$$

where N is the number of conduction electrons in the conductor.

4. Let the cross-sectional area of the conductor be A . Then its volume is $V = 2\pi aA$ and it contains $N = nV = 2\pi n a A$ electrons. Since $\cos \alpha = a/r_0$, we can rewrite expression (40.20) to obtain

$$B_{||} = \frac{\mu_0 \times envA \times 2\pi a^2}{4\pi r_0^3}$$

But $envA = i$ is the current in the conductor [see equation (39.17)]. Thus

$$B_{||} = \frac{2\mu_0 i \times \pi a^2}{4\pi r_0^3} \quad (40.21)$$

If the point in the field for which the induction is calculated is so far from the centre of the circular current that its distance $r = \sqrt{r_0^2 - a^2} \gg a$, it can be assumed with some degree of approximation that $r \approx r_0$. Then we can substitute the distance from the point in the field to the centre of the circular current for the distance from the point to the conductor in equation (40.21). Thus

$$B_{||} = \frac{2\mu_0 i \times \pi a^2}{4\pi r^3} \quad (40.22)$$

This formula closely resembles the expression for the strength of an electric field at a point on the axis of an electric dipole: $E_{||} = \frac{2p_e}{4\pi\epsilon_0 r^3}$ [see equation (10.10)]. Incidentally, the latter is also only valid for points of the field that are far from the centre of the dipole.

40.6. THE MAGNETIC MOMENT

1. Let us call the product of the current in a flat current-carrying loop by its enclosed area the *magnetic moment* of the loop. Thus

$$p_m = iA \quad (40.23)$$

For a circular current $p_m = \pi a^2 i$, where a is the radius of the circular loop. The SI unit for magnetic moment is the $\text{A}\cdot\text{m}^2$.

The magnetic moment can be regarded as a vector, ascribing to it the direction of the *outward normal* to the area enclosed by the loop. This direction and, consequently, that of the magnetic moment vector, can be determined by the corkscrew rule (Maxwell's rule): if we turn the handle of a corkscrew with right-hand thread in the direction of the current, the advancing motion of the screw indicates the direction of the magnetic moment vector (see Fig. 40.5).

2. Using the concept of the magnetic moment, we can rewrite equation (40.22) in the vector form:

$$\mathbf{B}_{\parallel} = \frac{2\mu_0 \mathbf{p}_m}{4\pi r^3} \quad (40.24)$$

Thus, at great distances from a current-carrying loop the magnetic field has the same structure as the electric field of a dipole (see Sec. 10.6). It is proportional to the magnetic moment and inversely proportional to the cube of the distance from the source of the field to the point in it.

Using this analogy we shall write the expression for the induction at a point lying in the plane of the circular current, i.e. on a perpendicular to the magnetic moment vector. By analogy with equation (10.11) we have

$$\mathbf{B}_{\perp} = -\frac{\mu_0 \mathbf{p}_m}{4\pi r^3} \quad (40.25)$$

A rigorous derivation confirms this result. The minus sign indicates that the direction of vector \mathbf{B}_{\perp} is opposite to that of the magnetic moment vector.

3. We should note that the structure of the magnetic field in the vicinity of the current-carrying loop differs in principle from the structure of the electric field of a dipole. An electric dipole consists of two unlike charges that are equal in magnitude. The lines of force of this electric field begin at the positive charge and end at the negative charge. The lines of induction of the magnetic field are closed and encircle the current-carrying loop. This is quite evident in Fig. 40.6, which illustrates the electric field of a dipole and the magnetic field of a loop. It is just as evident that at appreciable distances the lines of force and of induction of the two fields are of the same shape.

Note also that at great distances from the current-carrying loop, the structure of the field no longer depends upon the size and shape of the loop. If we increase the area of the loop and proportionally reduce the current we shall obtain the same field because the magnetic moment does not change. Moreover, it was found that the shape of the loop ceases to play any role. We can replace the round loop by one with a square, triangular or any other shape without changing the

field at large distances from the loop provided the magnetic moment is maintained constant.

4. We shall now introduce the concept of the magnetic moment of a circulating charge. If a charge q moves uniformly in a circle at

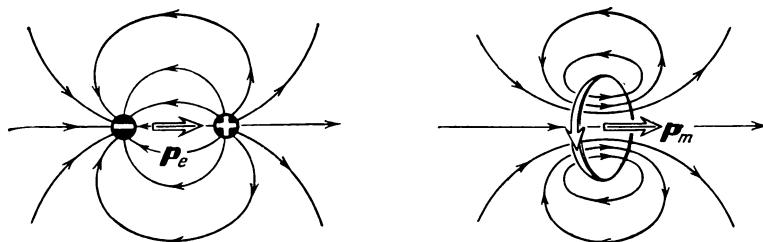


Fig. 40.6

a velocity v , it sets up a magnetic field that continuously changes in direction. The average value, however, of the induction vector on the axis will be the same as if we had a current $i = q/T$ in a circle,

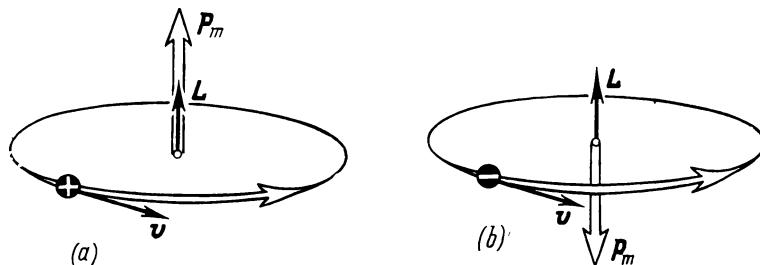


Fig. 40.7

where $T = 2\pi r/v$ is the time required for one revolution (period) of the charge. It is thus expedient to introduce the concept of the magnetic moment of a circulating charge. Thus

$$p_m = i\pi r^2 = \frac{q\pi r^2}{T} = \frac{\pi r^2 qv}{2\pi r} = \frac{qvr}{2}$$

Hence, the magnetic moment of a circulating charge is

$$p_m = \frac{qvr}{2} \quad (40.26)$$

or

$$p_m = \frac{q}{2m} mvr \quad (40.27)$$

But $L = mvr$ is the angular momentum. We see that the *magnetic moment of a circulating charge is proportional to the angular momentum of the body carrying the charge*. Thus

$$p_m = \frac{q}{2m} L \quad (40.28)$$

Both the angular momentum L and the magnetic moment p_m are vectors perpendicular to the orbital plane of the charge. If a positive charge circulates on an orbit, then vectors \mathbf{L} and \mathbf{p}_m have the same direction, which is like that of an advancing right-hand screw turned in the direction the charge revolves (Fig. 40.7a). If a negative charge (for instance, an electron) circulates on an orbit, the magnetic moment, according to equation (40.28), is directed opposite to the orbital moment (angular momentum) (Fig. 40.7b).

40.7. MAGNETIC FIELD STRENGTH

1. Besides the induction vector \mathbf{B} , which characterizes the force action of a magnetic field on a moving charge, a subsidiary vector \mathbf{H} is often used. It is called the *magnetic field strength* and is defined as the ratio of the magnetic induction in a vacuum B_0 to the permeability constant μ_0 . Thus

$$\mathbf{H} = \mathbf{B}_0 / \mu_0 \quad (40.29)$$

It should be noted that by analogy with electric field theory in which vector \mathbf{E} is called the field strength and subsidiary vector \mathbf{D} the electric displacement, it would be more suitable to call the force vector \mathbf{B} the magnetic field strength and the subsidiary vector \mathbf{H} the magnetic induction or magnetic displacement. This confusion in terminology arose before the physical sense of the field vectors \mathbf{B} and \mathbf{H} was sufficiently clear.

2. According to equations (40.21) and (40.29), the magnetic field strength on the axis of a circular current is

$$H_{||} = \frac{2\pi i a^2}{4\pi r_0^3} \quad (40.30)$$

At the centre of the circular current where $r_0 = a$ we have

$$H_c = i/2a \quad (40.31)$$

This relationship is the basis for the SI unit of magnetic field strength: the *ampere per metre* (A/m). This is the magnetic field strength at the centre of a round loop with a radius of 1 metre having a current of 2 amperes.

The unit used in the cgs system is called the *oersted* (Oe):

$$1 \text{ Oe} = \frac{1000}{4\pi} \text{ A/m} \cong 80 \text{ A/m}$$

40.8. THE MAGNETIC FIELD OF A SOLENOID

1. A *solenoid* is a cylindrical coil consisting of a great number of turns of wire wound in a close-packed helix on a form. It can be assumed that a solenoid is a system of circular currents of the same radius, which have a common axis.

The magnetic field strength on the axis of the solenoid is the sum of the strengths of the fields set up by all the circular currents

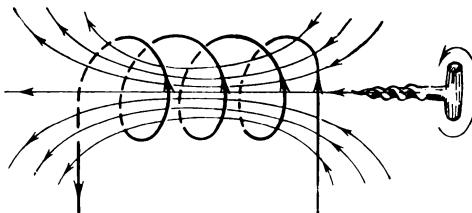


Fig. 40.8

(turns). The direction of the field strength vector (and magnetic induction vector) are determined by the corkscrew rule as illustrated in Fig. 40.8.

2. Let a solenoid consist of w turns of wire and have a length l . The density of the winding is $n = w/l$, the number of turns in unit length of the solenoid. Assume that the length of the solenoid is much greater than its diameter $l \gg a$. In this case the magnetic field inside the solenoid can be assumed uniform, i.e. its strength (and induction) are practically the same at all points. Thus

$$H = in \quad (40.32)$$

The field strength decreases near the ends of the solenoid and at the very ends is only one half that inside the coil:

$$H_{end} = in/2 \quad (40.33)$$

40.9. INVARIANCE OF THE ELECTRIC CHARGE

1. When we analyzed the nature of magnetic interaction we assumed that an electric charge has the same value in all frames of reference, i.e. the charge is invariant. Let us clear up the grounds for this conclusion.

If we assume that the charge is not invariant and changes with the velocity of the particle then, by analogy, we should surmise that the charge is expressed by an equation similar to those for length, time, mass, etc. (see Sec. 13.3). Thus

$$|\Delta q/q| \cong 1 - \sqrt{1 - v^2/c^2} \cong v^2/2c^2 \quad (40.34)$$

But the velocity of the electron in the atom (see Sec. 16.7) is about 10^6 m/s, hence

$$|\Delta q/q| \cong \frac{10^{12}}{2 \times 9 \times 10^{16}} \cong 10^{-5}$$

2. On the face of it this would seem to be a very small quantity that is difficult to detect. But a more careful analysis shows that this could have been detected.

As a matter of fact, the energy of electrons changes by several times when temperature changes or in chemical reactions (see Secs. 20.3 and 20.4). The velocity of the electrons changes accordingly. But the velocity of the nuclei (and their charges) remains unchanged. It follows, if our first assumption had been true, that a change in the velocity of the electron would violate the neutrality of the atom. This means that in one state the positive charge of the nucleus would be compensated by the negative charge of the electron shell while in some other state, when the velocity of the electrons is changed, this compensation would be violated and the atom would become charged. Though the change in the charge of one atom would be small, a macroscopic volume of gas would acquire a high charge.

Let us consider 1 cm³ of gas as an example. It contains $N = 2.7 \times 10^{19}$ atoms under standard conditions (see Sec. 26.9). According to equation (40.34), this gas, when heated, would acquire a surplus charge of

$$\Delta Q = N \Delta q \cong \frac{N e v^2}{2 c^2} \cong \frac{2.7 \times 10^{19} \times 1.6 \times 10^{-19} \times 10^{12}}{2 \times 9 \times 10^{16}} \cong 2 \times 10^{-5} \text{ C}$$

This would be accompanied by the appearance of an electric field with the potential

$$\varphi \cong \frac{\Delta Q}{4\pi e_0 r} \cong \frac{2 \times 10^{-5} \times 36\pi \times 10^9}{4\pi \times 10^{-2}} \cong 2 \times 10^7 \text{ V}$$

But an electric field with a potential of tens of millions of volts would be readily detected! However, nothing of the kind has been observed in experiments.

3. Therefore, the assumption that the electric charge is not invariant and changes with the velocity of the particle according to equation (40.34) leads to absurd consequences. Hence, the assumption is false.

The fact that the neutrality of the atom is not violated in either thermal processes or chemical reactions is conclusive proof of the invariance of the electric charge.

Chapter 41

CHARGES AND CURRENTS IN A MAGNETIC FIELD

41.1. THE LORENTZ FORCE

1. It was shown in the preceding chapter that a charge moving with a velocity v perpendicular to the induction vector \mathbf{B} is subject to the force $F = qvB$ [see equation (40.18)]. This is called the *Lorentz force*. The direction of the Lorentz force can be determined by means

of the *left-hand rule*: if the open left hand is placed so that vector \mathbf{B} points into the palm and vector \mathbf{v} is along the fingers, then the outspread thumb will show the direction of the force exerted by the magnetic field on a positive charge (Fig. 41.1). The force exerted on a negative charge moving in the same direction in the same magnetic field will be in the opposite direction.

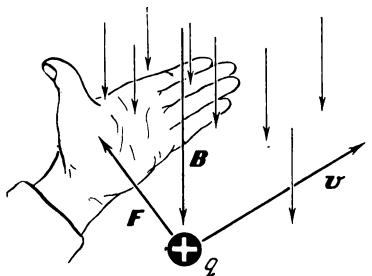


Fig. 41.1

Experiments show, also, that if the charge moves along the lines of induction, i.e. if the velocity and magnetic induction vectors are parallel or antiparallel, the magnetic field has no effect on the charge.

Hence, if the velocity vector of the charge makes an angle $\alpha \neq \pm\pi/2$ with the induction vector \mathbf{B} , the velocity vector can be resolved into two components: a longitudinal one $v_{||} = v \cos \alpha$ and a transverse one $v_{\perp} = v \sin \alpha$. The Lorentz force is determined by only the *transverse component of the velocity*:

$$\mathbf{F} = qv_{\perp}\mathbf{B} = qvB \sin \alpha \quad (41.1)$$

2. The expression for the Lorentz force can be written in the vector form if we introduce the concept of the *unit vector* \mathbf{n}_0 . The magnitude of the unit vector equals unity: $|\mathbf{n}_0| = 1$. Therefore, the magnitude of a scalar is not changed when it is multiplied by a unit vector. We shall chose the direction of the unit vector so that it is perpendicular to vectors \mathbf{v} and \mathbf{B} according to the left-hand rule.

The Lorentz force is written in vector form as follows:

$$\mathbf{F} = \mathbf{n}_0qvB \sin \alpha \quad (41.2)$$

We know that a charge is acted on by an electric field and not only by a magnetic field. Therefore, if a charge moves in an *electromagnetic field*, the force acting on it can be represented as the vector sum

of the electric and magnetic components. Thus

$$\mathbf{F} = \mathbf{F}_e + \mathbf{F}_m = q(\mathbf{E} + \mathbf{n}_0 v B \sin \alpha) \quad (41.3)$$

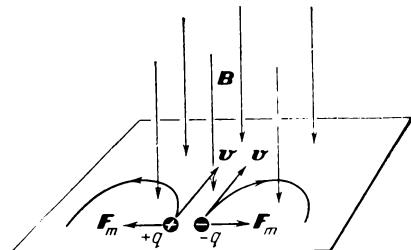
This expression defines the *generalized Lorentz force*.

3. One important difference should be noted between the electric and magnetic components of the generalized Lorentz force. As shown in Sec. 18.7 the work done in moving a charge in an electric field equals the product of the charge by the potential difference: $W = q(\varphi_1 - \varphi_2)$. The *magnetic force*, i.e. the Lorentz force in the narrow sense, *does no work*. Indeed, an element of work is $\Delta W = F \Delta l \cos \beta$, where β is the angle between the direction of the force and the direction of the displacement. But the Lorentz force is perpendicular to the motion of the charge ($\mathbf{F}_m \perp \mathbf{v}$). Hence, the angle $\beta = \pi/2$ and the work $\Delta W = 0$.

41.2. THE MOTION OF CHARGED PARTICLES IN A UNIFORM MAGNETIC FIELD

1. Suppose that a particle of charge q and rest mass m_0 is introduced with a velocity \mathbf{v} into a uniform magnetic field ($\mathbf{B} = \text{const}$) perpendicular to its lines of induction (Fig. 41.2). For concreteness we assume that the particle carries a positive charge. The particle is subject in the field to a Lorentz force that is perpendicular to its velocity vector. Therefore, the motion equation of the particle can be written as

$$\left. \begin{aligned} F_t &= \frac{\Delta(mv)}{\Delta t} = 0 \\ F_n &= \frac{mv^2}{r} = qvB \end{aligned} \right\} \quad (41.4)$$



The first equation indicates that the magnitude of the momentum does not change: $p = mv = \text{const}$. Since $p = \frac{m_0 v}{\sqrt{1 - v^2/c^2}}$, we come to the conclusion that *the magnitude of the velocity of particles does not change in a uniform magnetic field*.

We could have arrived at the same result in another way—from the law of conservation of energy. It was shown in the preceding section that magnetic forces perform no work. Therefore, the kinetic energy of the particle does not change in a uniform magnetic field. This indicates that the magnitude of the velocity remains constant.

2. The second equation of (41.4) shows that a normal (centripetal) force acts on the particle. Consequently, the direction of its velo-

Fig. 41.2

city will continuously change and it will move along a curvilinear path.

We shall show that this path is a circle. To do this we shall find the radius of curvature of the path:

$$r = \frac{mv}{qB} = \frac{p}{qB} \quad (41.5)$$

Since both the charge and magnetic induction are constant, and the magnitude of the momentum, as shown above, is also constant, $r = \text{const}$. But the locus of points located at a constant distance from a centre is a circle, i.e. a line of constant curvature ($r = \text{const}$). Thus, a *charged particle, introduced into a uniform magnetic field perpendicular to the lines of induction, moves at uniform velocity in a circle in a plane which is perpendicular to the magnetic induction vector.*

3. Let us calculate the period of revolution of the particle on this path:

$$T = \frac{2\pi r}{v} = \frac{2\pi m}{qB} = \frac{2\pi m_0}{qB \sqrt{1 - v^2/c^2}} \quad (41.6)$$

We see that at nonrelativistic velocities ($v \ll c$) the period of revolution of the particle does not depend on its velocity or the radius of the circle. But if the velocity of the particle approaches that of light (relativistic and ultrarelativistic velocities) the period increases with the velocity.

It should be pointed out that a test of the validity of equation (41.6) also tests the validity of one of the basic principles of the theory of relativity—the dependence of mass on velocity: $m = \frac{m_0}{\sqrt{1 - v^2/c^2}}$.

We gave this formula in Sec. 13.1 without proof. Numerous experiments that investigated the motion of charged particles in magnetic fields have checked formula (41.6) to a great degree of accuracy, thereby demonstrating the dependence of mass on velocity.

41.3. DETERMINING THE SIGN OF THE CHARGE OF ELEMENTARY PARTICLES

1. A negatively charged particle having the same velocity, mass and magnitude of charge as a positively charged particle will move along the same path but with only one difference. It will be turned in the opposite direction (Fig. 41.2). Hence, we can determine the sign of the charge of a particle by observing the direction in which it turns in a magnetic field.

2. This is how the *positron* (or antielectron) was discovered. This particle has the same mass as the electron but has a positive charge.

The positron was discovered in 1932 by C. D. Anderson (b. 1905) by the tracks it made in a Wilson cloud chamber (see Sec. 36.9), which was placed in a magnetic field. The photograph in Fig. 41.3 shows the track of a positron passing through a lead plate P located in the centre of the chamber. The magnetic field is directed perpendicular to the page toward the reader.

The direction of motion of the particle is determined from the curvature of the path. The path has a greater curvature below the

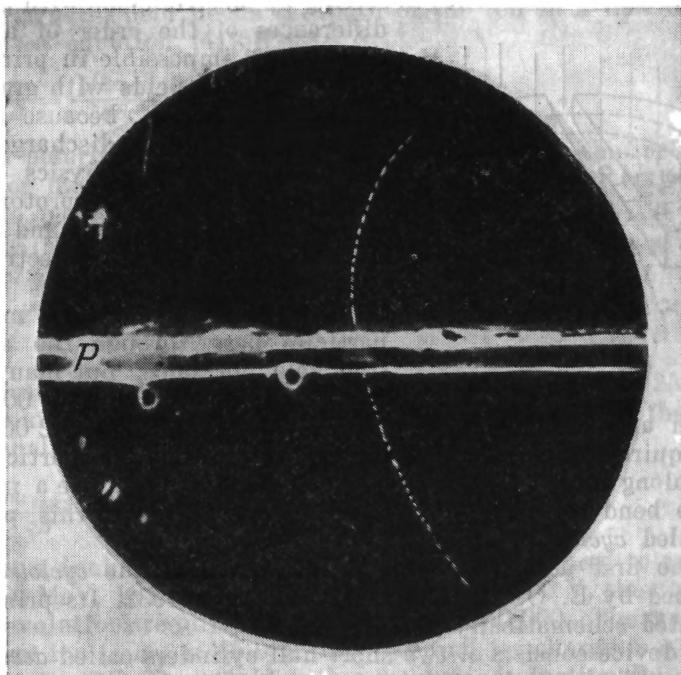


Fig. 41.3

plate than above it and, therefore, its velocity is higher below the plate. This indicates that the particle is moving upward and loses some of its energy when it passes through the lead. Determining the direction of the velocity vector in this way and knowing the direction of the magnetic induction vector, we can find the sign of the charge of the particle. It turned out to be positive in our case. The charges of other elementary particles are determined in a similar way from the photographs of the tracks they leave in a Wilson cloud chamber or in a bubble chamber.

41.4. THE CYCLOTRON

1. Many investigations in nuclear physics require beams of particles (electrons, protons or positive ions) with very high energy. Particles are accelerated to such high energies by *accelerators*.

The main element of an accelerator is an electric field with a substantial potential difference. The field is what accelerates the particles by imparting to them the kinetic energy $K = q\varphi$, where φ is the potential difference and q is the charge of the particle. But there are tremendous difficulties in setting up electric fields with potential differences of the order of a million volts.

It is impossible in principle to obtain electric fields with greater potential differences because of the occurrence of gas discharges. But research in nuclear physics requires particles (for instance, protons) with energies of ten thousands and hundred thousands of millions of electron volts (10 to 100 GeV).

The only solution is to make the particles pass through the accelerating field not once, but many times. Thus, if a proton passes 30 000 times

through an accelerating field with a potential of 22 000 V, it will acquire an energy of 660 MeV. To do this the particle must move along a closed path or a spiral. This will require a magnetic field to bend its path. Accelerators operating on this principle are called *cyclic*, or *circular*, accelerators.

2. The first accelerator of this type, called the *cyclotron*, was developed by E. O. Lawrence (1901-1958) in 1932. Its principle is illustrated schematically in Fig. 41.4.

This device consists of two short half cylinders called *dees* (owing to their "D" shape) located in a vacuum chamber. Near the centre of the chamber is a source of positive ions, which are injected into the gap between the dees. Powerful pumps maintain a high vacuum in the chamber to avoid collisions between the moving ions and molecules of air. The pressure is about one thousand millionth of standard atmospheric pressure (from 10^{-5} to 10^{-6} mm-Hg). The high-frequency alternating potential difference applied over the dees has a frequency of the order of 1 to 10 MHz [1 megahertz (MHz) = 10^6 cycles per s]. The average potential difference varies from several hundreds to several tens of thousands of volts.

The vacuum chamber is placed between the poles of a powerful electromagnet, which sets up a magnetic field with an induction of 1.3 to 1.6 T. The steel core of the electromagnet weighs several

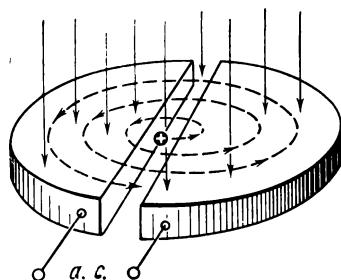


Fig. 41.4

thousand tons, the windings are supplied from generators rated at several megawatts.

3. The cyclotron operates as follows. In the gap between the dees a positive ion acquires the kinetic energy $K_1 = p_1^2/2m = q\varphi$, where φ is the potential difference between the dees and q is the charge of the ion. The momentum of the ion after a single passage through the field is

$$p_1 = \sqrt{2mq\varphi}$$

The magnetic field bends the path of the ion in a half circle of radius [see equation (41.5)]

$$r_1 = \frac{p_1}{qB} = \frac{\sqrt{2mq\varphi}}{qB} = \sqrt{\frac{m}{q}} \times \frac{\sqrt{2\varphi}}{B}$$

If the polarity of the electric field is reversed at the instant the ion, after making a half turn, again reaches the gap between the dees, the particle will acquire a new portion of energy. Its kinetic energy will become $K_2 = 2q\varphi$, its momentum $p_2 = \sqrt{4mq\varphi}$ and the radius of the next half turn will be

$$r_2 = \sqrt{\frac{m}{q}} \times \frac{\sqrt{4\varphi}}{B}$$

In N turns the ion will pass through the accelerating field $2N$ times. Its final kinetic energy will therefore be $K = 2Nq\varphi$ and the radius of its final orbit (dee radius) will be

$$R = \frac{mv_{\max}}{qB} = \frac{\sqrt{2mK}}{qB} = \sqrt{\frac{m}{q}} \times \frac{\sqrt{2N\varphi}}{B} \quad (41.7)$$

4. Let a doubly ionized beryllium ion be accelerated to an energy $K = 100$ MeV. If the accelerating potential is 500 V, then the number of revolutions required is $N = K/2q\varphi = 50\,000$. The mass of the beryllium ion is $m_0 = 15.0 \times 10^{-27}$ kg and its charge is $q = 2e = 3.2 \times 10^{-19}$ C. For a field with a magnetic induction $B = 1.53$ T, the dee radius should be

$$R = \frac{\sqrt{2 \times 15 \times 10^{-27} \times 100 \times 1.6 \times 10^{-13}}}{1.53 \times 3.2 \times 10^{-19}} = 1.4 \text{ m}$$

Note that the final velocity of the ion will be

$$v = \sqrt{\frac{2K}{m}} = \sqrt{\frac{2 \times 100 \times 1.6 \times 10^{-13}}{15 \times 10^{-27}}} = 4.6 \times 10^7 \text{ m/s}$$

which is about 15 per cent of the velocity of light in a vacuum.

5. For an ion to accelerate continuously it must pass through the gap between the dees at the instant the electric field is reversed. Therefore, the period with which the accelerating electric field changes

polarity must coincide with the period of revolution of the ion. This is called the *condition of synchronization*. According to equation (41.6), this condition can be expressed as

$$T_{\text{field}} = T_{\text{ion}} = \frac{2\pi m}{qB} \quad (41.8)$$

For the case of the beryllium ion we have

$$T = \frac{2\pi \times 15 \times 10^{-27} \text{ J}}{3.2 \times 10^{-19} \times 1.53} = 1.93 \times 10^{-7} \text{ s} \cong 0.2 \mu\text{s} \text{ (microsecond)}$$

The field polarity is to be reversed with the frequency

$$\nu = \frac{1}{T} = \frac{10^7}{1.93} = 5.2 \times 10^6 \text{ Hz} = 5.2 \text{ MHz}$$

41.5. ENERGY OF THE PARTICLE AND THE SYNCHRONIZATION CONDITION

1. In the preceding section we did not take into account the dependence of the mass of the ion on its velocity, i.e. we assumed that the ion travels with a velocity much less than that of light in a vacuum. Then the period of revolution of the ion in the magnetic field is constant and the equal period with which the accelerating electric field is reversed is also constant. Denoting this period by T_0 , we have

$$T_0 = \frac{2\pi m_0}{qB} \quad (41.9)$$

2. If the velocity of the accelerated particle approaches that of light in a vacuum, the mass of the particle will increase. Consequently, its period of revolution will likewise increase.

We divide equation (41.6) by equation (41.9) and take into consideration that

$$\frac{m}{m_0} = \frac{mc^2}{m_0c^2} = \frac{\mathcal{E}}{\mathcal{E}_0} = \frac{\mathcal{E}_0 + K}{\mathcal{E}_0}$$

where \mathcal{E} = total energy of the particle

\mathcal{E}_0 = its rest energy

K = its kinetic energy.

Then

$$\frac{T}{T_0} = \frac{m}{m_0} = \frac{\mathcal{E}_0 + K}{\mathcal{E}_0} = 1 + \frac{K}{\mathcal{E}_0}$$

From which

$$T = T_0 \left(1 + \frac{K}{\mathcal{E}_0} \right) = \frac{2\pi m_0}{qB} \left(1 + \frac{K}{\mathcal{E}_0} \right) \quad (41.10)$$

Thus, the period of revolution of the particle is appreciably changed only when its kinetic energy becomes commensurate with its rest energy.

In the problem dealt with above concerning the ion of beryllium the rest energy of the ion is

$$\mathcal{E}_0 = m_0 c^2 = 15 \times 10^{-27} \times 9 \times 10^{16} \text{ J} = \frac{15 \times 10^{-27} \times 9 \times 10^{16}}{1.6 \times 10^{-13}}$$

$$= 8400 \text{ MeV}$$

and the kinetic energy at the end of the acceleration cycle is $K = 100 \text{ MeV}$. Therefore, the period of revolution of the particle increases by about 1.2 per cent, which has almost no effect on the condition of synchronization. However, if a proton, whose rest mass (and rest energy) is only $1/8.94$ as much as the mass of the beryllium ion, is accelerated to the same kinetic energy, its period of revolution will increase by almost 11 per cent. This will drastically violate the condition of synchronization, the particles will get out of step with the field and the acceleration process will cease.

3. An attempt to accelerate protons in a cyclotron to the energy $K = 1 \text{ GeV}$ must fail. Indeed, the rest energy of the proton is

$$\mathcal{E}_0 = 1.67 \times 10^{-27} \text{ kg} \times 9 \times 10^{16} \text{ m}^2/\text{s}^2$$

$$= \frac{1.67 \times 10^{-27} \times 9 \times 10^{16}}{1.6 \times 10^{-10}} = 0.938 \text{ GeV}$$

We find, therefore, that $K \approx \mathcal{E}_0$ and $T \approx 2T_0$, i.e. the period of revolution of the proton at such energies is almost doubled in comparison with that of the nonrelativistic proton. If the period of reversal of the accelerating electric field is maintained constant, the protons will not reach the accelerating field (in the gap) at the required instant. This means that they will drop out of synchronization and cease to accelerate further.

4. Hence, a cyclotron is capable of accelerating only nonrelativistic particles, with a kinetic energy much less than their rest energy or, what is the same, a velocity much less than that of light. A cyclotron is quite unsuitable for obtaining relativistic and, all the more, ultrarelativistic particles.

The cyclotron is entirely incapable of accelerating electrons, whose rest energy is $\mathcal{E}_0 = 9 \times 10^{-31} \text{ kg} \times 9 \times 10^{16} \text{ m}^2/\text{s}^2 = 0.51 \text{ MeV}$. Even at low energies, less than 1 MeV, the electron already becomes a relativistic particle and drops out of synchronization. This is why cyclotrons can be used only for accelerating relatively heavy ions. An advantageous feature of the cyclotron is that it has a continuous mode of operation and can simultaneously accelerate many particles.

41.6. PROTON SYNCHROTRON

1. Relativistic particles with kinetic energies of tens of gigaelectron volts are produced today by installations called *synchrotrons*. A proton synchrotron which accelerates protons to a kinetic energy of 10 GeV was built in 1957 in the USSR at the Joint Institute for Nuclear Research in the city of Dubna. Since then synchrotrons have been developed in the USSR, USA and other countries. Protons are accelerated in them to higher and higher kinetic energies. The largest in the USSR to date is the proton synchrotron built near the city of Serpukhov.

2. Let us calculate the maximum momentum and the corresponding maximum radius of the orbit of the proton in the Serpukhov synchrotron. From the relationship (16.8) between the energy and momentum of a particle we have

$$p = mv = \frac{1}{c} \sqrt{\mathcal{E}^2 - \mathcal{E}_0^2} = \frac{1}{c} \sqrt{K(K + 2\mathcal{E}_0)}$$

The rest energy of the proton is $\mathcal{E}_0 = 0.938$ GeV and its kinetic energy in the Serpukhov synchrotron reaches $K = 76$ GeV. With an error within 1.5 per cent we can assume that the kinetic energy equals the total energy, i.e. $\mathcal{E} \cong K$, and that the momentum is $p \cong K/c$. Then the maximum orbit radius, according to equation (41.5), equals

$$R = \frac{mv_{\max}}{qB} = \frac{\sqrt{K(K + 2\mathcal{E}_0)}}{qBc} \cong \frac{K}{qBc} \quad (41.11)$$

With a magnetic induction of the field in the gap of $B = 1.07$ T we obtain

$$R = \frac{7.6 \times 10^{10} \times 1.6 \times 10^{-19}}{1.07 \times 1.6 \times 10^{-19} \times 3 \times 10^8} \cong 237 \text{ m}$$

It is practically impossible to build a cylindrical magnet of the cyclotron type with a diameter of about one half kilometre. This is why the magnetic field of the synchrotron is set up in a relatively narrow annular (ring-shaped) chamber and the *protons being accelerated travel around a circle of constant radius*. The vacuum chamber of the proton synchrotron in Dubna has a width of about one half metre. Owing to the new method of focusing the proton beam applied in Serpukhov, the width of the chamber is less than 10 cm.

The following consequences result from this feature of proton acceleration in the synchrotron.

3. In the first place, the synchrotron, in contrast to the cyclotron, cannot begin the proton acceleration process "from zero". The protons introduced into its chamber must already be travelling at a relativistic velocity. Protons entering the Dubna synchrotron are first acce-

Ierated by a linear accelerator to an energy of 9 MeV. At Serpukhov the initial energy of the protons is even higher: 100 MeV.

Secondly, in order to maintain a constant orbital radius when the energy of the particles increases, the magnetic induction of the field should be increased with the increase in momentum according to equation (41.5) so that their ratio remains constant. This is the essential difference between the synchrotron and the cyclotron, where the magnetic induction of the field is maintained constant while the particle is being accelerated.

Finally, the period of reversal of the accelerating electric field must also be varied accordingly. The period of revolution of the particle, as we know, is $T = 2\pi R/v$. But since the orbital radius is constant in the synchrotron, the period of revolution of the particle decreases with an increase in its velocity. The period of reversal of the accelerating electric field should be proportionally reduced.

4. Let us calculate certain characteristic parameters of the synchrotron. The initial kinetic energy of the proton is $K_1 = 9$ MeV; its final energy is $K_2 = 10$ GeV. The maximum magnetic induction of the field is $B_{\max} = 1.26$ T in the annular vacuum chamber.

According to equation (41.11), the orbital radius is

$$R = \frac{K_2}{ceB_{\max}} = \frac{10^{10} \times 1.6 \times 10^{-19}}{1.6 \times 10^{-19} \times 1.26 \times 3 \times 10^8} = 26.6 \text{ m}$$

The period of revolution of the particle is

$$T = \frac{2\pi R}{v} = \frac{2\pi Rm}{p} = \frac{2\pi R\mathcal{E}}{pc^2} \quad (41.12)$$

But the momentum is $p = \frac{1}{c} \sqrt{\mathcal{E}^2 - \mathcal{E}_0^2}$ and therefore

$$T = \frac{2\pi R\mathcal{E}}{c \sqrt{\mathcal{E}^2 - \mathcal{E}_0^2}} = \frac{2\pi R}{c \sqrt{1 - \mathcal{E}_0^2/\mathcal{E}^2}} \quad (41.13)$$

The initial energy is $\mathcal{E}_1 = \mathcal{E}_0 + K_1 = 938 + 9 = 947$ MeV, and the initial period is

$$T_1 = \frac{2\pi \times 26.6}{3 \times 10^8 \times \sqrt{1 - (938/947)^2}} = 4.05 \times 10^{-8} \text{ s} = 4.05 \mu\text{s}$$

The corresponding value for the final energy is $\mathcal{E}_2 = \mathcal{E}_0 + K_2 = 938 + 10 = 10.938$ GeV, and for the period

$$T_2 = \frac{2\pi \times 26.6}{3 \times 10^8 \times \sqrt{1 - (938/10.938)^2}} = 0.57 \times 10^{-8} \text{ s} = 0.57 \mu\text{s}$$

Thus, the frequency of reversal of the accelerating potential differences of the electric field should vary from 0.247 to 1.75 MHz.

In one revolution the proton passes once through the accelerating electric field having a maximum potential of 2.2 kV. For the proton to acquire a kinetic energy of 10 GeV, it must make $N = 10^{10}/2.2 \times \times 10^3 = 4.5 \times 10^6$ revolutions. The whole acceleration cycle takes 3.3 s.

5. In contrast to the cyclotron, having a continuous mode of operation, the synchrotron operates intermittently. The point is that ions can be continuously introduced into the cyclotron chamber, and they will be continuously accelerated because the magnetic induction of the field and the period of accelerating electric field reversal are maintained constant throughout the accelerating process.

Particles with a definite energy can be admitted into the synchrotron chamber only at the instant the magnetic field has its minimum value and the period of reversal of the accelerating potential is maximum. If particles with a given energy are introduced after the magnetic induction has increased and the period of reversal has been reduced, the particles will not satisfy the conditions of synchronization and will therefore fail to be accelerated.

Hence, only a definite portion (bunch) of particles are accelerated in the synchrotron. A new portion cannot be admitted into the chamber until the preceding portion has been accelerated.

6. A problem of prime importance is the production in the synchrotron of a stable bunch of particles which does not break up during acceleration. The possibility of obtaining such a stable bunch is based on the theory of phase stability of high-energy accelerators. This theory was first developed in 1944 in the USSR by V. I. Veksler (1907-1966) and independently in 1945 in the USA by E. M. McMillan (b. 1907). The essence of the principle is the following.

If only particles whose frequency *exactly* coincides with the frequency of field reversal were accelerated in the synchrotron, there would be no point in building the installation because the number of such equilibrium particles is negligibly small. Veksler showed that a bunch of particles will be accelerated if they approximately satisfy the synchronization condition $T_{\text{field}} \approx T_{\text{ion}}$. These particles may be accelerated faster than the equilibrium ones and overtake them, after which they are retarded by the field. When their velocity is reduced to somewhat less than that of the equilibrium particles, the field begins to accelerate them again. As a result the particles in a bunch are accelerated but their energy does not increase monotonically, it fluctuates about the value of the energy of the equilibrium particles. The number of such particles in a bunch is several orders of magnitude greater than the number of equilibrium particles.

41.7. THE ELECTRON CHARGE-TO-MASS RATIO

1. The *charge-to-mass ratio* of a particle is the quantity e/m_0 , i.e. the ratio of the charge of the particle to its rest mass. The charge of a particle can be determined with great precision. When we know the charge and the charge-to-mass ratio we can determine the rest mass.

The electric charge of any elementary particle, as can be shown by experiment, is always equal in magnitude to the charge of the

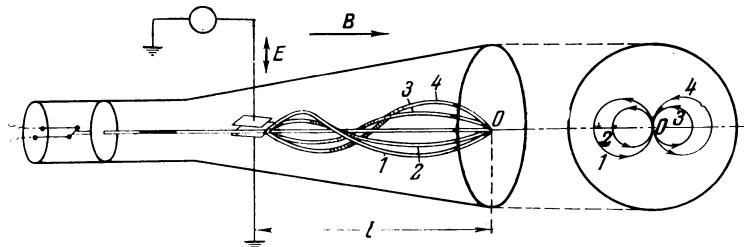


Fig. 41.5

electron (or is equal to zero). The charge of any ion is a multiple of the electronic charge.

2. Many methods are used to determine the electron charge-to-mass ratio. Let us consider one of them, the method of focusing in a longitudinal magnetic field.

The apparatus consists of an oscilloscope tube located inside a long solenoid that sets up a sufficiently uniform magnetic field with an induction B (Fig. 41.5). Electrons emitted from the cathode are accelerated by the electric field of the anode. The potential difference of this field is φ . We find the velocity of the electron from the condition $e\varphi = m_e v_0^2/2$, where m_e is the electron rest mass.

3. Let us resolve the velocity vector of the electron into two components: the transverse component $v_\perp = v \sin \alpha$ and the longitudinal component $v_{||} = v \cos \alpha$. The magnetic field has no effect on the longitudinal component; therefore, the path traversed by the electron along the field

$$l = v_{||} t = vt \cos \alpha$$

The velocity component v_\perp is subject to the Lorentz force, which, if there were no longitudinal component $v_{||}$, would make the electron move in a circle. As a result, the motion of the electron will be along a helix. At small angles ($\cos \alpha \approx 1$) the particles will be focused practically at a single point. Indeed, the electrons are focussed at one point if in traversing the distance $l \approx vt$ they manage to make one

or more whole revolutions, i.e. if $t = nT_0 = 2\pi m_e n/eB$ (where $n = 1, 2, 3, \dots$). At minimum induction the electron will make one revolution and, therefore, $l = 2\pi v m_e / eB$. Substituting $\sqrt{2e\varphi/m_e}$ for v , we have

$$\frac{e}{m_e} = \frac{8\pi^2\varphi}{B^2 l^2} \quad (41.14)$$

The anode potential, magnetic induction of the field and the distance between the deflecting electrodes and the screen can be readily determined by experiment or by calculations.

4. After careful analysis of all the methods of measuring the electron charge-to-mass ratio, the best experimental value is taken as

$$e/m_e = (1.7588028 \pm 0.0000054) \times 10^{11} \text{ C/kg}$$

from which the rest mass of the electron is

$$m_e = (9.109558 \pm 0.000054) \times 10^{-31} \text{ kg}$$

41.8. THE ION CHARGE-TO-MASS RATIO

1. Considerable difficulties are encountered in determining the charge-to-mass ratio of an ion because an ion beam contains particles with widely differing velocities that are, consequently, very hard to focus. It is necessary to separate ions of the same velocity out of the beam. This is accomplished with a *velocity filter*. In this device (Fig. 41.6), the electric field strength vector is perpendicular to the induction vector of the magnetic field.

A positive ion, moving through the crossed electric and magnetic fields, is subject to two forces: the electric force $F_e = qE$, deflecting the ion downwards, and the Lorentz force $F_m = qvB$, deflecting it upwards. The ion will pass through the slit diaphragm only if the direction of its velocity is not changed. This is possible only under the condition that the electric and Lorentz forces counterbalance each other: $F_e = F_m$ or $qE = qvB$. Thus, the only ions that pass through the slit are the ones that satisfy the condition $v = E/B$. The remaining ions are filtered out.

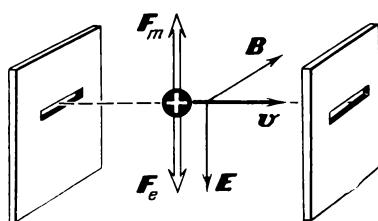


Fig. 41.6

The remaining ions are filtered out.

2. Ions having the same velocity are then focussed by means of electric or magnetic fields. In some mass spectrographs a transverse magnetic field is applied for focussing.

Ions passing out of the velocity filter move through a uniform transverse magnetic field. Their path is a half circle of radius $r = mv/qB_\perp$ [see equation (41.5)], where q/m is the charge-to-mass

ratio of the ion*, and B_{\perp} is the induction of the transverse magnetic field in the mass spectrograph. Since the velocity of the ions and the magnetic induction are controlled values, the radius of the half circle is fully determined by the charge-to-mass ratio of the ion. Beams with different ion charge-to-mass ratios are focussed at different points.

3. If we put a photographic plate at the place where the beams are focussed we obtain what is called a *mass spectrum* (for instance, Fig. 41.7). It follows from the relationship $q/m = v/rB_{\perp}$ that the deflection of any given line of the mass spectrum from the point of entry of the beam is inversely proportional to the charge-to-mass ratio of the ion.

To determine the charge-to-mass ratio of some ion it is necessary first to calibrate the mass spectrogram by means of ions with known charge-to-mass ratios. Carbon or oxygen are commonly used for this purpose.

It was found by means of a mass spectrograph that many substances are mixtures of *isotopes* (see Sec. 88.1). Mass spectrograms are shown in Fig. 41.7 for the isotopes of zinc, in which the masses of the nuclei are in the ratio 64 : 66 : 67 : 68 : 70, and also for those of germanium and tellurium.

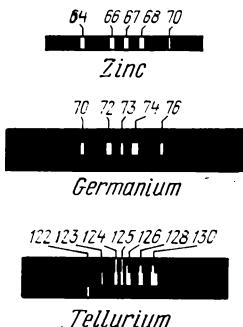


Fig. 41.7

41.9. A CURRENT-CARRYING CONDUCTOR IN A MAGNETIC FIELD

1. Experiment shows that when a conductor carrying a current is put in a magnetic field, the conductor is subject to a force whose direction is determined by the left-hand rule. This same rule is used to find the direction of the Lorentz force. The origin of this force should be quite clear. Current in a conductor is the ordered motion of electrons. The magnetic field exerts on each moving electron the force $f = evB \sin \alpha$ [see equation (41.1)]. The force exerted on the whole conductor is the vector sum of the elementary Lorentz forces exerted on each electron.

2. In the general case it is extremely difficult to calculate this force (sometimes called the *Ampere force*). We shall calculate the force exerted by a uniform magnetic field with an induction \mathbf{B} on a straight conductor with a current.

Under these conditions all the electrons are moving in the same direction and, consequently, all the elementary Lorentz forces,

* In experiments with ions $v \ll c$ and $m \approx m_0$.

exerted on the separate electrons, are directed to the same side. Hence, instead of vector addition it will be sufficient to simply add the magnitudes of all the elementary forces. If the average velocity of the electrons is \bar{v} , the length of the conductor is l and its cross-sectional area is A , the Ampere force is

$$F = Nf = nAl \times \bar{v}B \sin \alpha$$

After substituting the value of the current according to equation (39.17), we have

$$F = ilB \sin \alpha \quad (41.15)$$

where α is the angle between the conductor and magnetic induction vector. The direction of the Ampere force is determined by the left-hand rule (see Fig. 41.1); the direction of the current coincides with that of the positive charges.

41.10. A CURRENT-CARRYING LOOP IN A MAGNETIC FIELD

1. Let us place a rectangular loop carrying a current in a uniform magnetic field so that the angle between the normal to the loop and the direction of the magnetic induction vector equals α (Fig. 41.8a).

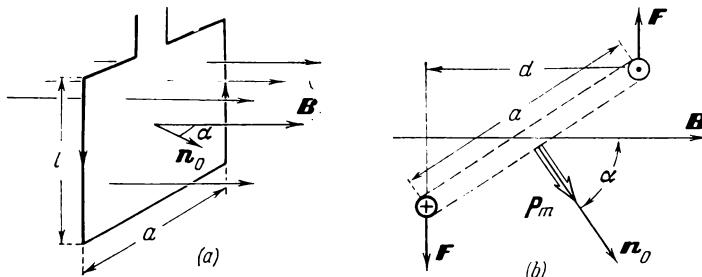


Fig. 41.8

Using the left-hand rule, we readily see that the forces exerted on the parts of the loop perpendicular to the induction vector develop a torque. In calculating this torque we shall refer to Fig. 41.8b which shows a cross section of the loop.

2. According to equation (22.14), the moment of the force couple (torque) is equal to the product of the Ampere force by the arm d . Thus

$$M = Fd = Fa \sin \alpha$$

where a is the width of the loop. Substituting the value of the Ampere force from equation (41.15), we have

$$M = ilBa \sin \alpha = iAB \sin \alpha \quad (41.16)$$

where $A = al$ is the area enclosed in the loop. According to equation (40.23), the product of the current by the area enclosed by a loop carrying the current is the magnetic moment p_m of the loop. Thus

$$M = p_m B \sin \alpha \quad (41.17)$$

3. If we compare equations (37.10) and (41.17), we shall see they are analogical. It was shown in Sec. 40.6 that a current-carrying loop sets up a magnetic field around itself that is similar to the electric field set up by a dipole.

We see that there is a striking analogy between the properties of an electric dipole $p_e = ql$ and a magnetic dipole $p_m = iAn_0$, i.e. a current-carrying loop. On this basis, we can write without proof, only by analogy with equation (37.12), the expression for the potential energy of a "magnetic dipole":

$$\mathcal{E}_{pot} = -p_m B \cos \alpha \quad (41.18)$$

On the basis of the same analogy we can conclude that in a nonuniform magnetic field a current-carrying loop is pulled into a region with a stronger field. Experiments confirm this assumption. According to equation (37.14), the force of attraction is

$$F = p_m \frac{\Delta B}{\Delta l} \quad (41.19)$$

It is proportional to the magnetic moment of the loop and to the gradient of magnetic induction of the nonuniform field.

Chapter 42

MAGNETIC MATERIALS

42.1. THREE TYPES OF MAGNETIC MATERIALS

1. The student knows that a substantial force of attraction is exerted on a piece of iron placed near a coil with a current. It is likewise known that the interaction force between two current-carrying coils is considerably increased if an iron core is inserted in them. Besides iron similar properties are found in various grades of steel, nickel, cobalt and certain special alloys: Permalloy, Alni, Magnico, etc. These substances are called ferromagnetic materials (from the Latin *ferrum* meaning iron).

In the middle of last century Faraday showed that these substances possess magnetic properties but that the degree and nature of their interaction with a field differ for different materials. Distinction

is made, in this connection, between paramagnetism, diamagnetism and ferromagnetism.

As to the intensity of their interaction with a magnetic field all substances can be divided into two classes. All diamagnetic and paramagnetic materials are weakly magnetic. At any temperature these substances interact weakly with a magnetic field. Strongly magnetic substances include ferromagnetic and antiferromagnetic materials, which at definite temperatures interact intensely with a field.

2. Paramagnetic materials are substances that are drawn into a region with a stronger field. If a paramagnetic liquid is poured

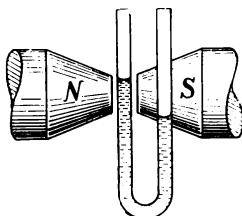


Fig. 42.1



Fig. 42.2

into a U-shaped glass tube, one leg of which is placed between the poles of a powerful electromagnet, the level of the liquid rises in this leg (Fig. 42.1). A small rod of paramagnetic material, suspended by a thin thread, aligns itself in a magnetic field along the lines of induction.

The properties of paramagnetic materials resemble those of ferromagnetic materials but the force of their interaction with a magnetic field is only one hundredth or one thousandth as much. This is why strong fields are required to observe paramagnetic effects. Paramagnetism is possessed by platinum, aluminium, tungsten and all the alkali and alkaline-earth metals. Paramagnetic gases are oxygen and, to a weak degree, nitric oxide.

3. Diamagnetic materials are substances that are repelled by a magnetic field. If we put a lighted candle between the poles of an electromagnet, its flame is repelled by the field (Fig. 42.2). A small rod of a diamagnetic material, suspended from a thread in a strong magnetic field, turns to a position perpendicular to the lines of induction (across the field).

We should note that the forces acting on a diamagnetic material in a magnetic field are one or two orders of magnitude less than those acting on a paramagnetic material in the same field. That is why exceptionally strong fields are required to demonstrate diamagnetic effects. Almost all the gases (except oxygen) are diamagnetic and also water, silver, gold, copper, diamond, graphite, bismuth and many organic compounds.

42.2. THE MAGNETIC MOMENT OF THE ATOM

1. At the beginning of the 19th century A. M. Ampère (1775-1836) came to the conclusion that the magnetic properties of matter could be explained if we assume that atoms are microscopic elementary magnets. Ampère could not, however, give grounds for the origin of the elementary magnetic moments. The nuclear model of atomic structure, which we have already used to explain the mechanism of deformation polarizability (see Sec. 38.5), will help us to understand the origin of the magnetic moments of atoms. Indeed, the electron, circulating in the atom, is an elementary current having a magnetic moment p_m^{orb} . Besides, the electron, circulating on its orbit, has an angular momentum called the *orbital angular momentum* L^{orb} . According to equation (40.28) there should be the following relationship between these two quantities:

$$p_m^{orb} = \frac{e}{2m_e} L^{orb} \quad (42.1)$$

where m_e = rest mass of the electron
 e = its charge.

The magnetic moment of the atom is the sum of the magnetic moments of the electrons revolving about the nucleus plus the magnetic moment of the nucleus itself. Experiments show, however, that the magnetic moment of the nucleus is about one two-thousandth as much as that of the electron (see Sec. 80.2). Therefore, in studying the magnetic properties of matter we neglect the magnetic moment of the nucleus assuming that the magnetic moment of the atom is the vector sum of the magnetic moments of the electrons circulating about the nucleus. We shall attempt to estimate the magnitude of this magnetic moment.

2. Let us choose a certain direction in space, for instance, the z -axis. This can mean that a certain magnetic field is directed along this axis. It has been shown in quantum mechanics that the projection of the orbital angular momentum in a given direction can take only definite values:

$$L_z^{orb} = m\hbar \quad (42.2)$$

where $\hbar = 1.05459 \times 10^{-34}$ J-s (see Ch. 14)

m = a whole number.

Using equation (42.1), we obtain for the projection of the orbital magnetic moment, which we denote by p_{mz}^{orb} , the following:

$$p_{mz}^{orb} = \frac{e}{2m_e} L_z^{orb} = m \frac{e\hbar}{2m_e} = m\mu_B \quad (42.3)$$

3. The quantity

$$\mu_B = \frac{e\hbar}{2m_e} = (9.274096 \pm 0.000065) \times 10^{-24} \text{ A}\cdot\text{m}^2 \quad (42.4)$$

is called the *Bohr magneton*. It is derived from fundamental constants—the charge and mass of the electron and Planck's constant—and is therefore a fundamental constant too.

The Bohr magneton is the minimum nonzero value of the projection of the magnetic moment of the electron in an arbitrary direction. Since the magnetic moment of the atom is the vector sum of the magnetic moments of the electrons, we can conclude that the projection of the magnetic moment of the atom onto some axis is either equal to zero or is a multiple of the Bohr magneton. Further on we shall see that experiments confirm this statement.

42.3. QUANTITIES CHARACTERIZING THE MAGNETIC FIELD IN MATTER

1. We introduced a special physical quantity, the polarization vector, to characterize the degree of polarization of a dielectric in an electric field. It equals the dipole moment of unit volume of the polarized dielectric (see Sec. 38.2). By analogy we now introduce a quantity that characterizes the degree of magnetization of a substance in a magnetic field. It is called the magnetization vector, or simply magnetization.

The *magnetization M* is defined as the magnetic moment per unit volume of the magnetized substance. It can be determined by finding the vector sum of the magnetic moments of all the atoms contained in unit volume of the magnetized substance. For the sake of simplicity we shall assume that the substance is homogeneous and that all of its atoms have the same magnetic moment \mathbf{p}_m . Then the magnetization is

$$\mathbf{M} = \frac{N\mathbf{p}_m}{V} = n\mathbf{p}_m \quad (42.5)$$

where n is the concentration of the atoms. The unit of magnetization coincides with the unit of the magnetic field strength:

$$[M] = [Np_m/V] = \text{A} \times \text{m}^2/\text{m}^3 = \text{A}/\text{m} = [H]$$

2. We shall restrict ourselves to the simplest case in which a homogeneous and isotropic substance completely fills the magnetic field. This can be accomplished by winding wire uniformly on a toroidal core of the substance being investigated and having current in the winding. The material of the core will, of course, be magnetized by the action of the magnetic field set up by the current.

If the substance is paramagnetic or diamagnetic, its magnetization vector will be found to be proportional to the magnetizing field strength. Thus

$$\mathbf{M} = \chi_m \mathbf{H} \quad (42.6)$$

Proof will be given in Secs. 42.4 and 42.5, where we shall deal with the mechanism for the magnetization of matter.

The proportionality factor χ_m in equation (42.6) is called the magnetic susceptibility of the substance. Obviously, the *magnetic susceptibility* is a dimensionless quantity.

3. The magnetization vector determines the magnetic induction \mathbf{B}_{sub} of the field set up by the magnetized substance. By analogy with the relationship between the induction of the field set up by currents and the field strength ($\mathbf{B}_{cur} = \mu_0 \mathbf{H}$), we can write the relationship between the magnetic induction of the field set up by the substance and the magnetization of the substance: $\mathbf{B}_{sub} = \mu_0 \mathbf{M}$.

The magnetic induction of the field in a substance that completely fills the field is the vector sum of the induction of the field set up by the magnetizing currents and the induction of the field set up by the magnetized substance. Thus

$$\mathbf{B} = \mathbf{B}_{cur} + \mathbf{B}_{sub} = \mu_0 (\mathbf{H} + \mathbf{M}) \quad (42.7)$$

Combining with equation (42.6) we have

$$\mathbf{B} = \mu_0 (\mathbf{H} + \chi_m \mathbf{H}) = \mu_0 (1 + \chi_m) \mathbf{H} = \mu_0 \mu \mathbf{H} \quad (42.8)$$

The quantity

$$\mu = 1 + \chi_m \quad (42.9)$$

is called the *magnetic permeability*, or relative magnetic permeability, of the substance.

42.4. DIAMAGNETISM

1. Let us try to find the explanation for the mechanism of the diamagnetic effect. For this we shall use the model of the helium atom. The nucleus of this atom has the charge $q = +2e$ and two electrons revolve about it. Experiments show that the helium atom has no magnetic moment. This can be explained by assuming that the two electrons revolve about the nucleus at the same velocity on like orbits but in opposite directions. Then their orbital magnetic moments will be equal in magnitude but of opposite sign so that the total magnetic moment of the atom equals zero (Fig. 42.3).

2. Next let us put the helium atom into a magnetic field having an induction $B_0 = \mu_0 H$. To simplify matters we shall assume that the induction vector is perpendicular to the planes of the electron

orbit. Now the electron is subject to two forces: the Coulomb force of attraction of the nucleus and the Lorentz force. Their resultant imparts centripetal acceleration to the electron. The motion equations for the two electrons will be of the form

$$\left. \begin{aligned} F_{\text{Coul}} - F_{\text{Lor}}^{(1)} &= \frac{2e^2}{4\pi\epsilon_0 r^2} - ev_1 B_0 = \frac{mv_1^2}{r} \\ F_{\text{Coul}} + F_{\text{Lor}}^{(2)} &= \frac{2e^2}{4\pi\epsilon_0 r^2} + ev_2 B_0 = \frac{mv_2^2}{r} \end{aligned} \right\} \quad (42.10)$$

It should be noted that in these equations we assume the magnetic field will change only the velocity of the electron but not its orbital radius.

3. It follows from equations (42.10) that the action of the magnetic field led to a reduction in the velocity of the first electron and an

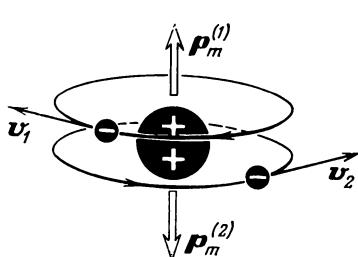


Fig. 42.3

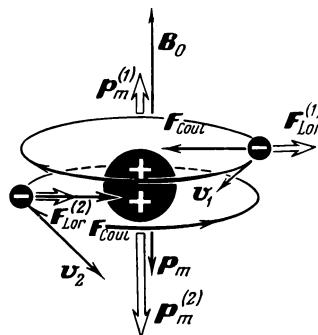


Fig. 42.4

increase in velocity of the second. Their magnetic moments are changed accordingly. Thus, a magnetic moment induced in the atom by the external magnetic field is

$$p_m = p_m^{(2)} - p_m^{(1)} = \frac{ev_2 r}{2} - \frac{ev_1 r}{2} = \frac{er}{2} (v_2 - v_1) \quad (42.11)$$

It should be pointed out that the induced magnetic moment p_m is directed opposite to the induction vector of the external magnetic field (Fig. 42.4).

We can find the difference in velocities by subtracting the upper of equations (42.10) from the lower. Thus

$$eB_0(v_2 + v_1) = \frac{m}{r} (v_2^2 - v_1^2)$$

or, after cancelling,

$$v_2 - v_1 = \frac{eB_0 r}{m} = \frac{er\mu_0 H}{m} \quad (42.12)$$

Substituting into equation (42.11), we obtain for the induced magnetic moment

$$\mathbf{p}_m = \frac{e^2 r^2 \mu_0 H}{2m} \quad (42.13)$$

Recalling that vectors \mathbf{p}_m and \mathbf{H} have opposite directions, we can rewrite equation (42.13) in vector form:

$$\mathbf{p}_m = -\frac{e^2 r^2 \mu_0}{2m} \mathbf{H} \quad (42.14)$$

4. According to equation (42.5), the magnetization vector is

$$\mathbf{M} = n \mathbf{p}_m = -\frac{e^2 r^2 n \mu_0}{2m} \mathbf{H} \quad (42.15)$$

We have found here that the magnetization vector really is proportional to the magnetizing field strength [see equation (42.6)]. The magnetic susceptibility turns out to be the negative quantity

$$\chi_m = -\frac{e^2 r^2 n \mu_0}{2m}$$

Let us estimate the magnetic susceptibility of helium. The electron orbital radius is $r = 0.5 \times 10^{-10}$ m and the atom concentration at ordinary pressures is $n = 2.69 \times 10^{25}$ m⁻³. We have

$$\chi_m = -\frac{1.6^2 \times 10^{-38} \times 0.5^2 \times 10^{-20} \times 2.69 \times 10^{25} \times 4\pi}{2 \times 9.1 \times 10^{-31} \times 10^7} = -1.2 \times 10^{-9}$$

Experiments yield the value $\chi_m = -2.25 \times 10^{-9}$ for the magnetic susceptibility of helium. Taking into account the crudeness of the model we employed, the agreement of the calculated and experimental data is to be considered quite good.

5. We were able to explain the mechanism of diamagnetic phenomena on a model of the helium atom. Generalizing, we can affirm that all substances whose atoms have no magnetic moment when there is no external field are diamagnetic materials.

Since the radii of the electron orbits do not differ greatly for the atoms of various substances, the magnetic susceptibilities of different substances in the same state of aggregation are approximately the same. For example, the magnetic susceptibility is -6.76×10^{-9} for nitrogen, -10.9×10^{-9} for argon, -2.25×10^{-9} for hydrogen, etc. The atom concentration of liquids and solids is about a thousand times more, and their magnetic susceptibility increases in the same proportion. The value for water is -9×10^{-6} , for silver -25×10^{-6} , etc.

42.5. PARAMAGNETISM

1. The mechanism for magnetizing diamagnetic materials, considered in the preceding section, reduces to the induction of an additional magnetic moment in the atom by the action of the external magnetic field. This phenomenon is typical for the atoms of any substance and therefore all substances should possess diamagnetic properties. What, then, is the origin of paramagnetic materials?

It was found that the atoms of paramagnetic materials have uncompensated magnetic moments, in contrast to diamagnetic materials in which the magnetic moments of the electrons in the atom exactly cancel. When an external magnetic field is applied to a paramagnetic material, the magnetic moments of its atoms tend to align with the lines of induction of the field. Thermal motion, naturally, interferes with this aligning process.

Evidently, the mechanism for magnetizing paramagnetic materials is quite similar to that for polarizing dielectrics having polar molecules (see Sec. 38.6). Using the same arguments as for deriving equation (38.25), we obtain the expression for the magnetization of paramagnetic materials:

$$\mathbf{M} = \frac{n p_m^2 \mathbf{B}_0}{3kT} = \frac{n p_m^2 \mu_0}{3kT} \mathbf{H} \quad (42.16)$$

2. The magnetic susceptibility of paramagnetic materials is positive:

$$\chi_m = \frac{n p_m^2 \mu_0}{3kT}$$

Evidently, at a constant concentration of the atoms (i.e. at constant density of the substance), the magnetic susceptibility of paramagnetic materials is inversely proportional to the absolute temperature (Curie's law):

$$\chi_m = \frac{\text{const}}{T}$$

This is experimentally confirmed.

We should note that expression (42.16) is valid provided that $p_m B_0 / kT \ll 1$, i.e. if the energy of a dipole in a magnetic field $\mathcal{E}_m = p_m B_0$ is much less than the energy of thermal motion of the particle $\epsilon = kT$. Since $p_m \cong \mu_B \cong 9.27 \times 10^{-24} \text{ A}\cdot\text{m}^2$ and the maximum attainable magnetic induction of a field is $B_{\max} \cong 1.3 \text{ T}$, the condition is always satisfied at room temperature. Only at very low temperatures can the value $p_m B_0 / kT \cong 1$ be obtained; the paramagnetic substance would have to be cooled to the temperature

$$T \cong \frac{p_m B_0}{k} = \frac{\mu_B B_{\max}}{k} = \frac{9.27 \times 10^{-24} \times 1.3}{1.38 \times 10^{-23}} \cong 1 \text{ K}$$

which is close to absolute zero. At this temperature the magnetization of the paramagnetic substance is near to saturation because all the magnetic moments of the atoms line up along the field and the weak thermal motion cannot violate this order.

3. Let us estimate the magnetic susceptibility of paramagnetic materials at room temperature. For gases $n = 2.69 \times 10^{25} \text{ m}^{-3}$; we take room temperature to be $T \cong 300 \text{ K}$. Then

$$\chi_m \cong \frac{2.69 \times 10^{25} \times 9.27^2 \times 10^{-48} \times 4\pi}{3 \times 1.38 \times 10^{-23} \times 300 \times 10^7} \cong 2.3 \times 10^{-7}$$

Experimental data on the magnetic susceptibility at 20 °C indicate 18.6×10^{-7} for oxygen and 0.8×10^{-7} for nitric oxide (NO). The susceptibility for solid paramagnetic materials should be about a thousand time higher. According to experimental measurements, it is 2.58×10^{-4} for platinum, 8.3×10^{-6} for sodium, 2.3×10^{-5} for lithium and much less for other substances.

Hence, the magnetic susceptibility of paramagnetic materials is about a hundred times higher than for diamagnetic substances. This explains why no diamagnetic properties are displayed by paramagnetic substances. The strong paramagnetic effect was found to overwhelm weaker diamagnetic phenomena.

42.6. FERROMAGNETISM. THE CURIE TEMPERATURE

1. In contrast to paramagnetic and diamagnetic materials in which the magnetization vector is proportional to the strength of the magnetizing field, in ferromagnetic substances this vector is a very complex nonlinear function of the field strength. This relationship is shown in Fig. 42.5 for nickel. For better illustration of the curves the scale has been distorted: the range of field strength from 0.1×10^3 to $2 \times 10^3 \text{ A/m}$ has been extended. We see in the graph that the magnetization of nickel first increases rapidly with the strength of the magnetic field. Then magnetic saturation is reached in which magnetization no longer depends, practically, on the strength of the magnetizing field.

2. Since the magnetization of a ferromagnetic material is not proportional to the strength of the magnetizing field, equations (42.6) and (42.8) are not valid. Formally, however, they can be applied if we take into account the fact that for ferromagnetic substances the

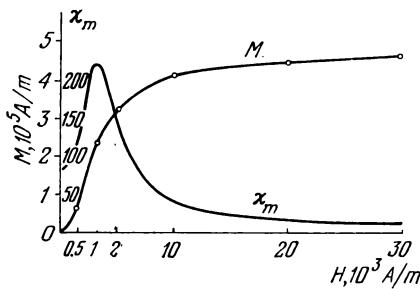


Fig. 42.5

magnetic susceptibility $\chi_m = M/H$ and the magnetic permeability $\mu = B/\mu_0 H$ are not constants but are quite complex functions of the magnetizing field strength.

3. The next important difference of ferromagnetic materials is that their ferromagnetism is observed only in crystals, while paramagnetism and diamagnetism are displayed in any state of aggregation. In the liquid and gaseous states, ferromagnetic materials behave like ordinary paramagnetic substances. Hence the ferromagnetism of certain substances is not determined by any special properties of their atoms, which differ in no way from the atoms of paramagnetic substances, but by the special structure of their crystal lattices. The magnetic properties of ferromagnetic materials should change with changes in their lattice structure. This statement is confirmed by experiments.

Each ferromagnetic substance was found to have a definite temperature, the *Curie temperature*, above which ferromagnetism disappears and the substance becomes only paramagnetic. The Curie temperature is 770 °C for iron, 360 °C for nickel and only 70 °C for Permalloy (an alloy of 70% Fe and 30% Ni).

4. Academician Landau showed that the Curie temperature of ferromagnetic materials is the temperature of a second-order phase transition.

The structure of substances is changed in first-order phase transition. In melting a crystal, for instance, long-range order is changed to short-range order. As a result the internal energy and entropy of the body are changed, expending, for this purpose, a certain amount of energy—the heat of phase transition (see Ch. 36). Second-order phase transitions, for example, the disappearance of ferromagnetism when a body is heated above the Curie temperature, are not accompanied by changes in the state of aggregation. The crystal lattice is not broken up though its type of symmetry is changed. Since the internal energy and entropy of a body are not changed in second-order phase transitions, no heat of phase transition is required.

Landau showed that such properties of substances as the specific heat, coefficient of linear expansion, compressibility and others change with jumps in second-order phase transitions. This is confirmed by experiments. Graphs showing the variation of the magnetization and specific heat of some substance are given in Fig. 42.6. We see that at a temperature of 4.2 K the magnetization drops sharply. This, therefore, is the Curie temperature of the given substance. The specific heat first rises rapidly and then drops at this temperature.

5. Ferromagnetism is inherent only in crystals which possess anisotropy (see Sec. 32.1). Hence, a monocrystal of a ferromagnetic material should have anisotropy of magnetization: its magnetic

properties should differ in different directions. This assumption can be confirmed experimentally.

Shown in Fig. 42.7 are curves of the magnetization of a monocrystal of iron having a body-centered cubic lattice (see Sec. 33.4). It is found that the magnetization of the monocrystal for a given field strength is maximum (a) for a direction along an edge of the cube, less (b) in the direction of a diagonal of a face of the cube and least (c)

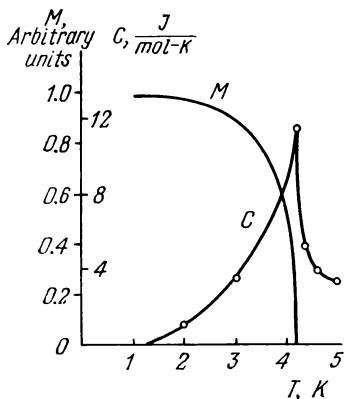


Fig. 42.6

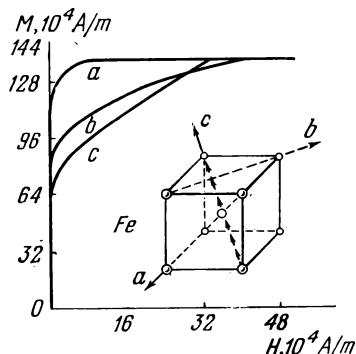


Fig. 42.7

along a body diagonal. For this reason the edge of the cube is called the direction of easiest magnetization and the diagonal, the direction of hard magnetization.

Naturally, no anisotropy of magnetization is observed in a polycrystal because the separate microcrystallites are randomly oriented (see Sec. 32.2).

42.7. HYSTERESIS

1. A typical property of ferromagnetic materials is *hysteresis*. The essence of this phenomenon is that the magnetization of a ferromagnetic material depends not only on the strength of the magnetizing field at the given instant but on the previous magnetization of the specimen. Hence, it is impossible to indicate what magnetization of the ferromagnetic substance corresponds to a given value of the magnetizing field strength if we do not know in what condition the substance was before being put into the field. Of course this also refers to the values of the magnetic susceptibility and permeability.

The graph of Fig. 42.5 represents the course of initial magnetization when the ferromagnetic material is first heated above the Curie temperature and, thereby, completely demagnetized and then cooled

and magnetized. The magnetization curve will be of entirely different shape if the material has been previously magnetized.

2. Let us make a core of toroid shape from a demagnetized ferromagnetic material and wind a wire uniformly on it. Varying the current in the winding, we, thereby, vary the strength of the magnetizing field. We increase the field strength to the value H_s (Fig. 42.8). This field corresponds with the saturation magnetization M_s . Next we reduce the current in the winding, thereby reducing the strength

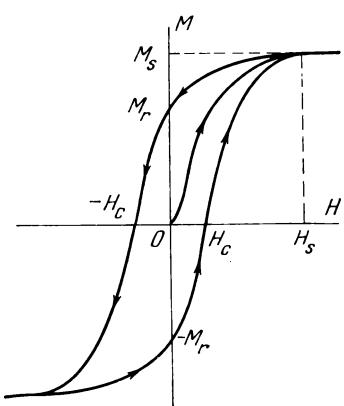


Fig. 42.8

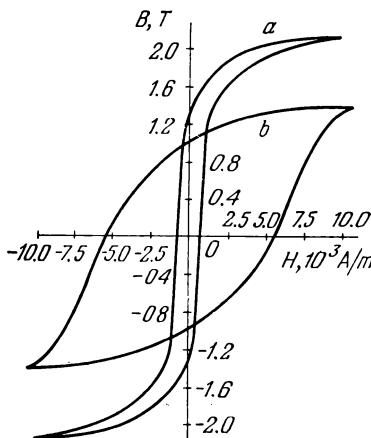


Fig. 42.9

of the magnetizing field. We shall see that the magnetization of the core in demagnetization is at all times greater than it was when the core was being magnetized.

When the current in the winding reaches zero, the magnetizing field also disappears. But the magnetization of the ferromagnetic substance does not vanish. The core retains a certain *remanent magnetization*, M_r . Only when the current in the toroid winding is reversed and increased until it sets up a field of the strength $-H_c$ will the magnetization of the core vanish. The strength H_c of the demagnetizing (reverse) field is called the *coercive force*, or *coercivity*.

The curve shown in Fig. 42.8 is called the *hysteresis loop*.

3. Ferromagnetic materials may have different properties that depend upon their chemical composition, and also the kind of heat treatment and processing they undergo (hardening, annealing, rolling, cold forging, pressworking, etc.). Substances with low coercivity are called *soft* magnetic materials; high-coercivity materials are said to be *hard*. Hysteresis loops are illustrated in Fig. 42.9 for two materials: soft-magnetic iron (a) and high-coercivity hardened steel (b).

The shape of the hysteresis loop determines the field of application of a ferromagnetic material. Soft magnetic materials are used for making the cores of electromagnets, in which it is important to have a high value of the maximum field induction and a low coercive force. These materials are also used for the cores of transformers and alternating current machines (generators and motors); owing to their low coercivity their magnetization is easily reversed. The same requirements are made to the magnet cores of particle accelerators.

Hard-magnetic materials are used for making permanent magnets. Owing to the high value of their coercive force and the comparatively high remanent induction, these magnets can maintain strong magnetic fields for a long time.

Permanent magnets of high-coercivity alloys are applied in magnetoelectric (moving-coil) measuring instruments, electrodynamic loudspeakers and microphones, in small generators (for instance, those for bicycles), in micromotors, etc.

42.8. THE DOMAIN STRUCTURE OF FERROMAGNETIC MATERIALS

1. The features of ferromagnetic substances can be explained on the basis of the classical theory of ferromagnetism proposed by P. Weiss (1865-1940). This theory contends that at temperatures below the Curie temperature ferromagnetic substances are made up of microscopic regions called *domains*. In each domain the atoms have magnetic moments aligned strictly in a single direction, that of easy magnetization (see Sec. 42.6). In this manner each domain is magnetized to saturation regardless of whether there is any external magnetic field or of its magnitude.

The domains range from 10^{-2} to 10^{-3} cm in size and can, therefore, be observed in a microscope. For this purpose a drop of oil with a suspension of fine particles of ferromagnetic powder is applied to the polished surface of the ferromagnetic material (Akulov-Bitter method). The particles of the ferromagnetic powder concentrate on the domain boundaries where the magnetic field is most nonuniform (Fig. 42.10).

When there is no external magnetic field, the domains in a monocrystal are arranged so that their magnetic fields close one another and the total external magnetic field equals zero. These domains are shown schematically in Fig. 42.11. The arrows show the directions of magnetization inside the domains. As was shown by L. D. Landau and E. M. Lifshitz (b. 1915), such a system of domains in a monocrystal characterizes a state of minimum energy of the magnetic field. This, from the viewpoint of thermodynamics, ensures stable equilibrium of the system.

2. On the basis of these conceptions let us explain the mechanism of magnetization of a polycrystalline specimen. A polycrystal con-



Fig. 42.10

sists of separate grains having directions of easiest magnetization that are randomly oriented. Each grain is divided into several domains directed along the line of easiest magnetization (Fig. 42.12a).

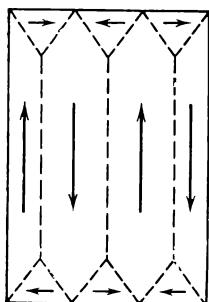


Fig. 42.11

If we impose a weak magnetic field on the specimen, there will always be many domains in which the direction of their magnetization vector coincides with that of the external field. These domains possess minimum energy and are, therefore, in a state of stable equilibrium. Neighbouring domains may possess maximum energy. If so, it will prove favourable for the magnetic moments of certain atoms to become reoriented and to join the domains whose energy is minimal. This process is called domain wall displacement (Figure 42.12b).

As the strength of the imposed magnetic field increases, the domain boundaries are displaced more and more. The domain walls begin to encounter various crystal defects, such as dislocations or interstitial atoms (see Sec. 32.4). These defects hinder directional changes in the magnetic moments of atoms located near the defect (they hinder domain wall motion). As the imposed field continues to become stronger, the potential energy of the atoms (the magnetic dipoles)

in the vicinity of the defect increase until it exceeds the activation energy (see Sec. 34.3). Then the magnetic moments of the atoms suddenly reverse their direction. In other words, the domain wall breaks away with a jump and moves to the next defect. Hence, the magnetization of a polycrystalline substance is not a gradual process, but takes place in jumps. This phenomenon is called the *Barkhausen effect, or jump.*

This effect can be detected experimentally. If a magnet is brought close to a bundle of nickel wire, stepwise magnetization of the nickel occurs, leading to stepwise changes (in jumps) of the magnetic

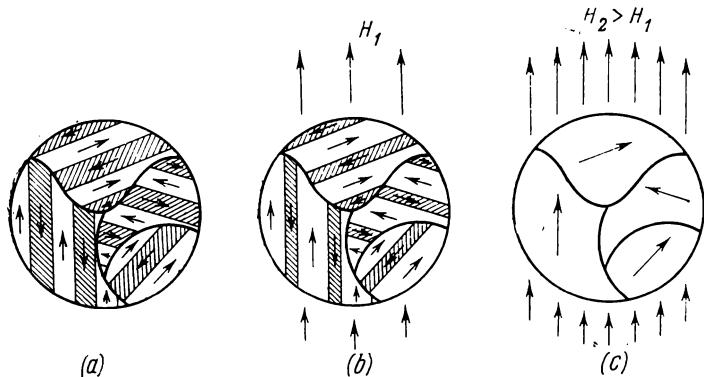


Fig. 42.12

field. Owing to the phenomenon of electromagnetic induction (see Ch. 43), current pulses, quite audible in a telephone receiver, are induced in the wire wound on the nickel bundle.

The stepwise nature of the magnetizing process of a polycrystalline ferromagnetic substance leads to irreversible losses of energy (motion of dislocations and other defects, the production of elastic waves whose energy is transformed into internal energy, etc.). Hence, the magnetization process in strong fields is also irreversible. When the strength of the magnetic field is reduced, the original structure of the substance is not restored. The demagnetized domains have other shapes and orientation than they had in the process of magnetization. This is the cause of hysteresis phenomena.

3. In sufficiently strong fields all the domain walls are moved to the grain boundaries; each grain is magnetized along that direction of easiest magnetization that makes the least angle with the direction of the external field (Fig. 42.12c). With a still stronger magnetizing field, the magnetic moments of the grains begin to turn into alignment with the field. This process requires very strong fields. Consequently, the magnetization changes slowly in this region and the growth of the magnetization curve is retarded.

Finally, when the magnetic moments of all the grains are aligned with the field, magnetic saturation is reached. The magnetization of the substance no longer increases with the strength of the imposed field.

4. We see that all the features of the magnetization process of ferromagnetic materials can be explained on the basis of their domain structure. The Curie temperature is the one above which the domain structure is broken up.

Two questions remain to be answered. What forces spontaneously align the magnetic moments of all the atoms within a domain along the direction of easy magnetization? Why do certain substances have ferromagnetic properties and others do not? Only the quantum theory of magnetic phenomena could find answers to these questions.

42.9. THE EINSTEIN AND DE HAAS EXPERIMENT

1. In the preceding sections we made use of the concept that the atoms of paramagnetic and ferromagnetic materials have magnetic moments and that in an external magnetic field these moments are

aligned along the field. We found in Sec. 42.2 that the ratio of the orbital magnetic moment to the orbital angular momentum equals one half of the electron charge-to-mass ratio. It was obviously important to check the validity of these concepts experimentally. The idea of such an experiment was proposed in 1915 by Einstein. He conducted the experiment in the same year with W. J. de Haas (1878-1960).

The essence of this experiment is the following. We suspend a rod of a ferromagnetic material by a thin thread in a coil so that the axis of the rod and its suspension exactly coincide with the direction of the induction vector of the magnetizing field. A current is passed through the coil winding, magnetizing the rod to saturation. Then the current is reversed in the coil,

magnetizing the rod to saturation in the opposite direction (Fig. 42.13). This alternating magnetization process consists in reversing the magnetic moments of the atoms. But at the same time the angular momentum of the electrons is reversed.

The system of electrons and positive ions that form the crystal lattice of the rod is a closed one because it is not subject to any mechanical moments. But then, according to the well-known conservation law (see Sec. 22.7), the total angular momentum of the electrons and ion lattice must remain constant. And since the angular momentum of the electrons changes when the rod is alternately

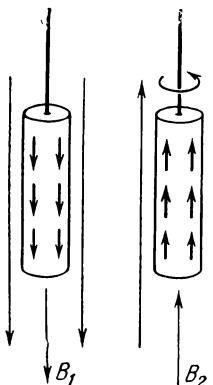


Fig. 42.13

magnetized the rod itself must acquire an angular momentum of such a value that the algebraic sum retains its initial value. As a result the rod turns about an axis parallel to the field.

This was experimental proof that atoms have a magnetic moment due to the motion of electrons in them. According to equation (42.1), the magnetic moment produces a mechanical angular momentum.

2. In the same year, S. J. Barnett (1873-1956) showed that there is an effect which in a sense is the reverse of the Einstein and de Haas effect. He found that if an iron rod is rotated rapidly, it will be magnetized. This experiment also yielded the ratio of the magnetic moment to the angular momentum.

3. A careful analysis of the data obtained in this experiment showed that in both the Einstein-de Haas and Barnett experiments the ratio of the magnetic moment to the angular momentum has the expected sign, corresponding to the negative charge of the electron, but that the numerical value turned out to be twice as large as given by equation (42.1), i.e.

$$\frac{p_m}{L} = \frac{e}{m_e} \quad (42.17)$$

The following explanation of this result may be proposed. Let us assume that the electron has not only an orbital magnetic moment for which equation (42.1) is valid, but another, the *spin magnetic moment*, which satisfies the relationship (42.17). And if we take into consideration the fact that orbital magnetic moments can be compensated for, then the results of the Einstein-de Haas and Barnett experiments can be easily explained.

The existence of a spin magnetic moment of electrons was found to follow directly from the Stern-Gerlach experiment.

42.10. THE STERN-GERLACH EXPERIMENT

1. In 1921 O. Stern and W. Gerlach (b. 1889) conducted a series of experiments in which they directly measured the magnetic moments of the atoms of various chemical elements. This was done by passing a beam of atoms through a magnetic field with a very high induction gradient $\Delta B/\Delta z$. This gradient was obtained by employing an electromagnet with special pole tips (Fig. 42.14).

Atoms of the metal are vaporized in oven 1 and pass through slit 2 to form a narrow beam. When there is no magnetic field, this beam strikes the target at its centre leaving a trace on photographic plate 3.

If the magnetic field is switched on, then the force $F_z = p_{mz} \frac{\Delta B}{\Delta z}$ [see equation (41.19)] is applied to the atoms along the z -axis. Here p_{mz} is the projection of the magnetic moment onto the z -axis. This force deflects the atoms up or down depending on whether the projec-

tion of the magnetic moment is in the direction of or opposite to the axis. The projection of the magnetic moment of the atom can be calculated from the known gradient of field induction and the measured deflection of the beam.

2. According to classical concepts, any orientation of the magnetic moment with respect to the z -axis is equally probable so that the projection of the magnetic moment may have any value from $-p_m$ to $+p_m$, including zero. This would correspondingly vary the force

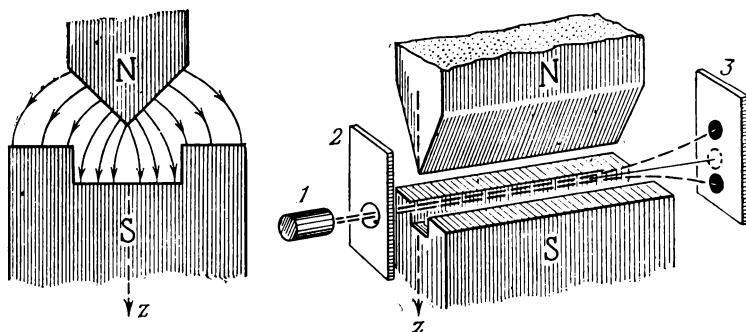


Fig. 42.14

acting on the atoms. Since there are very many atoms in the beam, the trace left on the photographic plate, according to classical concepts, should be a wide band. But no such band is obtained in the experiments. Hence, the assumption of equally probable orientation of the magnetic moments in random directions, lying at the basis of classical physics, is false.

3. The experiments showed that the atoms of some substances for example, mercury, are not deflected at all by the magnetic field. This means that these atoms have no magnetic moments and therefore the substances must be diamagnetic, which is the case.

Especially interesting results were obtained with the alkali metals lithium, sodium and potassium (and also aluminium and some others). Here the atom beam was divided in two, each being deflected the same amount but in opposite directions (Fig. 42.14). This means that the z -projection of the magnetic moment of the given atom can have *only two values*, equal in magnitude and opposite in direction. When the deflection was measured, it was found that in the given case the projection of the magnetic moment of the atom equals the Bohr magneton. Thus

$$p_{mz} = \pm \mu_B$$

$$(42.18)$$

42.11. ELECTRON SPIN

1. Let us analyze in more detail the result obtained with the alkali metals, for example, lithium. Lithium is the third element of the Periodic Table, therefore, three electrons revolve about its nucleus (see Sec. 73.2). But, at the same time, lithium, like all the alkali metals, is a monovalent element. This leads us to the conclusion that one of the electrons of lithium differs in some way from the other two. As will be shown in Sec. 73.1, the two electrons are on the same orbit as the two electrons of helium while the third electron is on another orbit.

Electrons on the helium orbit do not contribute to the magnetic moment of the atom since their magnetic moments are compensated (see Sec. 42.4). Consequently, the magnetic moment of the lithium atom is due to the uncompensated magnetic moment of the valence electron. But since the atoms of all alkali metals have a single valence electron, they should all have an uncompensated magnetic moment. As a result, all alkali metals should be paramagnetic substances, which they have been found to be.

2. However, a careful analysis indicates that this uncompensated magnetic moment is not an orbital moment.

In 1925, after studying the difficult questions that atomic physics had been unable to answer to that date (including the explanation of the results of the Stern-Gerlach experiment), S. A. Goudsmit (b. 1902) and G. E. Uhlenbeck (b. 1900) came to the conclusion that the electron has its own angular momentum, which was named *spin*, implying that it (the electron) rotates about its axis (see Sec. 72.5).

3. Let us direct our attention to the role of spin in the origin of ferromagnetism. By means of calculations beyond the scope of this book, Frenkel and Heisenberg showed that in the electrostatic interaction of electrons of a substance the interaction energy consists of two parts: the energy of the Coulomb field and another component, a form of interaction called *exchange interaction*. This last is a purely quantum mechanical effect having no classical analog (see Sec. 74.3).

It was found that in parallel spin orientation the energy of exchange interaction is negative, corresponding to forces of attraction. But minimum energy is the condition for stable equilibrium of the system (see Sec. 19.6). Therefore, as a result of the exchange interaction the spin moments of the atoms should align in a single direction if there was no interference from thermal motion.

The exchange interaction energy of paramagnetic materials is less than their energy of thermal motion. As a result these materials cannot become magnetized spontaneously. As to ferromagnetic substances their exchange interaction energy below the Curie temperature is substantially greater than the energy of thermal motion.

Consequently, the spins of a large group of atoms are all oriented in the direction of easiest magnetization, thereby forming regions of spontaneous magnetization, i.e. domains.

42.12. ANTIFERROMAGNETISM

1. In his research on the properties of ferromagnetic substances Landau came to the conclusion in 1933 that there must exist substances in which the magnetic moments at low temperatures are oriented not parallel to one another as in ferromagnetic substances

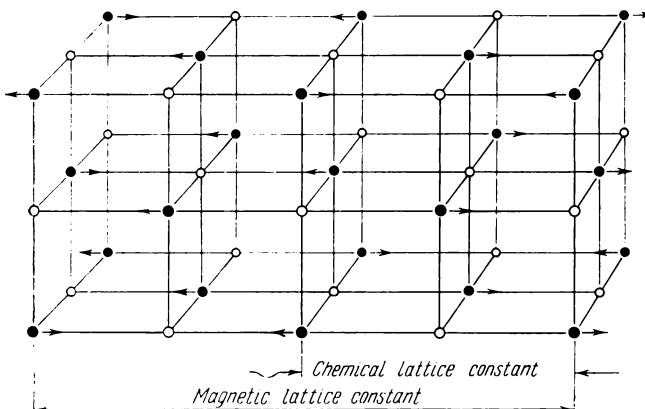


Fig. 42.15

but in antiparallel alignment, i.e. with the magnetic moments of adjacent atoms oriented oppositely. It was discovered experimentally in 1938 that the compounds MnO, MnS, Cr₂O₃, NiCr and a number of others display antiferromagnetic properties.

The crystal lattice of manganese oxide MnO is shown schematically in Fig. 42.15. The light circles without arrows represent the oxygen atoms, which have no magnetic moment; the dark circles represent the paramagnetic manganese atoms and the arrows indicate the direction of their magnetic moments. We see that the crystal has two lattice constants: the chemical lattice constant equal to the edge of the cube which can, by parallel displacement, construct the whole crystal lattice, and the magnetic lattice constant equal to the edge of the cube that can, by parallel displacement, construct the whole magnetic lattice. It was shown experimentally, using the phenomena of neutron diffraction (see Sec. 69.2), that the magnetic lattice constant is twice as long as the chemical lattice constant.

2. Antiferromagnetic materials have a temperature, analogical to the Curie temperature of ferromagnetism, above which the magnetic

order illustrated in Fig. 42.15 is destroyed and the material becomes paramagnetic. This is called the *Néel temperature* T_N . Like the Curie temperature it is a temperature of a second-order phase transition.

While the behaviour of ferromagnetic and antiferromagnetic materials is generally the same above the phase transition temperature—both become paramagnetic substances—they behave entirely

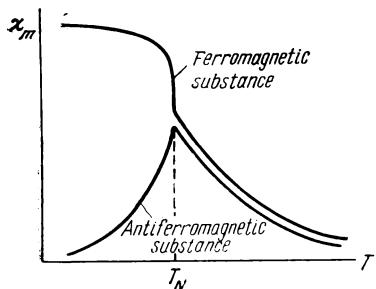


Fig. 42.16

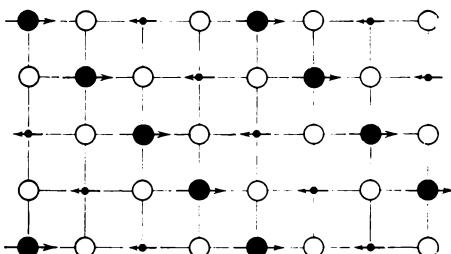


Fig. 42.17

differently below this temperature. Owing to spontaneous magnetization and the formation of domains below the Curie temperature, the magnetic susceptibility of ferromagnetic substances increases in a jump by hundreds and thousands of times. Then, down to temperatures near to absolute zero, their magnetic susceptibility depends only slightly on the temperature. In antiferromagnetic materials below the Néel temperature, magnetic order is established, the number of antiparallel moments increases and the magnetization is sharply reduced. This correspondingly reduces the magnetic susceptibility. As the temperature is lowered, magnetic order is improved because the energy of thermal motion, interfering with ordered orientation of the magnetic moments, is reduced. Consequently, as the temperature of the antiferromagnetic substance approaches absolute zero, its magnetic susceptibility also tends to zero (Fig. 42.16).

3. Let us imagine a substance having antiferromagnetic order in which a half of the manganese atoms are replaced by other paramagnetic atoms having a large magnetic moment, for instance, iron atoms. In Fig. 42.17 these atoms are shown by dark circles of larger diameter, with longer arrows. Below the Néel temperature such a substance is spontaneously magnetized because the difference between the magnetic moments of two adjacent paramagnetic atoms does not equal zero.

Materials with such properties actually exist; they are called ferrites and their properties are called ferrimagnetism (note the spel-

ling). Ferrites are made by sintering powders consisting of an intimate mixture of ferric oxide (Fe_2O_3) and the oxides of other metals, such as lithium, nickel and manganese. Depending on the composition

and treatment, ferrites with a variety of properties can be obtained. For example, the coercivity of nickel-zinc ferrites is about 1 A/m , while that of cobalt ferrites is about $2 \times 10^5 \text{ A/m}$. The hysteresis loop of magnesium-manganese ferrites is almost rectangular (Fig. 42.18). This enables them to be used for the memory circuits in computers.

The magnetic induction of ferrites is much lower than that of ferromagnetic materials. But ferrites possess one important advantage: while metallic ferromagnetic materials are all good conductors of electric current, ferrites have a relatively low electric conductivity.

This permits ferrites to be employed as the cores of devices operating with high-frequency current, in which it is impossible to apply steel cores.

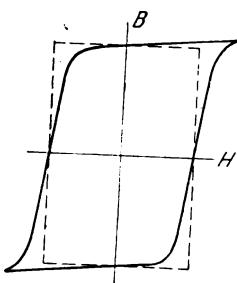


Fig. 42.18

Chapter 43

ELECTROMAGNETIC INDUCTION

43.1. FARADAY'S DISCOVERY

1. On the basis of the fact that a magnetic field is set up around a conductor with current, Michael Faraday came to the conclusion, as early as 1821, that a current could be produced in a conductor by means of a magnetic field. Witness to this is an entry for that year in his diary in which he set himself the task of converting magnetism into electricity. But ten years of intensive investigation were required before he solved this problem. Faraday discovered that if two windings are made on a core and the current is varied in one of the windings (for instance, by closing or opening the circuit), current is generated in the secondary winding.

Next he found that the current in the secondary winding is sharply increased if an iron core is employed. Then it turned out that the first winding with current can be replaced by a magnet and that current is produced in the second winding either when the winding is moved with respect to the core or when the core is moved in a stationary winding.

2. Faraday's discovery initiated a new era both in the theory of electromagnetism and in its practical applications. Grounds were

finally established for the unity of electric and magnetic phenomena, which served as the stimulus for the development of the theory of the electromagnetic field. This theory was worked out by Maxwell in the seventies of the last century. Faraday's discovery also became the foundation for the development of modern electrical engineering since all electrical machines used today are based on the phenomenon of electromagnetic induction.

43.2. ELECTROMAGNETIC INDUCTION AND THE LORENTZ FORCE

1. The mechanism by which an induced current is produced in a moving conductor can be explained by means of the Lorentz force. Suppose a conductor in which there are free positive and negative charges moves at the velocity v perpendicular to the induction vector B (Fig. 43.1). Then the charges will be subject to the Lorentz force $F_m = qvB$ which is directed, according to the left-hand rule (see Sec. 41.1), along the conductor.

The Lorentz force *divides the charges*: positive charges are accumulated at one end of the conductor and negative charges at the other. These charges set up an electrostatic Coulomb field inside the conductor. If the conductor is not connected to any circuit then the motion of the charges due to the action of the Lorentz force will continue until the electric force acting on the charges as a result of the Coulomb field becomes equal to the Lorentz force.

We see that the action of the Lorentz forces is similar to the action of a certain electric field opposing the Coulomb field. Since this field is not due to the division of the charges, but to nonelectrostatic (in our case, Lorentz) forces, it is an *extraneous field*. The strength of the extraneous field is

$$E^* = F_m/q = vB \quad (43.1)$$

2. Let us now connect the conductor moving in a magnetic field to an external load as shown in Fig. 43.2. We will find that the potential difference between the ends of the conductor is less than when the circuit was open. The strength of the Coulomb field inside the conductor is correspondingly reduced.

The strength of the extraneous field depends only on the velocity of motion of the conductor and on the induction of the magnetic field. Therefore, it will not change when the circuit is closed. Thus the strengths of the Coulomb and extraneous fields are equal before

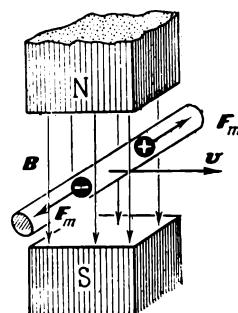


Fig. 43.1

the conductor is connected to a circuit, but the strength of the Coulomb field is less than that of the extraneous field in the closed circuit.

3. Let us follow out the direction of the current carriers. The direction of the current is determined by the direction in which the positive charges move along the field, i.e. from a region of higher potential to one of lower potential.

In the external part of the circuit the positive charges move from the region of potential φ_1 to the region of potential φ_2 by the action

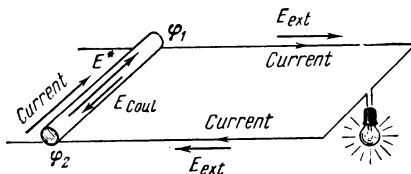


Fig. 43.2

of the electric field of a strength E_{ext} . The work done by this field in moving a charge is

$$W_{ext} = q(\varphi_1 - \varphi_2) \quad (43.2)$$

In the internal part of the circuit the positive charges move from the region with the lower potential φ_2 to the one with the higher potential φ_1 . This is possible because the extraneous forces (Lorentz forces) are higher in this case than the Coulomb forces which oppose charge motion. The work done in moving a charge in the internal part of the circuit equals the difference in the work done by the extraneous and Coulomb fields. Coulomb forces are conservative and the work they do does not depend upon the shape of the path (see Sec. 18.7). Hence the work done in moving a charge in the internal part of the circuit is

$$W_{int} = ql(E^* - E_{Coul}) = qlvB - q(\varphi_1 - \varphi_2) \quad (43.3)$$

43.3. INDUCED ELECTROMOTIVE FORCE

1. Combining equations (39.11) and (43.1) we obtain the expression for the induced emf:

$$\mathcal{E} = E^*l = vBl \quad (43.4)$$

It should be pointed out that we considered a case in which the velocity vector is perpendicular to the induction vector of the magnetic field. We shall leave it as an exercise for the student to show that in the general case

$$\mathcal{E} = Blv \sin \alpha \quad (43.5)$$

where α is the angle between the direction of the velocity vector and the magnetic induction vector.

2. Let us calculate the work done in moving a charge along a closed circuit. Combining equations (43.3), (43.4) and (43.2) we have

$$\begin{aligned} W_{cir} &= W_{ext} + W_{int} = q(\varphi_1 - \varphi_2) + qlvB - q(\varphi_1 - \varphi_2) \\ &= qlvB = q\mathcal{E} = W_{extr} \end{aligned}$$

from which

$$\mathcal{E} = W_{cir}/q \quad (43.6)$$

We see that the work done in moving a charge along a closed circuit is equal to the work done by extraneous forces. This enables us to give another definition of the emf: the electromotive force is equal to the ratio of the work done in moving a charge along a closed circuit, making a single round, to the magnitude of the charge.

43.4. THE INDUCTION PHENOMENON IN A STATIONARY CONDUCTOR

1. We shall use the arrangement shown in Fig. 43.3 to explain the mechanism of electromagnetic induction in a stationary conductor. Here reference frame xyz is fixed to a conducting loop, and frame

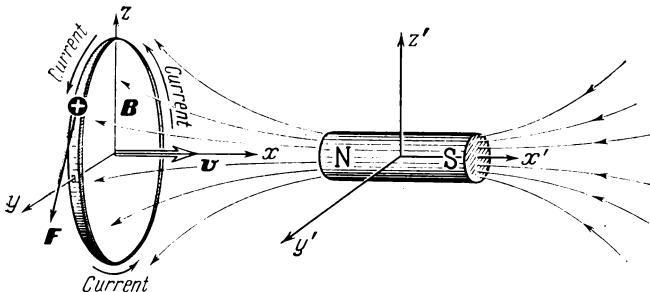


Fig. 43.3

$x'y'z'$ to the source of a magnetic field, for instance to a permanent magnet. Suppose that the loop and magnet approach each other with the velocity v .

In reference frame xyz , linked to the loop, the charges are stationary with respect to the frame and are therefore not subject to the Lorentz force. Nevertheless, as the magnet approaches the loop a current is induced in the latter. This result wholly agrees with the relativity principle according to which all inertial reference frames are equivalent and phenomena (current in the given case) are deter-

mined only by the relative velocity with which the loop and magnet approach each other. But what forces in this case cause circulation of charges around the loop?

2. We know that the motion of charges may be due to the action of either electric or Lorentz forces. Since Lorentz forces cannot be acting in reference frame xyz , linked to the loop, we conclude that the current in the loop is produced by an *induced electric field* which is not in reference frame $x'y'z'$.

An induced electric field has the following special features:

(a) *It is not a Coulomb field*; it is not set up by some distribution of charges, but by the approach of a magnet (i.e. source of a magnetic field) to the conducting loop.

(b) In contrast to the lines of force of a Coulomb field, which begin at the positive and end at the negative charges, the *lines of induction of an induced field are closed on themselves*. For this reason such a field is said to be a *circuital field*.

(c) *An induced field is not conservative* because the work done by this field in moving a charge along a closed circuit does not equal zero. Consequently, the energy characteristic of an induced field is the induced electromotive force (induced emf) and not the potential.

43.5. STRENGTH OF AN INDUCED FIELD

1. Let us determine the strength of an induced electric field. In reference frame $x'y'z'$ (Fig. 43.3) the charge is subject to the Lorentz force $F'_m = qvB'$, and in reference frame xyz to the electric force $F_e = qE$. We derived the relationship between these forces in Sec. 40.2. It is $F' = F\sqrt{1 - v^2/c^2}$. Thus $qvB' = qE\sqrt{1 - v^2/c^2}$ or

$$E = \frac{vB'}{\sqrt{1 - v^2/c^2}} \quad (43.7)$$

2. In reference frame $x'y'z'$ the electromotive force is equal in magnitude to the work done by extraneous (in our case, Lorentz) forces in moving a unit charge: $\mathcal{E}' = F'l/q = vB'l$. In reference frame xyz the emf is equal in magnitude to the work done by the induced electric field in moving a unit charge once around a closed circuit, i.e. $\mathcal{E} = Fl/q = El$. But the length of the conductor is the same in both frames of reference because the loop is perpendicular to the direction of motion. Hence the strength of the induced field is expressed by formula (43.7). Thus

$$\mathcal{E} = \frac{vB'l}{\sqrt{1 - v^2/c^2}} = \frac{\mathcal{E}'}{\sqrt{1 - v^2/c^2}} \quad (43.8)$$

In the phenomenon of electromagnetic induction the velocity of the conductors is considerably less than that of light and so the emf's in the two frames of reference practically coincide.

43.6. THE ELECTROMAGNETIC FIELD AND THE RELATIVITY PRINCIPLE

1. An analysis of the phenomenon of electromagnetic induction from the viewpoint of two different inertial reference frames leads us again to the same conclusion we reached in Sec. 40.3, namely, that the division of a unified electromagnetic field into electric and magnetic components depends on the reference frame in which the field is described. Indeed, in the reference frame with respect to which the charges are at rest (see Fig. 40.2), their interaction is accomplished by means of a field that we call electrostatic. There is no magnetic field in this reference frame. In the other frame of reference, with respect to which the charges are moving, the interaction between them is accomplished by means of a field that can be divided into two components: electric and magnetic. We can say that there are two fields, electric and magnetic, in this frame of reference, and that the interaction between the charges is accomplished in this frame by means of both fields.

We find a similar case in the arrangement illustrated in Fig. 43.3. In the reference frame linked to the magnet there is only a magnetic field and the motion of the charges in the loop is due to the Lorentz forces. In the reference frame of the loop, besides the magnetic field there is a circuital electric field which produces the current.

2. Again we come to the conclusion that "electric and magnetic fields do not exist independently of the state of motion of the reference frame". This is Einstein's statement taken from his first paper on the theory of relativity, published in 1905 and titled "On the Electrodynamics of Moving Bodies".

According to Einstein, the force acting on a charged body is calculated by multiplying the charge by the electric field strength at the point where the body is located. The expression for the electric field strength, used in the calculation, should be transformed to the reference frame in which the body being investigated is at rest.

43.7. FARADAY'S LAW OF INDUCTION

1. It remains to consider the phenomenon of electromagnetic induction for the case when the conductor in which the current is induced and the source of the magnetic field are both at rest in the given frame of reference. In Faraday's experiment with two coils wound on some kind of core, current appears in the secondary winding only upon changes in the magnetic field set up by the primary winding. Note that the way in which the field is changed is of no significance. One coil can be moved with respect to the other; the current can be turned on and off in the primary winding with a switch or the current in this winding can be varied with a rheostat. The result will in all

cases be the same: a current is induced in the secondary winding upon any changes in the magnetic field. We already know that in a conducting loop the circulation of the charges is caused by the induced electric field. Hence, we reach the conclusion that any changes in the magnetic field are accompanied by the establishment of a circuital electric field. This interpretation of induction phenomena was proposed by Maxwell, and it turned out to be especially fruitful.

2. Let us transform equation (43.5) so that it becomes suitable for calculating the induced emf in cases when the conductor is stationary

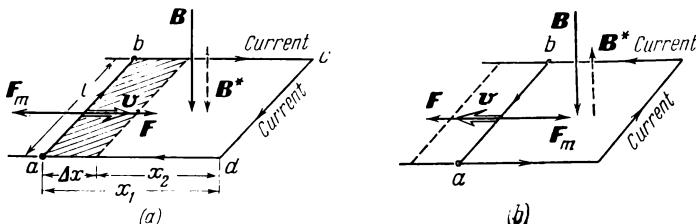


Fig. 43.4

and the magnetic field changes. For this purpose we shall find it necessary to introduce a new scalar characteristic of the magnetic field: the *flux of the magnetic induction vector*, or simply, the *magnetic flux* Φ .

Suppose the plane surface $abcd$ of area A to be located in a uniform magnetic field perpendicular to the lines of induction (Fig. 43.4a). The flux of the induction vector \mathbf{B} through the given surface A is the product of these quantities. Thus

$$\Phi = BA \quad (43.9)$$

The SI unit of magnetic flux is the *weber* (Wb): this is the flux threading through a surface of 1 m^2 area located perpendicular to the lines of induction of a magnetic field with an induction of 1 T . Thus

$$1 \text{ Wb} = 1 \text{ T} \times 1 \text{ m}^2$$

Another unit used for measuring flux is the *maxwell* (Mx): $1 \text{ Mx} = 10^{-8} \text{ Wb}$.

3. When conductor ab is moved with the velocity $v = \Delta x/\Delta t$, the emf induced in the circuit is $\mathcal{E} = Blv = Bl \Delta x/\Delta t$. The numerator of the fraction can be transformed as follows: $Bl \Delta x = Bl(x_2 - x_1) = BA_2 - BA_1 = \Phi_2 - \Phi_1 = -\Delta\Phi$ (since $A_2 < A_1$). The minus sign appeared because the current decreases. The induced emf is

$$\mathcal{E} = -\frac{\Delta\Phi}{\Delta t} \quad (43.10)$$

4. This expression was derived for the case when the conductor moves with respect to the source of the magnetic field. In this special case, equations (43.10) and (43.5) are equally valid; the emf can be calculated by either one. But expression (43.10) is also suitable for describing the phenomena of electromagnetic induction when the conductor is stationary and the electric field is induced by changes in the magnetic flux.

Thus in all cases the emf of electromagnetic induction *equals the rate of change of the magnetic flux taken with the opposite sign*. This is *Faraday's law of induction*.

43.8. LENZ'S LAW

1. Let us find the meaning of the sign in equation (43.10). We shall consider the emf to be positive if the direction of the induced electric field forms a system with the direction of vector \mathbf{B} that complies with the right-hand rule (Fig. 43.4a). Here the magnetic flux decreases as the conductor moves, the rate of change of the magnetic flux is $\Delta\Phi/\Delta t < 0$ and the emf is $\mathcal{E} = -\Delta\Phi/\Delta t > 0$.

The student will readily perceive that when the motion of the conductor is reversed, the magnetic flux increases, the rate of change of the magnetic flux is $\Delta\Phi/\Delta t > 0$ and the emf will have the opposite sign (Fig. 43.4b).

2. The sign of the emf and the corresponding direction of the induced current can be determined by means of a very general rule formulated in 1834 by H.F.E. Lenz (1804-1865). To derive this rule let us refer again to Fig. 43.4 where the dashed arrow indicates the direction of the induction vector \mathbf{B}^* of the magnetic field set up by the induced current. When the magnetic flux decreases, the directions of vectors \mathbf{B} and \mathbf{B}^* coincide, thereby opposing the reduction of the magnetic flux (Fig. 43.4a). In the second case, when the flux increases, vector \mathbf{B}^* is directed opposite to vector \mathbf{B} , thereby opposing the increase of the magnetic flux (Fig. 43.4b). In this way the *induced current will appear in such a direction that its magnetic field opposes the change in magnetic flux that produces the induction*. This is called *Lenz's law*.

43.9. ELECTROMAGNETIC INDUCTION AND THE LAW OF CONSERVATION OF ENERGY

1. Lenz's law is closely related to the law of conservation of energy and is actually a consequence of this general law of nature. Indeed, the induced electric field does work in moving charges along a circuit. But work is a measure of the change in energy (see Secs. 16.4 and 21.1). What changes in energy accompany the phenomenon of electromagnetic induction?

Let us refer once again to Fig. 43.4. Evidently, the motion of the conductor at velocity v is a result of the action of external force F , directed along the velocity. This induces a current in the conductor which, like any current, is subject to the Ampere force $F_m = iBl$ [equation (41.15)]. Since the conductor is moving at uniform velocity the external force and the Ampere force must be equal in magnitude ($F = F_m$) but of opposite direction.

2. Let us compare the work done by the external force with that done by the induced field in moving charges along the circuit. The work done by the external force is

$$W_{\text{mech}} = Fvt = iBlvt = qBlv \quad (43.11)$$

The work done by the induced field [equations (39.5) and (43.5)] is

$$W_{\text{el}} = q\mathcal{E} = qBlv \quad (43.12)$$

Hence, if there are no losses, the work done by the external force is found to be equal to that done by the induced field in moving charges along the circuit. Therefore, from the energy point of view, electromagnetic induction is the process of converting mechanical energy into the energy of the induced electric field. This principle is the basis for the operation of induction generators of electric current.

43.10. SELF-INDUCTION

1. Let us calculate the magnetic flux threading through the turns of a solenoid winding. The flux threading through one turn is equal, according to definition, to the product of the area enclosed in the turn by the induction of the magnetic field, i.e. BA . Since the solenoid has w turns, the total flux is

$$\Phi = wBA \quad (43.13)$$

But $w = nl$, where l is the length of the solenoid and n is the density of the winding. On the other hand, according to equations (40.32) and (42.8), $B = \mu\mu_0H = \mu\mu_0in$. Substituting into equation (43.13) we have

$$\Phi = \mu\mu_0n^2Ali = Li \quad (43.14)$$

2. The quantity L which for a long solenoid equals

$$L = \mu\mu_0n^2Al \quad (43.15)$$

is called the *inductance* of the solenoid. The unit of inductance is the *henry* (H). It is the inductance of a solenoid in which a current of one ampere sets up a flux of one weber:

$$1 \text{ H} = 1 \text{ Wb}/1 \text{ A}$$

If there are no ferromagnetic materials in the field or when such a material is in a state of magnetic saturation, the inductance of a solenoid is determined by its size and the density of the winding.

3. It is evident from equation (43.14) that when there is no change in the current in the solenoid winding, the magnetic flux threading through the winding will also remain unchanged and no induction phenomena will occur. If the current is changed, the flux is correspondingly changed and an induced emf will appear in the solenoid. Here the current induces an emf in its own circuit and the phenomenon is therefore called *self-induction*. We can determine this emf by combining equations (43.10) and (43.14):

$$\mathcal{E} = -\frac{\Delta\Phi}{\Delta t} = -\frac{\Phi_2 - \Phi_1}{t_2 - t_1} = -L \frac{i_2 - i_1}{t_2 - t_1} = -L \frac{\Delta i}{\Delta t}$$

Thus the self-induced emf is proportional to the rate of change of the current in the winding:

$$\mathcal{E}_L = -L \frac{\Delta i}{\Delta t} \quad (43.16)$$

The sign of the self-induced emf is determined by Lenz's law. If the current in the circuit is increasing, the induced field should oppose this increase, and therefore the sign of the self-induced emf should be opposite to the sign of the potential difference at the ends of the winding. If the current in the circuit is decreasing, the induced field should oppose this decrease, and therefore the sign of the emf is the same as that of the potential difference at the ends of the winding.

43.11. ENERGY OF AN ELECTROMAGNETIC FIELD

1. It was shown in Secs. 37.7 and 38.4 that an electric field stores energy which is distributed in it with the volume density

$$u_e = \frac{\Delta U_e}{\Delta V} = \frac{\epsilon \epsilon_0 E^2}{2} = \frac{D^2}{2\epsilon \epsilon_0}$$

We shall derive a similar expression for the energy of a magnetic field.

Suppose the current in a certain coil increases from zero to some value I . Accordingly, the magnetic flux increases from zero to the value $\Phi = LI$ (Fig. 43.5). As shown in the preceding section an emf will be induced which opposes the increase in current. Hence the external current source must do work to overcome the induced field.

As a result, the energy of the current source is converted into energy of the magnetic field. The change in the energy of the magnetic field is equal in magnitude to the element of work done in moving charge

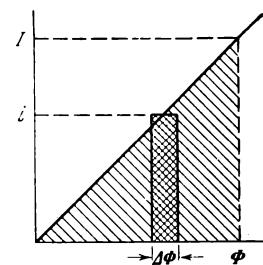


Fig. 43.5

Δq along the closed circuit:

$$\Delta U_m = -\Delta W = -\Delta q \mathcal{E} \quad (43.17)$$

But since $\mathcal{E} = -\Delta\Phi/\Delta t$ [equation (43.10)] and the current $i = \Delta q/\Delta t$ [equation (39.13)], we obtain

$$\Delta U_m = \Delta q \frac{\Delta\Phi}{\Delta t} = i\Delta\Phi \quad (43.18)$$

In the graph (Fig. 43.5) this quantity is represented by the area of the cross-hatched rectangle. The total energy of the field is equal numerically to the area of the hatched triangle. Thus

$$U_m = \frac{I\Phi}{2} = \frac{LI^2}{2} \quad (43.19)$$

2. Maxwell found that the energy of a magnetic field is spread throughout its whole volume. Therefore, the energy density is

$$u_m = \frac{\Delta U_m}{\Delta V} = \frac{\mu\mu_0 H^2}{2} = \frac{B^2}{2\mu\mu_0} \quad (43.20)$$

To derive this relationship we express the inductance in equation (43.19) in terms of its value in equation (43.15) and the current in terms of the field strength according to equation (40.32). Thus

$$U_m = \frac{LI^2}{2} = \frac{\mu\mu_0 n^2 V \times H^2}{2n^2} = \frac{\mu\mu_0 H^2}{2} V$$

Dividing both sides by the volume V we obtain equation (43.20).

3. The force with which a steel plate is attracted to the poles of an electromagnet can be calculated by analogy with the method used in Sec. 37.8 to find the force of interaction between the plates of a capacitor [equation (37.26)]. Thus

$$F = -\frac{B^2}{2\mu\mu_0} A \quad (43.21)$$

This expression is valid only for the case when the induction of the magnetic field in the gap between the core of the electromagnet and the steel plate is the same as the induction of the field in the core. In other words, if the field is not dissipated. To avoid such leakage the gap must be very narrow, much less than the diameter of the electromagnet core.

4. As we have repeatedly indicated, the division of a unified electromagnetic field into its electric and magnetic components depends upon the reference frame. Therefore the expression for the energy density of a magnetic field should be written in the form

$$u = \frac{\epsilon\epsilon_0 E^2}{2} + \frac{\mu\mu_0 H^2}{2} \quad (43.22)$$

Only in special cases that were dealt with above may one of the components equal zero.

43.12. CLOSING A CIRCUIT WITH INDUCTANCE

1. Let us arrange a direct current circuit consisting of two branches. In one branch we connect an incandescent lamp, a coil with a high inductance L and a resistor R . In the other we connect the same kind of lamp and a resistor with the same resistance. When we close the switch we shall see that in the branch with the resistor the lamp lights up instantly to its full brightness. In the other branch, with the coil, the lamp lights up gradually and reaches full brightness only after a certain time Δt has elapsed. Thus in a circuit having inductance, the current builds up gradually as shown in Fig. 43.6.

The reason for this is that a self-induced emf is produced as the current increases in the coil. According to Lenz's law this emf opposes the increase in the current. Ohm's law for this circuit is written as

$$i = \frac{\mathcal{E} + \mathcal{E}_L}{R} = \frac{\mathcal{E}}{R} - \frac{L}{R} \frac{\Delta i}{\Delta t}$$

The time Δt is required for the current to reach its maximum value $I_M = \mathcal{E}/R$, after which it ceases to rise further.

2. Let us estimate the time required for the current to build up. According to the law of conservation of energy, the work done by the current source in moving a charge along the circuit is partly expended in increasing the internal energy of the conductor (Joule heat) and partly to increase the energy of the magnetic field. Thus

$$W = Q + U_m$$

or

$$\mathcal{E} I_{av} \Delta t = I_{av}^2 R \Delta t + \frac{LI_M^2}{2}$$

where $I_M = \mathcal{E}/R$ = maximum current in the circuit

I_{av} = average current during the time Δt .

Assuming with some approximation that $I_{av} \cong I_M/2 \cong \mathcal{E}/2R$, we have

$$\frac{\mathcal{E}^2 \Delta t}{2R} \cong \frac{\mathcal{E}^2 R \Delta t}{4R^2} + \frac{LI_M^2}{2R^2}$$

It follows that the current builds up to its maximum value during a time interval equal to

$$\Delta t \cong 2L/R = 2\tau_L \quad (43.23)$$

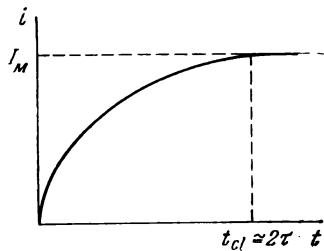


Fig. 43.6

Analogical to the case with the charging of a capacitor (see Sec. 39.8), the quantity $\tau_L = L/R$ is called the *inductive time constant* for a circuit consisting of a resistor and a coil. It is also called the *relaxation time*. This quantity characterizes the time required for a constant current to be established in the circuit.

Chapter 44

ELECTRICAL CONDUCTION IN SOLIDS

44.1. EXPERIMENTAL BASIS FOR THE ELECTRON THEORY OF CONDUCTION IN METALS

1. We have repeatedly resorted to the concept that conduction in metals is due to the availability of free electrons which move easily between the ions located at the points of the crystal lattice. These conceptions were based on the following experiments.

The first experiments were conducted in 1901 by E.C.V. Riekie (1845-1915). Three cylinders, two of copper and one of aluminium, with carefully ground end faces, were put end on end and connected to a current source.

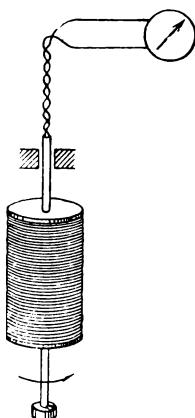


Fig. 44.1

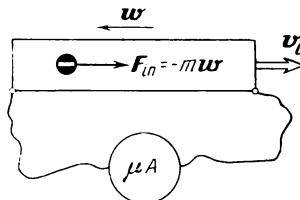


Fig. 44.2

A steady current of 0.1 A existed in the conductors (cylinders) for a whole year so that the total charge flowing through them exceeded 3.5×10^6 coulombs. Nevertheless, no chemical changes occurred in the conductors. This means that current in metals is due to the motion of particles that have nothing to do with the chemical structure of the metals. These particles cannot, of course, be the ions, which differ for various substances. They can only be electrons.

2. The decisive experiment was conducted in 1917 by two physicists, T. D. Stewart and R. C. Tolman (1881-1948), who measured the charge-to-mass ratio of particles that bring about conduction

in metals. The experiment consisted essentially in the following. A long conductor (wire) was wound on a spool and connected to the terminals of a galvanometer (Fig. 44.1). The spool was rotated rapidly and then suddenly braked.

During the braking process the reference frame linked to the conductor is noninertial. As has been previously shown (see Secs. 24.1, 24.2 and 24.3) a body in a noninertial reference frame is subject to the inertial force $F_{in} = -mw$, where w is the acceleration of the reference frame. Due to the action of this force the free particles in the metal will move forward in the direction of motion of the conductor in the same way as passengers are thrown forward when a railway car is suddenly braked. This will produce a current in the conductor of a direction depending on the sign of the particles (Fig. 44.2). Here the force of inertia is an extraneous force and thereby an emf. Thus

$$\mathcal{E} = \frac{W_{extr}}{q} = \frac{F_{in}l}{q} = -\frac{mwl}{q} \quad (44.1)$$

where q = charge of a free particle

m = its mass

l = length of the conductor.

Since the length of the conductor and the acceleration in braking (i.e. deceleration) are known, the charge-to-mass ratio of the particle could be determined by measuring the emf. It was found, however, that measurements of the emf involve considerable errors, and that the total charge Q , flowing through the galvanometer during the braking process, can be measured with much higher accuracy.

3. According to the definition of the current [equation (39.13)], $Q = it$, where t is the braking time. But, according to Ohm's law,

$$i = \frac{\mathcal{E}}{R} = -\frac{mwl}{qR}$$

where R is the total resistance of the conductor and galvanometer. The acceleration is

$$w = \frac{v - v_0}{t} = -\frac{v_0}{t}$$

because the conductor stops after being braked.

Hence the charge-to-mass ratio of the particle is

$$\frac{q}{m} = \frac{v_0 l}{QR} \quad (44.2)$$

The sign of the charge on the particles is determined from the sign of charge Q flowing through the galvanometer.

4. The experiment showed that free particles accomplishing conduction in metals are negatively charged. Their charge-to-mass ratios

were found to be: 1.60×10^{11} C/kg for copper, 1.49×10^{11} C/kg for silver and 1.54×10^{11} C/kg for aluminium. As we see, the deviation between the experimental data and the charge-to-mass ratio of the electron, equal to 1.76×10^{11} C/kg, is from 9 to 15 per cent. This is well within the limits of possible errors for such complicated experiments. Hence we can assert with sufficient assurance that current in metals is the result of the ordered motion of free electrons.

44.2. THE HALL EFFECT

1. In Sec. 39.2 we derived the relationship between the current density and the concentration of free electrons in the conductor.

It was found that this concentration could be measured by means of a phenomenon discovered in 1879 by E. H. Hall (1855-1938).

The Hall effect consists essentially in the following. We take a plate of homogeneous material (e.g., a metal) in the form of a parallelepiped and, by means of some current source, set up a uniform electric field in it along the axis of abscissas. This will produce a current in this direction. Any plane perpendicular to the x -axis is an equipotential surface, and the potential difference between faces 1 and 2 equals zero (Fig. 44.3).

If we now place the plate in a magnetic field so that the induction vector \mathbf{B} is parallel to the y -axis, a transverse potential difference is produced between faces 1 and 2. It is called the *transverse Hall potential difference* and equals

$$\Delta\varphi_H = R_H \frac{iB}{d} \quad (44.3)$$

where d = thickness of the plate

i = current

B = induction of the magnetic field

R_H = Hall coefficient.

2. The transverse potential difference is due to the following. The ordered motion of the electrons in the plate is in the direction perpendicular to the induction vector of the magnetic field; their average velocity is v . The moving electrons are subject to the sideways Lorentz force $F_m = eBv$ [equation (41.1)] which deflects them towards the upper face of the plate (Fig. 44.3). As a result, surplus charges are built up at the faces of the plate: a negative charge at the upper face and a positive charge at the lower. This sets up an electric field

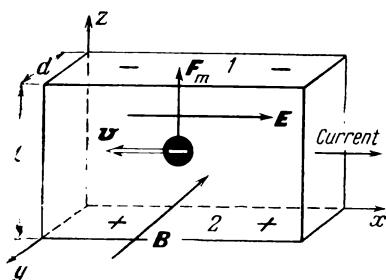


Fig. 44.3

of strength E_z along the z -axis that opposes the motion of electrons towards the upper face. Equilibrium is reached when the electric force counterbalances the Lorentz force. Hence, $F_m = F_e$ or

$$eB\bar{v} = eE_z \quad (44.4)$$

The Hall potential difference is

$$\Delta\Phi_H = E_z l = \bar{v} B l \quad (44.5)$$

If in equation (39.17) we express the average velocity of ordered motion of the electrons in terms of the current: $\bar{v} = \frac{i}{enA} = \frac{i}{endl}$, and substitute into equation (44.5), we obtain

$$\Delta\Phi_H = \frac{1}{en} \frac{iB}{d} \quad (44.6)$$

3. Combining equations (44.3) and (44.6) we find that the Hall coefficient is determined by the concentration of electrons and the charge of the current carriers:

$$R_H = \frac{1}{en} \quad (44.7)$$

Thus, after determining the Hall coefficient experimentally we can readily calculate the concentration of free electrons in a substance.

4. The mobility b of a particle is the ratio of its average velocity of ordered motion to the electric field strength [equation (39.26)]:

$$b = \frac{\bar{v}}{E} = \frac{\gamma}{en} = \gamma R_H \quad (44.8)$$

Hence if the conductivity and Hall coefficient of a metal are known we can calculate the mobility of the conduction electrons. The results of such experiments conducted on certain metals are listed in Table 44.1. Here ρ is the density of the metal, A is its atomic mass and N_A is Avogadro's number (see Sec. 26.9).

5. It is evident from the table that the monovalent metals lithium and sodium have an average of one free electron per atom. Trivalent aluminium has about three free electrons, and copper, which may be either monovalent or bivalent, has somewhat less than one and a half per atom.

The mobility of electrons is low and their velocity of ordered motion, or drift, is likewise low. Since the electric field in metals rarely exceeds 1 V/m, the average velocity of ordered motion $\bar{v} = bE$ is not over a millimetre per second, and is frequently much lower.

The Hall coefficient was measured in a wide temperature range. The test results demonstrated that in metals it does not depend on the temperature. Consequently, the concentration of free electrons is also independent of the temperature. This means that thermal motion plays no role in the formation of free electrons in metals.

Table 44.1

	Hall coefficient R_H , m ³ /C	Electrical conductivity ν , $\frac{1}{\text{ohm-m}}$	Concentration of conduction electrons $n = \frac{1}{eR_H} \cdot \frac{1}{\text{m}^3}$
Aluminium	0.30×10^{-10}	4.08×10^7	20.8×10^{28}
Copper	0.55×10^{-10}	6.41×10^7	11.4×10^{28}
Lithium	1.70×10^{-10}	1.17×10^7	3.67×10^{28}
Sodium	2.50×10^{-10}	2.48×10^7	2.50×10^{28}
Zinc oxide	1.6×10^{-5}	3.0×10^2	3.9×10^{28}
	Concentration of atoms $n_A = \frac{\rho N_A}{A} \cdot \frac{1}{\text{m}^3}$	Average number of free electrons per atom $v = \frac{n}{n_A}$	Electron mobility $b = \nu R_H \cdot \frac{m^2}{V-s}$
Aluminium	6.03×10^{28}	3.46	1.2×10^{-3}
Copper	8.49×10^{28}	1.34	3.5×10^{-3}
Lithium	4.60×10^{28}	0.80	2.0×10^{-3}
Sodium	2.54×10^{28}	0.99	6.2×10^{-3}
Zinc oxide	$\cong 4 \times 10^{28}$	$\cong 10^{-5}$	4.8×10^{-3}

6. Much more complex phenomena are observed when the Hall experiment is performed on substances called *semiconductors*. These include selenium, silicon, germanium and the oxides of many metals.

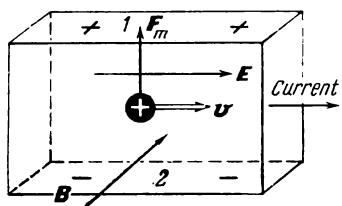


Fig. 44.4

Table 44.1 lists data on one semiconductor, zinc oxide. We see that its Hall coefficient is about 10^5 as much as for metals, its electrical conductivity is only 10^{-5} as much and its concentration of free electrons is also less in about the same proportion. Consequently, only a negligible number of atoms are ionized in this substance. But this is not the main point. The primary feature of semiconductors is

that the Hall coefficient drops sharply with a raise in temperature. Hence the concentration of free electrons increases with the temperature of the semiconductor. The second typical feature of semiconductors is that in some of them the Hall effect has the opposite sign. For the same directions of the current and magnetic field induction as shown in Fig. 44.3, the upper face of the plate is positively charged (Fig. 44.4). The only reasonable explanation of this fact is the assumption that

conduction in these substances is accomplished by the motion of positive charges rather than negative ones (see Sec. 77.2).

This leads us to a paradox that could not be explained by classical electronic theory. The quantum theory of semiconductor conductivity will be dealt with in more detail in Chapter 77.

44.3. ELECTRON GAS

1. P. Drude (1863-1906) developed his electron theory of conduction in metals in 1900. It was further developed on the basis of the Maxwell-Boltzmann distribution by H. Lorentz. According to the Drude-Lorentz theory, the totality of free electrons can be regarded as some ideal gas, called the *electron gas*. Drude and Lorentz postulated that the properties of electron gas are similar to those of an ordinary ideal gas.

As we shall see further on, the analogy between electron gas and an ideal gas implies extreme simplification of the matter and leads to many difficulties, especially for metals (see Ch. 75). Nevertheless, in the present chapter we shall deal with the totality of conduction electrons as an ideal gas. This enables us to explain by elementary methods the mechanism of various phenomena occurring in metals and semiconductors.

2. We shall calculate the average energy of thermal motion of the electrons by the same formula used to find the average kinetic energy of the molecules of an ideal monatomic gas [see equation (26.8)]. Thus

$$\bar{e} = \frac{m\bar{u}^2}{2} = \frac{3}{2} kT \quad (44.9)$$

At room temperature ($T \approx 300$ K), we obtain for the average velocity

$$\bar{u} = \sqrt{\frac{3kT}{m}} \approx \sqrt{\frac{3 \times 1.38 \times 10^{-23} \times 300}{9.1 \times 10^{-31}}} \approx 10^5 \text{ m/s}$$

According to quantum mechanical concepts, equation (16.26) should be employed to calculate the average kinetic energy of a conduction electron. Then the velocity of random motion of the conduction electrons is

$$u = \sqrt{\frac{2\bar{e}}{m_e}} = \sqrt{\frac{\hbar^2 n^{2/3} \times 2}{2m \times m}} = \frac{\hbar n^{1/3}}{m} \approx \frac{10^{-34} \times 10^{10}}{10^{-30}} \approx 10^6 \text{ m/s}$$

We obtained practically the same result as for the classical conceptions.

As we see, the average velocity of thermal motion of electrons exceeds their average velocity of ordered motion (see Secs. 39.2 and 44.2) by about 10^8 times! Proceeding from this fact we can conceive of the mechanism for the establishment of a current. While there is

no electric field in the conductor, the electrons move randomly in all possible directions. As soon as a field is set up the electrons begin to move slowly in the direction opposite to that of the lines of force of the field. This slow ordered motion is frequently called *drift*; it has practically no effect on the thermal motion of the particles.

3. We should not confuse the drift speed with the velocity of the electric current. Current velocity is the velocity at which electric energy is transmitted along wires. Experiments and theory indicate that the current velocity is very high; it is equal to the velocity of propagation of electromagnetic waves (see Sec. 59.1). Energy is transmitted from the source to the user by the electromagnetic field. The conductor with the drifting electrons serves only as a "guide" for the energy flux.

44.4. DERIVATION OF OHM'S LAW FROM ELECTRONIC THEORY

1. The model of the electron gas enables us to explain why Ohm's law holds in metals and semiconductors. Such a derivation is based on the following idea.

Suppose there is an electric field of strength E in a metal. Then the electron is subject to the force $F = eE$ and it will move with the acceleration $a = F/m = eE/m$ until it collides with an ion.

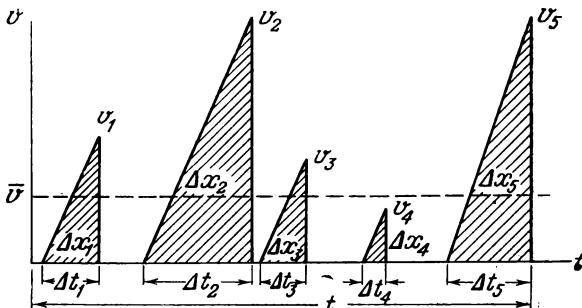


Fig. 44.5

Since the electron also participates in the thermal motion, its collisions with ions will not be regular. Moreover, we do not know how long the electron is in the bound state with the ion. Therefore the graph of its ordered motion can be represented approximately as shown in Fig. 44.5. During the time Δt_1 the electron passes a distance $\Delta x_1 = a\Delta t_1^2/2$ along the conductor; during the time Δt_2 , the distance $\Delta x_2 = a\Delta t_2^2/2$; etc. During a substantial time interval t it will travel the distance

$$x = \Delta x_1 + \Delta x_2 + \dots + \Delta x_n = \frac{a}{2} (\Delta t_1^2 + \Delta t_2^2 + \dots + \Delta t_n^2) \quad (44.10)$$

The average velocity of ordered motion of the electron along the conductor (drift speed) is

$$\bar{v} = \frac{x}{t} = \frac{at}{2} \quad (44.11)$$

where

$$\tau = \frac{\Delta t_1^2 + \Delta t_2^2 + \dots + \Delta t_n^2}{t}$$

is called the *mean free time* of the electron. Since $a = eE/m$, we have

$$\bar{v} = \frac{e\tau E}{2m} \quad (44.12)$$

Substituting the average drift speed [equation (44.11)] into equation (39.18) we obtain

$$j = \frac{e^2 n \tau}{2m} E = \gamma E \quad (44.13)$$

In metals and semiconductors the conductivity $\gamma = e^2 n \tau / 2m$ depends neither on the current density nor on the field strength. Hence equation (44.13) indicates that the current density in these substances is proportional to the field strength. But this is an expression of Ohm's law [equation (39.24)].

2. The mobility of the electron (see Sec. 39.5) is

$$b = \frac{\bar{v}}{E} = \frac{e\tau}{2m} \quad (44.14)$$

Using experimental data on electron mobility listed in Table 44.1 in equation (44.14) we can calculate the mean free time. It will be about 10^{-14} s. This means that the electron has about 10^{14} collisions per second with the ions of the crystal lattice.

3. From equation (44.13) we can obtain the expression for the conductivity:

$$\gamma = \frac{e^2 n \tau}{2m} = \frac{e^2 n \lambda}{2m u} \quad (44.15)$$

and for the resistivity

$$\rho = \frac{1}{\gamma} = \frac{2m}{e \cdot n \tau} = \frac{2m u}{e^2 n \lambda} \quad (44.16)$$

An analysis of expressions (44.15) and (44.16) explains some special features of metals and semiconductors, in particular, the dependence of the electrical conductivity on the temperature and on the structure of the crystal lattice.

44.5. THE CONDUCTIVITY OF METALS AND SEMICONDUCTORS

1. It is known from experiments that the electrical conductivity of metals is about five orders of magnitude greater than that of semiconductors. The reason for this enormous difference can be explained on the basis of electronic theory.

The distance between the crystal lattice points of all solids is about the same. Therefore, the mean free path of the conduction electrons cannot differ greatly in various solids. On the other hand, the Hall coefficient for semiconductors is 10^5 times as much as for metals. Consequently, the conduction electron concentration of metals is the same number of times greater. But their electrical conductivity is also this number of times larger (see Table 44.1), thereby fully agreeing with equation (44.15).

2. The resistivity of alloys is several times as much as that of pure metals (Table 44.2).

Table 44.2

		ρ (at 0 °C), ohm-mm ² /m	α (from 0° to 100 °C), 10^{-3} K^{-1}
Metals	Aluminium	0.0245	4.5
	Iron	0.089	6.5
	Copper	0.0156	4.3
	Platinum	0.0981	3.92
	Rhodium	0.043	4.4
	Nickel	0.0614	6.8
	Chromium	0.127	—
Alloys	Platinum-rhodium (90/10)	0.187	1.66
	Nickel-chromium (80/20)	1.03	1.0
	Steel	0.12 to 0.45	3.3 to 1.5
	Constantan	0.48	0.02

The reason why the electrical conductivity of a body depends upon the degree of perfection of its crystal lattice will be explained on the basis of quantum mechanical conceptions in Sec. 75.9. It was found that the presence of impurities and defects in the lattice, and certain features of the thermal vibrations of the particles in a solid reduce the mean free path of electrons, thereby reducing the electrical conductivity of the substance.

3. Experiments show that the resistivity of semiconductors is drastically reduced when the temperature is raised and when they contain impurities. In this they differ in principle from metals. The reason is that the conduction electron concentration in semiconductors considerably increases in the presence of impurities and when the

temperature is raised. This follows from experiments for determining the Hall coefficient of these substances. This increase in conduction electron concentration predominates over the effect of reducing the mean free path, and the resistivity decreases.

The reason why the concentration of conduction electrons does not depend on the temperature and impurities in metals, but does in semiconductors cannot be explained on the basis of classical electronic theory. Quantum mechanical concepts on this effect will be discussed in Chapter 75.

44.6. DERIVATION OF THE JOULE-LENZ LAW

1. In Sec. 39.7 we obtained equation (39.33) for calculating the change in internal energy of a conductor with a current. However, no light was thrown on the mechanism by which the conductor is heated. On the basis of electronic theory this phenomenon can be explained as follows.

An electron accelerated by the electric field acquires, at the end of its mean free path, the excess kinetic energy $K = mv_{\max}^2/2$, where $v_{\max} = a\tau$ is its maximum velocity of ordered motion. At the end of its free path the electron collides with an ion whose mass is several thousand times greater than that of the electron.

As a result of the collision, the excess energy is transmitted to the ion, which begins to vibrate more intensively. Thus, the increase in the internal energy of a conductor results from the bombardment of the ion lattice by electrons that have been accelerated by the field.

2. We shall calculate the quantity of heat involved in the following way. Let the length of the conductor be l and its cross-sectional area A . Then the number of conduction electrons in the conductor is $N = nAl$, where n is their concentration. The change in internal energy is equal to the product of the surplus kinetic energy K of the electron by the number N of electrons and by the number Z of collisions that the electron undergoes in time t . Thus

$$Q = NZK \quad (44.17)$$

The number of collisions is $Z = t/\tau$, where τ is the mean free time of the electron. Substituting into equation (44.17) and making the necessary transformations, we have

$$Q = \frac{ma^2\tau^2}{2} nAl \frac{t}{\tau} = \frac{e^2n\tau}{2m} E^2 Alt \quad (44.18)$$

But the field strength can be expressed in terms of the voltage over a length of the conductor, making use of equation (39.10): $E = u/l$. At the same time, according to equation (44.16), $\frac{e^2n\tau}{2m} = \frac{1}{\rho}$, where ρ

is the resistivity of the conductor. Thus

$$Q = \frac{A}{\rho l} u^2 t = \frac{u^2 t}{R} \quad (44.19)$$

This equation and equation (39.33) are identical.

3. The mechanism for the transmission of energy from the electron gas to the crystal lattice is not near as simple as it is described above. Why, indeed, in a collision with an ion, does the electron transmit to it precisely the excess energy of ordered motion, rather than less or more energy? We know that when there is no current the conduction electrons have a tremendous amount of kinetic energy that is evaluated by the formula for the energy of a localized particle (see Sec. 16.7). Since the conduction electron concentration in metals is $n \cong 10^{29} \text{ m}^{-3}$ (see Sec. 44.2), we have

$$\bar{K} = \frac{\hbar^2 n^{2/3}}{2m} \cong \frac{10^{-68} \times 2 \times 10^{19}}{2 \times 9 \times 10^{-31}} \cong 10^{-19} \text{ J} \cong 1 \text{ eV}$$

The maximum excess energy of ordered motion of the electron, produced by the action of electric fields with a strength of the order of $E \cong 1 \text{ V/m}$, is equal to

$$K_{\text{exc}} = \frac{mv^2}{2} = \frac{e^2 E^2 \tau^2}{2m} \cong \frac{2.56 \times 10^{-38} \times 1 \times 10^{-28}}{2 \times 9 \times 10^{-31}} \cong 10^{-38} \text{ J}$$

Hence the excess energy of ordered motion of the electron is an absolutely negligible fraction of its total energy ($K_{\text{exc}}/\bar{K} \cong 10^{-17}$). Why is the thermal effect due precisely to this excess energy?

On the other hand it can be shown on the basis of the laws of conservation of energy and momentum (see Secs. 17.2, 17.3 and 17.4) that in the collision of an electron with an ion only a small part of the kinetic energy K_{el} of the electron is converted into kinetic energy K_{ion} of the ion. It is evident that $\frac{\Delta K_{\text{ion}}}{K_{\text{el}}} = \frac{m}{M+m}$ for an

inelastic collision and $\frac{\Delta K_{\text{ion}}}{K_{\text{el}}} = \frac{4mM}{(m+M)^2}$ for an elastic collision where m and M are the masses of the electron and of the ion. Since $m/M \cong 10^{-4}$, we see that in both cases, practically, $\frac{\Delta K_{\text{ion}}}{K_{\text{el}}} \cong \frac{m}{M} \cong 10^{-4}$. Why is this not the part of the kinetic energy of the electron that is transmitted? Only 10^{-17} of its energy is transmitted, which is very much less.

4. Qualitatively, this phenomenon is explained as follows.

In the crystal lattice the conduction electrons form what would appear to be two "gases"—quantum and classical—which fill the same volume. Their temperatures differ essentially. The temperature of the lattice is of the order of 300 to 500 K; the temperature of the

electron gas is $T_{deg} = \hbar^2 n^{2/3} / 3km \approx 5000$ K (see Sec. 26.8). Notwithstanding this great difference in temperatures, there is no heat exchange between the electron gas and the lattice in the absence of current. The point is that, owing to the uncertainty relation, the energy of the conduction electrons cannot become less than $\bar{K} = \hbar^2 n^{2/3} / 2m$ [see equation (16.26)] which is the *minimum* energy of a localized particle.

Matters are considerably different when a current is established. Owing to the action of the electric field the electrons acquire excess kinetic energy of ordered motion. Even though this excess energy is extremely small and the corresponding increment of temperature of the electron gas is only

$$\Delta T_{el} = \frac{2\Delta K_{el}}{3k} = \frac{e^2 E^2 \tau^2}{3km} \approx \frac{2.56 \times 10^{-38} \times 1 \times 10^{-28}}{3 \times 1.38 \times 10^{-23} \times 9 \times 10^{-31}} \approx 10^{-13} \text{ K}$$

it is precisely this excess energy that violates thermodynamic equilibrium. The process of heat exchange is initiated between the "superheated" electron gas and the crystal lattice, and the excess energy obtained by the gas from the electric field is transmitted to the lattice. This heats up the conductor, i.e. it leads to the generation of Joule-Lenz heat.

44.7. CONTACT POTENTIAL DIFFERENCE

1. At the end of the 18th century A. Volta (1745-1827) observed that a contact potential difference is established between two different metals when they are in contact with each other. The origin of this potential jump can be explained by the electronic theory.

Let us consider two metals with different electron concentration ($n_1 > n_2$). Since the two metals have the same temperature at their place of contact, the average velocity of thermal motion of their electrons is also the same at this place. For simplicity let us suppose that their mean free paths are approximately the same as well. Owing to their thermal motion the electrons diffuse into the two metals (see Secs. 25.4 and 25.5). The number of electrons transferred from the first metal to the second is equal to one sixth of the electrons contained in a layer of thickness λ , i.e. $N_1 = \frac{1}{6} n_1 A \lambda$. The number of electrons transferred from the second metal to the first is $N_2 = \frac{1}{6} n_2 A \lambda$. But since we specified that $n_1 > n_2$, the second metal acquires more electrons than it gives up to the first metal.

Hence, one metal becomes positively charged as a result of diffusion, and the other negatively. This sets up an electric field that opposes diffusion. Diffusion in this case plays the same role as an extraneous field would at the place of contact. The flow of electrons

from one metal to the other ceases when the established potential difference counterbalances the emf of the extraneous field.

2. Let us qualitatively assess the contact potential difference from the following considerations. The current in the contact layer is

$$i = \frac{\Delta q}{\Delta t} = \frac{e(N_1 - N_2)}{\tau} = \frac{e\lambda A}{6\tau} (n_1 - n_2) = \frac{eA\bar{u}}{6} (n_1 - n_2)$$

According to Ohm's law the potential difference equals the product of the current by the resistance of the contact layers. This resistance is $R = R_1 + R_2 = (\rho_1 \lambda / A) + (\rho_2 \lambda / A)$, where ρ_1 and ρ_2 are the resistivities of the layers. Combining with equation (44.16) we have

$$R = \frac{2m\bar{u}^2}{e^2 A} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

Thus

$$\Delta\varphi = iR = \frac{m\bar{u}^2 (n_1 - n_2)}{3e} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \quad (44.20)$$

After substituting $3kT$ for $m\bar{u}^2$ and certain simple transformations we obtain

$$\Delta\varphi = \frac{kT}{e} \left(\frac{n_1}{n_2} - \frac{n_2}{n_1} \right) \quad (44.21)$$

We see that the contact potential difference depends only on the temperature and the chemical compositions of the metals. Quantum mechanical conceptions of this phenomenon are discussed in Sec. 78.1.

44.8. THERMOELECTRICITY

1. Let us arrange a closed circuit of two metals and maintain the same temperature at all points. We will find that under this condition there will be no current in this circuit due to contact potential jumps because its emf is zero:

$$\mathcal{E} = \Delta\varphi_{12} + \Delta\varphi_{21} = \frac{kT}{e} \left(\frac{n_1}{n_2} - \frac{n_2}{n_1} + \frac{n_2}{n_1} - \frac{n_1}{n_2} \right) = 0$$

This result could have been foreseen from the first and second laws of thermodynamics. Indeed, a current does work and therefore a source of energy is required to produce a current. This cannot be the energy of the environment because, according to the second law of thermodynamics, there is no heat exchange between bodies at the same temperature (see Ch. 28).

2. The result will be quite different if the temperatures of the junctions are not the same. Assume, for instance, that $T_1 > T_2$.

Then the expression for the emf becomes

$$\mathcal{E} = \Delta\varphi_{12} + \Delta\varphi_{21} = \frac{k}{e} \left(\frac{n_1}{n_2} - \frac{n_2}{n_1} \right) (T_1 - T_2) \quad (44.22)$$

The emf here is evidently not equal to zero, and a thermoelectric current is established in the circuit. This phenomenon was discovered in 1821 by T. Seebeck (1770-1831).

Equation (44.22) can be rewritten as

$$\mathcal{E} = \alpha (T_1 - T_2) = \alpha \Delta T \quad (44.23)$$

where $\alpha = \frac{k}{e} \left(\frac{n_1}{n_2} - \frac{n_2}{n_1} \right)$ = constant characterizing the properties of contact between the given two metals
 \mathcal{E} = thermoelectromotive force
 ΔT = difference in the temperature of the two junctions.

3. The thermoelectromotive force arising from the contact of two metals is very small. Though equation (44.22) was derived by means of quite crude approximations, it can be used for a qualitative estimate of this thermoelectromotive force. Taking $\Delta T = 1$ K and the concentrations n_1 and n_2 , for instance, from the data in Table 44.1 for copper and aluminium, we obtain

$$\begin{aligned} \alpha &= \frac{k}{e} \left(\frac{n_1}{n_2} - \frac{n_2}{n_1} \right) \cong \frac{1.38 \times 10^{-23}}{1.61 \times 10^{-19}} \left(\frac{21}{11} - \frac{11}{21} \right) \\ &\cong 1.2 \times 10^{-4} \text{ V/K} = 120 \text{ muV/K} \end{aligned}$$

Experimental data for a thermoelectric couple of these metals is only 1/30 as much: $\alpha = 3.4$ muV/K.

If we replace one of the metals by a semiconductor, the thermoelectromotive force should substantially increase because the electron concentrations in metals and semiconductors differ by 4 or 5 orders of magnitude. Equation (44.22) is inapplicable here, even for qualitative calculations, but the fact is confirmed experimentally that a metal-semiconductor thermocouple gives rise to a much higher thermoelectromotive force than a metal-metal one.

4. Thermocouples are commonly employed for temperature measurements, especially where the application of liquid thermometers (for instance, mercury thermometers) is excluded. Advantageous features of thermocouple thermometers are their high sensitivity; the possibility of connecting them to recorders that register the variation in temperature with time; small size, low inherent heat capacity and their extensive range (from $+2000^\circ\text{C}$ to -200°C).

Thermocouples can also be used as current sources. True, they have a low efficiency (less than 1 per cent) and a low emf. If, however, certain definite types of semiconductors are used and the ther-

mocouples are connected in series to form a thermopile, an emf of the order of several dozens of volts can be obtained. Such thermopiles were developed under the leadership of Academician A. F. Ioffe. They can operate from any heat source, for instance a kerosine lamp, and deliver power sufficient to supply a radio receiver.

The *Romashka* (daisy) nuclear power installation operates on the same principle. The *Romashka* is a small-size nuclear reactor designed to operate on 49 kg of uranium. The temperature in the active zone reaches 1770°C. Its walls are lined with several thousand silicon-germanium plates, which serve as semiconductor thermocouples and are connected together to form a thermopile. The thermoelectromotive force arises from the difference in the temperature inside the reactor and that of the outside atmosphere. This installation, developed in 1964 in the Kurchatov Institute of Atomic Energy under the leadership of Academician M. D. Millionshchikov (1913-1973), successfully passed all performance tests and operates very efficiently.

5. From the viewpoint of thermodynamics a thermocouple is an analogy of a heat engine (see Ch. 29). Here we have two bodies at different temperatures, one serving as the heater (or heat source) and the other as the cooler (or heat sink), and the working medium is the electron gas. In contrast to an ordinary heat engine where a part of the internal energy of the heater is converted into mechanical energy, in the thermocouple it is converted into energy of an electric current.

6. We know that if we deliver mechanical energy from an external source to a heat engine it will operate as a heat pump or refrigerator (see Sec. 29.8). By analogy we could expect a difference in temperatures between the junctions if we pass a current from an external source through a thermocouple. Such an effect is actually observed. It is called the *Peltier effect* in honour of J. Peltier (1785-1845) who discovered it in 1834.

To obtain a more significant temperature difference, semiconductor thermocouples should be employed. Thermoelectric refrigerators using the Peltier effect in semiconductors were developed in 1954 by a group of scientists of the Leningrad Institute of Semiconductors under the leadership of Academician A. F. Ioffe.

44.9. THE WORK FUNCTION

1. Electrons moving about in a metal sometimes escape beyond its bounds, forming an "electron cloud" above the metal. A part of these electrons return to the metal, and others escape again. This phenomenon is very much like the evaporation of liquids (see Secs. 35.1 and 35.2). The surface of the metal and the electron cloud form a double electric layer (barrier layer) similar to a parallel-plate capacitor. This layer is several interatomic distances thick ($d \cong 10^{-10}$ to

10^{-9} m). The potential difference in this layer is called the *surface potential jump at the metal-vacuum boundary*, or the *contact-potential difference* between the metal and a vacuum.

This quantity can be evaluated by the following approximate calculation. Suppose that an electron is emitted from a metal. This will cause a positive charge of a magnitude equal to that of the electron to appear in the metal. This positive charge is called the electrostatic image of the electron (Fig. 44.6).

It was found that the interaction of the electron with the metal is equal to the interaction between the electron and its image. Thereby, it is reduced to the interaction of point charges which we know how to calculate. According to equation (37.4) we have for the contact potential jump (at $d \cong 5 \times 10^{-10}$ m)

$$\varphi_{\text{con}} \cong \frac{e}{4\pi\epsilon_0 \times 2d} \cong \frac{1.6 \times 10^{-19} \times 36\pi \times 10^9}{4\pi \times 2 \times 5 \times 10^{-10}} \cong 1.4 \text{ V}$$

2. For an electron in a metal to escape beyond the bounds of the metal surface it must do work against the forces of attraction of its image and against the forces of repulsion of the negatively charged electron cloud. This work is called the *work function* W_0 . It is equal, therefore, to the minimum energy that must be imparted to a conduction electron to enable it to emerge from the metal into a vacuum. We can assume that the work function is

$$W_0 = e\varphi_{\text{con}} \quad (44.24)$$

Our approximate calculation yields to value $W_0 \cong 1.5 \text{ eV} = 2.4 \times 10^{-19} \text{ J}$ for the work function. This result is correct insofar as the order of magnitude is concerned, as is evident from the table:

	$W_0, \text{ eV}$		$W_0, \text{ eV}$
Lithium, sodium, potassium	2.3	Barium	2.5
Zinc	4.2	Calcium	2.7
Tungsten	4.5	Cesium	1.9
Platinum	5.3	Barium on tungsten	1.1
		Cesium on tungsten	1.4

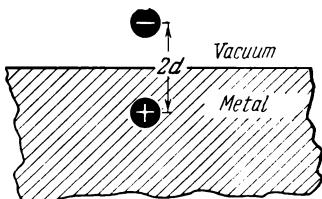


Fig. 44.6

3. The potential energy of a conduction electron in a metal is less than the energy of a free electron. Assuming that the potential energy of the electron in a vacuum equals zero, its energy in a metal is $U = -W_0 = -e\varphi_{\text{con}}$. This energy can be represented graphically,

plotting the coordinate of the electron on the axis of abscissas and its potential energy on the axis of ordinates (Fig. 44.7). Here the thickness d of the "electron cloud" is greatly exaggerated. The graph has the shape of a trough and is called the "potential trough".

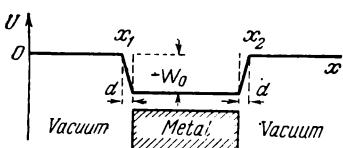


Fig. 44.7

Therefore, when we wish to indicate that the electron in a metal has negative potential energy we usually say it is in a potential trough.

4. For a conduction electron to escape from a metal its kinetic energy must be greater (or, at least, equal to) the work function. The electron may acquire this energy in various ways.

One way is by allowing light to fall on the metal. After receiving energy from the light wave the electron may have sufficient energy (equal to the work function) to escape from the metal. This phenomenon is called *photoelectric emission* and will be discussed in Secs. 68.1, 68.2 and 68.3.

A second way is to bombard the metal surface with particles having an energy of several hundred electron volts.

The third way is to raise the temperature of the metal. The emission of electrons by a heated body is called *thermionic emission* (see Sec. 47.1).

Chapter 45

HEAT CAPACITY AND THERMAL CONDUCTIVITY OF SOLIDS

45.1. HEAT CAPACITY

1. It was shown in Secs. 27.4 and 27.5 that in gases it is necessary to distinguish between the specific heat at constant volume c_V and at constant pressure c_p , which differ considerably ($\gamma = c_p/c_V \cong 1.7$ to 1.4). These specific heats differ only slightly in solids: from 3 per cent for copper and gold to 8 per cent for alkali metals. For this reason we shall not differentiate in the following between these quantities, and simply refer to the heat capacity or specific heat of a solid.

In discussing the expression for the internal energy of an ideal gas we came to the conclusion that it can be represented by the sum of the kinetic energies of the molecules plus a constant addend, because the sum of the potential energies of the molecules and their rest energy is a constant (see Sec. 27.1). Matters are quite different in a solid. Here the particles vibrate at the points of a crystal lattice,

and in this vibrational process kinetic energy is continuously being converted into potential energy and vice versa. Hence the average values of the kinetic and potential energy are equal in a solid (see Sec. 49.3). Thus

$$\bar{\varepsilon}_{pot} = \bar{\varepsilon}_{kin} = \frac{3}{2} kT \quad (45.1)$$

The average vibrational energy of the particle is the sum of its potential and kinetic energies:

$$\bar{\varepsilon} = \bar{\varepsilon}_{kin} + \bar{\varepsilon}_{pot} = 3kT \quad (45.2)$$

2. The internal energy of one kilomole of a substance is equal to the product of the average vibrational energy of one atom by Avogadro's number. Using equation (45.2) we obtain

$$U_m = N_A \bar{\varepsilon} = 3N_A kT = 3RT \quad (45.3)$$

where $R = 8.31 \times 10^3 \text{ J/kmol-K} \cong 1.98 \text{ kcal/kmol-K}$ is the universal gas constant (see Sec. 26.9).

By definition the heat capacity per kilomole equals

$$C_m = \frac{U_2 - U_1}{T_2 - T_1} = 3R \cong 5.94 \frac{\text{kcal}}{\text{kmol-K}} \quad (45.4)$$

This is the *Dulong and Petit law*, established in 1819 by P. L. Dulong (1785-1838) and A. T. Petit (1791-1820). It states that

the heat capacity per kilomole of chemically simple crystalline bodies is approximately 6 kcal per kilomole-kelvin.

3. The *specific heat* is the ratio of the heat capacity per kilomole to the atomic mass of an element:

$$c = \frac{C_m}{A} = \frac{3R}{A} = \frac{3N_A k}{N_A m_0} = \frac{3k}{m_0} \quad (45.5)$$

where m_0 is the mass of the atom.

The *heat capacity per unit volume* is the ratio of the heat capacity per kilogram atomic weight to the volume of one kilogram atomic weight:

$$C = \frac{C_m}{V_m} = \frac{3N_A k}{N_A v} = \frac{3k}{v} \quad (45.6)$$

where v is the volume of a unit cell of the crystal lattice.

The heat capacity per kilogram atomic weight of various substances, both insulators and metals, is shown in Fig. 45.1. An analysis of experimental data leads us to the conclusion that the Dulong and

Petit law is very approximate; it holds only at sufficiently high temperatures. At low temperatures, the heat capacity of solids decreases rapidly and tends to zero at absolute zero.

4. Classical theory cannot explain the dependence of the heat capacity of solids on the temperature. This was first explained in 1905 by Einstein on the basis of quantum mechanical conceptions. The equation he derived was qualitatively correct in describing the dependence of heat capacity on temperature shown in Fig. 45.1. A more exact theory, also based on quantum mechanical conceptions, was developed in 1912 by P. W. Debye

(1884-1966). He showed, in particular, that each substance has a definite characteristic temperature Θ (*Debye temperature*), beginning with which the Dulong and Petit law is valid.

Thus

$$\Theta = \frac{2\hbar a}{kd} \quad (45.7)$$

where a = velocity of sound

d = characteristic distance between particles in the crystal lattice.

Let us calculate the Debye temperature for aluminium ($a = 6400$ m/s and $d = 4$ Å). Thus

$$\Theta \cong \frac{2 \times 10^{-34} \times 6400}{1.38 \times 10^{-23} \times 4 \times 10^{-10}} \cong 230 \text{ K}$$

which agrees well with experimental data (see Fig. 45.1). The low values of the heat capacity of the diamond at quite high temperatures indicates that its Debye temperature is much higher than 1200 K (actually $\Theta \cong 2000$ K). This is due to the very high velocity of sound in the diamond and the small distances between the atoms ($a = 16\ 000$ m/s and $d = 1.54$ Å).

45.2. HEAT CAPACITY OF METALS

The fact that the Dulong and Petit law is applicable to metals and semiconductors is entirely unexpected. In deriving equation (45.3) we assumed that the internal energy of the crystal is determined by only the vibrational energy of the particles at the lattice points. But metals and semiconductors contain conduction electrons whose

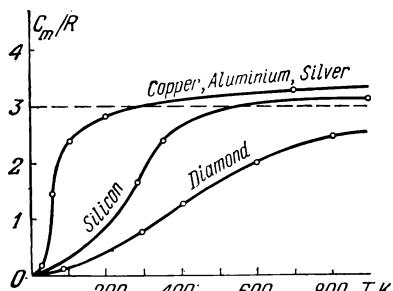


Fig. 45.1

totality we regard as an ideal gas. Consequently, the internal energy should be equal to the sum of the energy of the lattice plus that of the electron gas.

Assuming that the average energy of thermal motion of the electrons is $\bar{\varepsilon}_{el} = \frac{3}{2}kT$ and that the number of free electrons per atom is v , we obtain for one kilogram atomic weight:

$$U_m = N_A \bar{\varepsilon} + v N_A \bar{\varepsilon}_{el} = N_A \left(3kT + \frac{3}{2} v kT \right) = 3RT \left(1 + \frac{v}{2} \right)$$

From which we obtain the expression for the molar specific heat:

$$C_m = 3R \left(1 + \frac{v}{2} \right) \quad (45.8)$$

Using the data of Table 44.1 we will find that the contribution of conduction electrons to the heat capacity of semiconductors is negligibly small, because here $v \approx 10^{-5}$. As to metals (see Sec. 75.3), their heat capacity should considerably exceed that of insulators. As a matter of fact, according to equation (45.8), we obtain $C_m = 7.4R$ for aluminium, $C_m = 5R$ for copper, $C_m = 4.2R$ for lithium, etc. This, however, does not agree with experimental data which indicate that the Dulong and Petit law is valid for metals as well.

45.3. THERMAL CONDUCTIVITY OF INSULATORS

1. If a certain temperature difference is maintained at the ends of a rod, energy in the form of heat will be transmitted from the hot end to the cold one. Let the length of the rod be l and its cross-sectional area A . The temperature of the heated part is denoted by T_1 ; that of the cold part by T_2 . The rod is located along the axis of abscissas. The quantity $\frac{\Delta T}{\Delta x} = \frac{T_2 - T_1}{x_2 - x_1}$ is called the *temperature gradient*. It shows how fast the temperature varies along the rod. The quantity $\frac{\Delta Q}{A \Delta t}$ indicates the amount of heat transferred across unit cross-sectional area in unit time; it is called the *heat flow*.

It was found that the *heat flow is proportional to the temperature gradient* (Fourier's law of heat conduction):

$$\frac{\Delta Q}{A \Delta t} = -K \frac{\Delta T}{\Delta x} \quad (45.9)$$

The minus sign appeared because the heat flow is a positive quantity and the temperature gradient is negative.

The quantity K is called the *thermal conductivity* of the substance through which heat is transferred. The SI unit of thermal conductivi-

ty is the $\frac{\text{watt}}{\text{m-kelvin}}$ (W/m-K). Nonsystem units are

$$1 \frac{\text{kcal}}{\text{h-m-deg C}} = 1.16 \frac{\text{W}}{\text{m-K}}$$

$$1 \frac{\text{cal}}{\text{s-cm-deg C}} = 419 \frac{\text{W}}{\text{m-K}}$$

Typical values of the thermal conductivity of some insulators are listed in Table 45.1.

Table 45.1

Substance	Velocity of sound a , km/s	Thermal conducti- vity K , W/m-K	d , Å	$\frac{3ka}{d^2}$
Plastics	2.0 to 2.6	0.1 to 0.4	5	0.3
Brick	3.6	0.6 to 1.1	—	—
Glass	5.7 to 4.3	1.0 to 0.8	5	0.9 to 0.7
Sodium chloride	4.75	7.1	2.82	2.4
Ice (-5°C)	3.23	2.2	3.2	1.3
Water (20°C)	1.47	0.61	3.1	0.64
Acetone (25°C)	1.17	0.16	5	0.19
Carbon tetrachloride	0.93	0.11	5	0.15

2. To analyze the mechanism of thermal conduction we shall consider two adjacent layers of atoms (Fig. 45.2). Let the temperature of the left layer be $T_1 = T$ and that of the right layer $T_2 = T - \Delta T$.

Particles located in these layers have the corresponding average energies $\bar{\epsilon}_1 = 3kT$ and $\bar{\epsilon}_2 = 3k(T - \Delta T)$. Since the vibration of the particles in one layer is more intensive than in the other, energy is transmitted from the left layer to the right one.

One particle transmits the energy $\Delta\epsilon = \bar{\epsilon}_2 - \bar{\epsilon}_1 = -3k\Delta T$. A layer of area A contains $N = A/d^2$ particles, where d is the distance between the particles.

Consequently, the quantity of heat is $\Delta Q = N\Delta\epsilon = -3kA\Delta T/d^2$. The heat flow is

$$\frac{\Delta Q}{A\Delta t} = -\frac{3k\Delta T}{d^2\Delta t} = -\frac{3k}{d^2} \frac{\Delta x}{\Delta t} \frac{\Delta T}{\Delta x} = -\frac{3ka}{d^2} \frac{\Delta T}{\Delta x} \quad (45.10)$$

The quantity $a = \Delta x/\Delta t$ is the rate of transfer of thermal vibrations. In principle, it does not differ from the rate of transfer of

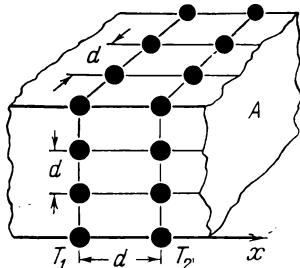


Fig. 45.2

any other elastic vibrations from layer to layer, i.e. from the velocity of sound.

Combining equations (45.9) and (45.10) we obtain for the thermal conductivity

$$K = \frac{3ka}{d^2} = \frac{3kad}{v} = Cad \quad (45.11)$$

where C is the heat capacity per unit volume of the substance (see Sec. 45.1).

Expression (45.11) enables us to explain qualitatively the anisotropy of heat conduction in certain monocrystals (see Sec. 32.1). The main point is that the distances between atoms and the velocity of sound differs in different directions in monocrystals. Consequently, the thermal conductivity also differs.

3. Thermal conductivities calculated by equation (45.11) are listed in the last column of Table 45.1. We see that for plastics, amorphous bodies and liquids, theory agrees well with experimental data. But for crystals the experimental values are several times larger than the theoretical ones.

This can be explained by using considerations advanced by Debye in 1914. Sound waves sensed by the human ear have a frequency in the range from about 20 Hz to 20 kHz, ultrasonic frequencies range from 20 kHz to 100 MHz. These waves are practically unscattered when they propagate in crystals, just as light is not scattered when it passes through transparent media such as glass or water. Thermal vibrations have much higher frequencies, of the order of 10^{10} to 10^{13} Hz (hypersonics). It was found that such high-frequency waves are intensively scattered in crystals like light is scattered in turbid (cloudy) media. The more intense the scattering of sound waves of a frequency corresponding to the thermal vibrations of the crystal, the slower the heat transfer process and the lower the thermal conductivity of the substance. Starting from these conceptions, Debye derived for the thermal conductivity:

$$K = \frac{1}{3} Ca\Lambda \quad (45.12)$$

where Λ is the average distance a sound wave travels before being scattered. If we introduce the concept that sound waves propagate in the form of elementary portions called sound field quanta, or *phonons*, then Λ is the *mean free path of a phonon*. Combining equations (45.11) and (45.12) we find that $\Lambda \cong 3d$.

4. Sound waves are scattered by irregularities of the crystal lattice, in a way similar to the scattering of conduction electrons (see Sec. 44.5). Irregularities of the crystal lattice are due to its defects and to the special features of thermal vibrations of the particles at the lattice points (see Sec. 75.9). In amorphous bodies and

liquids there is short-range order and, consequently, a great many defects. For this reason, their thermal conductivity is low, as is confirmed experimentally (see Table 45.1). In crystals, where the packing order is considerably better and there are less defects, a wave can propagate without being scattered a much greater distance. Calculations yield the values $\Lambda = 23 \text{ \AA} \cong 8d$ for sodium chloride at 0°C , $\Lambda = 40 \text{ \AA} \cong 13d$ for quartz along the optic axis and at the same temperature, etc. This explains the discrepancy between experimental data and our elementary calculation for crystals by equation (45.11).

Thermal vibration of the particles at the points of the crystal lattice is intensified with a rise in the temperature of the substance, and the distance the sound waves propagate without scattering is reduced. This should lead to a reduction in the thermal conductivity in the same way as the electrical conductivity is decreased when the temperature is raised (see Sec. 44.5). Experiments confirm this conclusion, but at relatively high temperatures. At low temperatures, thermal conductivity is reduced due to the sharp drop in heat capacity (see Sec. 45.1).

45.4. THERMAL CONDUCTIVITY OF METALS

1. Comparative data on the thermal conductivity and electrical conductivity of certain metals at $T = 273 \text{ K}$ are listed in Table 45.2.

We see that the thermal conductivity of metals is very high. It does not reduce to the thermal conductivity of the lattice; therefore, there must be another mechanism here for heat transfer. It was found that in pure metals heat conductivity is practically due entirely to the electron gas. The contribution of the thermal conductivity of the lattice is appreciable only in highly impure metals and alloys in which electric conductivity is low.

Table 45.2

Substance	$K, \text{ W/m-K}$	$\gamma, \text{ ohm}^{-1}\text{-m}^{-1}$	$K/\gamma T$
Copper	385	64.1×10^6	2.2×10^{-8}
Aluminium	228	40.8×10^6	2.1×10^{-8}
Sodium	135	23.8×10^6	2.1×10^{-8}

For the derivation we shall consider two layers of metal, each of a thickness equal to the mean free path of the electron. During the time $\tau = \lambda/u$, $N = \frac{1}{6} nA\lambda$ electrons pass from one layer to the

other. Electrons going from the left-hand layer to the right-hand one carry the energy $W_1 = N\bar{\epsilon}_1$. The counterflow carries the energy $W_2 = N\bar{\epsilon}_2$. According to classical electronic theory, the average energy of the electron is $\bar{\epsilon} = \frac{3}{2}kT$. Hence

$$Q = W_1 - W_2 = N(\bar{\epsilon}_1 - \bar{\epsilon}_2) = \frac{1}{6}nA\lambda \times \frac{3}{2}k(T_1 - T_2) \quad (45.13)$$

The heat flow is

$$\frac{\Delta Q}{A\Delta t} = \frac{1}{4}n\lambda k \frac{T_1 - T_2}{x_2 - x_1} \times \frac{x_2 - x_1}{\Delta t}$$

Since $x_2 - x_1 = 2\lambda = 2\bar{u}\Delta t$, where \bar{u} is the average velocity of thermal motion of the electrons, then

$$\frac{\Delta Q}{A\Delta t} = -\frac{1}{2}n\lambda\bar{u}k \frac{\Delta T}{\Delta x} \quad (45.14)$$

Combining equations (45.14) and (45.9) we have the expression for the thermal conductivity of the electron gas:

$$K = \frac{1}{2}n\lambda\bar{u}k \quad (45.15)$$

2. We cannot calculate the thermal conductivity by this equation because we do not know the mean free path of the electron. However, dividing equation (45.15) by equation (44.15) we obtain

$$\frac{K}{\gamma} = \frac{n\lambda\bar{u}k \times 2m\bar{u}}{2e^2n\lambda} = \frac{k m \bar{u}^2}{e^2}$$

But $m\bar{u}^2 = 3kT$ and therefore

$$\frac{K}{\gamma} = \frac{3k^2T}{e^2} = 2.23 \times 10^{-8}T \quad (45.16)$$

This expression is called the Wiedemann-Franz law in honour of G. H. Wiedemann (1826-1899) and R. Franz who discovered in 1853 that at any temperature the ratio of the thermal conductivity to the electric conductivity is the same for all metals. The dependence of this ratio on the temperature was shown by Lorentz.

3. Comparing the experimental data of Table 45.2 with equation (45.16) we can see that the thermal conductivity of metals is determined practically by only the thermal conductivity of the electron gas which, at temperatures far from absolute zero, is hundreds of times greater than that of the crystal lattice.

As we see, classical electronic theory not only gives a qualitative explanation of the mechanism of thermal conductivity in metals, but also qualitative relationships, the Wiedemann-Franz law, which agrees well with experiments.

4. Let us sum up. The idea of Drude and Lorentz proposing that the totality of conduction electrons in metals be regarded as an ideal gas turned out to be especially fruitful. It proved possible on this basis to explain many phenomena associated with the electrical and thermal conduction of metals. In all cases we assumed that the electron gas complies with Maxwell-Boltzmann statistics like any ordinary monatomic ideal gas. As a result, the average energy of thermal motion of a free electron is found to be $\bar{\epsilon}_{el} = \frac{3}{2} kT$.

But this assumption which, for instance, excellently explains special features of the thermal conductivity of metals (Wiedemann-Franz law), does not yield correct values for the heat capacity of metals. The validity of the Dulong and Petit law for metals can be explained if we further assume that the electron gas does not participate in heat capacity. But then we will be unable to explain features of the thermal conduction of metals.

This contradiction indicates that classical electronic theory is unsound in principle. Only the ideas of quantum mechanics were capable of solving this problem. We shall return to it in Chapter 75.

Chapter 46

ELECTRICAL CONDUCTIVITY OF ELECTROLYTES

46.1. ELECTROLYTIC DISSOCIATION

1. Experiments reveal that *electrolytes*—solutions of salts, acids and alkalis in water—are good conductors of electric current. Current in electrolytes is accompanied by *electrolysis*, which is the deposit of substances on electrodes immersed in the electrolyte. In contrast to metals, which have electron conduction, electrolytes are conductors in which the current is due to the motion of *ions*.

Neither distilled water nor the substance to be dissolved (while still in the solid state) have any free ions, and are therefore good insulators. Consequently, the ions appear only as a result of interaction between the molecules of the dissolved substance and the water molecules. The process of decomposition of the molecules of the dissolved substance (the solute) into ions by the action of the solvent is called *electrolytic dissociation*.

2. The reason for this dissociation is that water molecules have a large dipole moment ($p_e = 6.1 \times 10^{-30}$ C-m) as a result of which a strong electric field is set up around the molecule at distances of the order of the distance between molecules in liquids ($r \cong 1 \text{ \AA} =$

$= 10^{-10}$ m). Combining equations (10.11) and (37.9) we have

$$\varphi \cong \frac{p_e}{4\pi\epsilon_0 r^2} \cong \frac{6 \times 10^{-30} \times 36\pi \times 10^9}{4\pi \times 10^{-20}} \cong 6 \text{ V}$$

The energy of interaction between an ion of sodium or chlorine and a water molecule is approximately equal to that between ions in the NaCl molecule. This is why the molecules are decomposed into ions by thermal collisions when salt is dissolved. Thus



The arrows indicate that the process proceeds in both directions: along with the *dissociation* of molecules into two ions there is also the reverse process of *recombination* of the ions into a neutral molecule.

Positive ions moving to the cathode are called *cations*; negative ions are called *anions*, i.e. moving toward the anode.

It should be noted that if the molecules of the dissolved substance do not dissociate into ions, the solution will not be a conductor. Examples are aqueous solutions of sugar and glycerin. They are good insulators.

3. The molecules of water envelope each ion, forming a *solvation sheath* (Fig. 46.1). In the first place this hinders recombination of the ions, and at low concentrations all the solute molecules in the solution are dissociated. In the second place, the motion of the ions meets with great difficulties. Actually, it is not the ion but the solvate moving about in the solution. Such a solvate is a sphere consisting of an ion enveloped by the solvation sheath in which there are several layers of solvent molecules.

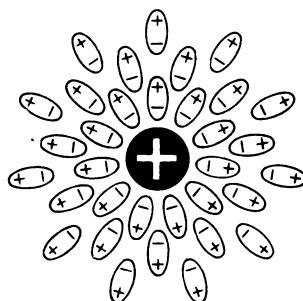


Fig. 46.1

46.2. OHM'S LAW AND THE ELECTRICAL CONDUCTIVITY OF ELECTROLYTES

1. If two electrodes are immersed in an electrolyte and a potential difference is applied by connecting them to a current source, ordered motion of the solvates will be initiated. Positively charged solvates will move toward the cathode with the velocity v_+ and negatively charged solvates toward the anode with the velocity v_- . By analogy with equation (39.18) we can write the expression for the current density in an electrolyte:

$$j = q_+ n_+ v_+ + q_- n_- v_- \quad (46.1)$$

where q = charge of the ion

n = ion concentration.

It follows from the law of conservation of electric charge that the sum of the charges of the positive and negative ions equals zero. Hence

$$q_+n_+ = q_-n_- = qn = \alpha qn_0 \quad (46.2)$$

where n_0 is the concentration of the molecules of the dissolved substance. The dissociation coefficient $\alpha = n/n_0$ is the ratio of the concentration of ions to the concentration of molecules of the substance.

The current density is

$$j = qn(v_+ + v_-) \quad (46.3)$$

2. The velocity of ordered motion of the solvate can be determined from the following considerations. The solvate is subject to the electric force $F_e = qE$ and to the friction drag $f = 6\pi\eta rv$, where r is the radius of the solvate and η is the viscosity of the liquid (see Sec. 11.8). The solvate moves at constant velocity when the friction drag counterbalances the electric force: $qE = 6\pi\eta rv$. Then the velocity of the solvate is

$$v = \frac{qE}{6\pi\eta r} \quad (46.4)$$

and its mobility [see equation (39.26)] is

$$b = \frac{v}{E} = \frac{q}{6\pi\eta r} \quad (46.5)$$

Combining equations (46.5) and (46.3) we find that Ohm's law should hold for electrolytes. Thus

$$j = qn(b_+ + b_-)E = \gamma E \quad (46.6)$$

This is confirmed by experiment.

3. The electrical conductivity of an electrolyte is

$$\gamma = qn(b_+ + b_-) = \frac{\alpha q^2 n_0}{6\pi\eta} \left(\frac{1}{r_+} + \frac{1}{r_-} \right) \quad (46.7)$$

It is much less than that of metals. Acid solutions have the highest electrical conductivity because the radius of the solvate formed around an ion of hydrogen (proton) is only from one fifth to one tenth that of other solvates.

The electrical conductivity of electrolytes increases rapidly with the temperature. This is due to the increased degree of dissociation of the solution and the reduction in viscosity of the electrolyte.

46.3. FARADAY'S LAWS

1. An ion is a particle (atom, molecule or group of atoms) that has lost or gained one or several electrons. If the valency of a given element is Z then the charge of the ion is $q = Ze$.

The number of ions deposited on the electrode equals the total charge Q that passed through the electrolyte divided by the charge of the ion. Thus

$$N = \frac{Q}{q} = \frac{Q}{Ze}$$

For a chemical element the mass m_0 of the ion is equal to the atomic mass A of the element divided by Avogadro's number N_A . The mass of the substance deposited on the electrode equals the mass of the ion multiplied by the number of ions. Thus

$$m = m_0 N = \frac{A}{N_A} \frac{Q}{Ze} = \frac{1}{N_A e} \frac{A}{Z} Q \quad (46.8)$$

2. Equation (46.8) contains two laws which were formulated by Faraday in 1832 on the basis of a series of experiments.

Faraday's first law states that the mass of the substance deposited on (or liberated at) an electrode is proportional to the electric charge passing through the electrolyte. Thus

$$m = KQ = Kit \quad (46.9)$$

where m = mass of the substance

Q = charge

i = current

t = time.

The quantity $K = m/Q$ indicates the mass of substance deposited on an electrode when an electric charge of one coulomb passes through the electrolyte. It is called the *electrochemical equivalent* of the substance.

Faraday's second law states that the electrochemical equivalents of the elements are proportional to their chemical equivalents. Thus

$$K = \frac{1}{F} \frac{A}{Z} \quad (46.10)$$

where A = atomic mass of the element

Z = its valency

A/Z = its chemical equivalent, i.e. the ratio of the atomic mass of an element to its valency.

3. Faraday's constant

$$F = N_A e = (9.648670 \pm 0.000054) \times 10^7 \text{ C/kmol}$$

is equal in magnitude to the electric charge that must be passed through an electrolyte to deposit (or liberate) one kilogram equivalent of a substance on (or at) an electrode. Indeed, if we put

$m = A/Z$ in equation (46.8) we obtain $Q = F$. For a monovalent substance this means that the Faraday constant is equal to the electric charge that must be passed through the electrolyte for one kilomole of the substance to be deposited on the electrode.

Determining the Faraday constant experimentally and knowing Avogadro's number we can calculate the charge of a monovalent ion, equal to the charge of the electron, by the equation

$$e = \frac{F}{N_A} \quad (46.11)$$

46.4. THE GALVANIC CELL

1. Experiment shows that if a metal plate is immersed in an electrolyte a potential difference is established between the metal and the electrolyte. This phenomenon is similar to the contact potential difference between two different metals (see Sec. 44.7), except that in an electrolyte it is due to the diffusion of ions and not to electrons as in the contact of metals.

The diffusion of the ions is accompanied by a chemical reaction between the metal and the electrolyte that leads to a change in the internal energy of the reacting substances. This change is proportional to the number of reacting atoms of the metal or, in other words, to the mass of the metal dissolved in the electrolyte. Thus (see Secs. 20.4 and 36.7)

$$\Delta U = \lambda m \quad (46.12)$$

where U = internal energy

λ = energy of the chemical reaction per unit mass

m = mass of the substance.

2. It was found that the chemical reaction proceeds only when the circuit is closed and there is a current in it. Energy evolved by the chemical reaction is converted by this arrangement into the energy of an electric current.

The preceding considerations enable us to calculate the emf of a *galvanic*, or *voltaic*, *cell*. It follows from the law of conservation of energy that the energy of the chemical reactions taking place at the anode and cathode is equal to the work done by extraneous forces in moving a charge along the circuit. Thus

$$\Delta U_{an} + \Delta U_{cath} = W_{extr} = Q\mathcal{E} \quad (46.13)$$

Zinc is commonly used as the cathode of galvanic cells, and energy is liberated when it goes into solution in the electrolyte, i.e. $\Delta U_{cath} > 0$. The reducing reaction at the anode is accompanied by the absorption of energy, $\Delta U_{an} < 0$. Combining equations (46.12)

and (46.13) we have

$$\lambda_{cath}m_{cath} - \lambda_{an}m_{an} = Q\mathcal{E} \quad (46.14)$$

But, according to Faraday's law, $m/Q = K$, the electrochemical equivalent. Thus

$$\mathcal{E} = \lambda_{cath}K_{cath} - \lambda_{an}K_{an} \quad (46.15)$$

3. To evaluate the emf of a cell we shall make use of the following data: the energy per unit mass in the interaction between zinc and sulphuric acid is $\lambda = 6.96 \times 10^6$ J/kg and the electrochemical equivalent of zinc is $K = 3.39 \times 10^{-7}$ kg/C. Neglecting the reaction at the anode, we obtain the approximate value

$$\mathcal{E} \cong \lambda K = 6.96 \times 10^6 \times 3.39 \times 10^{-7} = 2.36 \text{ V}$$

Measurements of emf yield the following data: $\mathcal{E} \cong 2$ V for the Grenet cell and for the lead-acid storage battery, 1.8 V for the alkali storage battery and 1.3 V for the Leclanché dry cell. As we can see, we correctly determined the order of magnitude.

Chapter 47

CURRENT IN A VACUUM

47.1. THERMIONIC EMISSION

1. Vacuum is an excellent insulator because it has no free electric charges. An evacuated volume can be made to conduct electricity artificially if free charges from some source are introduced into the volume. This is accomplished by the phenomenon of *thermionic emission*, the radiation of electrons from the surface of a heated metal.

This phenomenon can be observed by means of the following apparatus. In a glass bulb two electrodes are secured: one is a thin *filament* serving as the cathode and the other is a cylindrical electrode called the *plate*, which serves as the anode (Fig. 47.1a). A high vacuum is maintained in the bulb. The symbol for this tube shown in Fig. 47.1b is used in diagrams.

2. Let us connect this tube into the circuit shown in Fig. 47.2. Here *Sw* is the switch, *FB* is the filament battery, *PB* is the plate battery and *mA* is a milliammeter. If we open the switch there will be no current in the circuit even if the potential of the plate is raised to 300 or 400 V. But when we close the switch the filament serving as the cathode is heated to incandescence. Then a current is established in the circuit. This leads us to the conclusion that negatively charged particles are given off by the incandescent filament.

A measurement of their charge-to-mass ratio indicated that they are electrons. They are called *thermoelectrons*.

3. The phenomenon of thermionic emission is similar to the evaporation of a liquid (see Secs. 35.1 and 35.2). Just as molecules with energy greater than the energy of evaporation fly away from a liquid, so are electrons with energy greater than the work function emitted

from a metal. The number of such particles increases rapidly with the temperature.

Finally, saturated vapour forms above the surface of the liquid

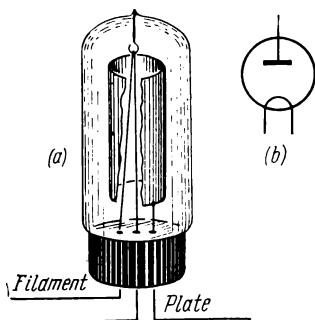


Fig. 47.1

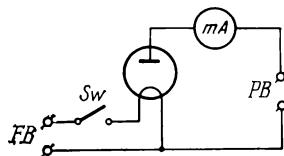


Fig. 47.2

in the same way that an *electron cloud* is formed around the filament.

4. The problem of direct conversion of the internal energy of heated bodies into the energy of an electric current can be solved by thermionic emission. This is the principle of operation of the Topaz nuclear power installation, developed in 1970-71 in the USSR.

The Topaz is a small-size uranium nuclear reactor (see Secs. 82.9 and 82.10) with a heating capacity of about 150 kW. It consists of a massive moderator of zirconium hydride in which through electricity-generating channels have been cut. These channels are in the form of five electricity-generating elements connected in parallel. Located at the centre of each channel is a small cylinder of uranium coated with an air-tight shell of a molybdenum alloy. This rod serves as the cathode. The anode is a cylinder separated from the cathode by a clearance of fractions of a millimetre. The heat evolved in the nuclear reaction in the uranium heats the surface of the cathode to a high temperature. The stream of electrons emitted from the surface of the cathode is captured by the anode. A current is established in the external load due to the potential difference set up between the cathode and anode.

The Topaz reactor is much more powerful than the Romashka installation (see Sec. 44.8). Experimental models have operated with a power output of 5 to 10 kW for 1000 hours. Evidently, installations of this type will be applied in the near future in everyday practice along with other types of low-power nuclear electric power plants.

47.2. THE DIODE AND ITS CHARACTERISTICS

1. The vacuum tube discussed in the preceding section is called a *two-electrode electronic tube*, or *valve*, or simply a *diode*.

Tubes with directly and indirectly heated cathodes are used. In tubes with a directly heated cathode, the tungsten filament serves as the cathode. To obtain considerable emission the filament is heated to a temperature from 2000 to 2500 K. In the indirectly heated tubes the cathode is a small nickel tube coated with a layer of barium, strontium or calcium oxide. The work function from the surface of such a cathode is much less than when tungsten is used. This enables

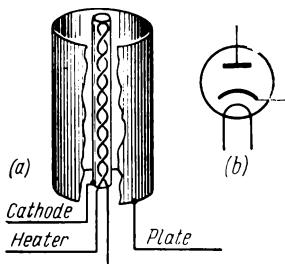


Fig. 47.3

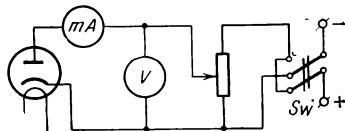


Fig. 47.4

the temperature of incandescence to be reduced to 1000 K. The cathode is heated by a small spiral filament inside the nickel tube (Fig. 47.3a). The symbol used in circuit diagrams for a tube with indirect heating is shown in Fig. 47.3b.

2. Let us wire the circuit illustrated in Figure 47.4. The resistor is connected into the circuit so that the potential over the anode can be varied, by moving the sliding contact of the resistor, from zero (at the lower position of the contact) to the voltage over the poles of the plate battery. When thus connected the resistor is called a potentiometer. Change-over switch *Sw* enables us to obtain either a positive potential over the plate with respect to the cathode (in the position shown in the diagram) or a negative potential (if the contacts are switched over).

If we smoothly change the plate potential and register its value each time together with the corresponding current, we shall obtain the function expressing the dependence of the current on the potential. A graph of this function is called the *current-voltage characteristic* of the diode. Three such characteristics, plotted for different temperatures of the cathode, are given in Fig. 47.5.

3. The first thing we notice is the *nonlinearity* of the characteristic. This means that the current is not proportional to the plate potential or, in other words, Ohm's law does not hold here. The reason for this is the following.

As we know from equation (39.18) the current density is $j = env.$ In metals and electrolytes the concentration of free charges (electrons or ions) is independent of the current and the average velocity of the charges is proportional to the field strength. This is the basis for Ohm's law (see Sec. 44.4). Matters are entirely different in an electronic tube. Owing to the electron cloud around the cathode, the dependence of the velocity of the charges on the field strength is much more complex, and the concentration of charges in the electron cloud decreases with an increase in the current. As a result, the dependence of the current on the potential at the rising part of the characteristic is expressed by the three-halves power law:

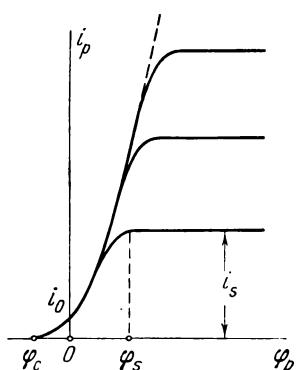


Fig. 47.5

called the *saturation current*, can be calculated on the basis of the following. The maximum current density in the tube is equal to the product of the charge of the electron by the number of electrons emitted per second from the surface of the cathode. Thus

$$j_s = eG \quad (47.2)$$

where G is the rate of vaporization. We can calculate it by formula (35.4), making use of the fact that according to quantum mechanical conceptions $\alpha = 2$ for an electron gas. Substituting into equation (47.2) we obtain the *Richardson-Dushman equation*:

$$j_s = BT^2 e^{-\frac{W_0}{kT}} \quad (47.3)$$

where W_0 = work function of the electron (see Sec. 44.9)

T = temperature of the metal

B = constant depending upon the metal of which the cathode is made.

5. If a potential φ_c , negative with respect to the cathode, is applied to the plate, the tube will be cut off, i.e. the plate current will equal zero. This means that the kinetic energy of the electrons is insuffi-

cient to overcome the field that repels them from the plate. Thus

$$\frac{mv^2}{2} \ll e\Phi_c \quad (47.4)$$

By measuring the cutoff potential φ_c we can determine the maximum kinetic energy (and velocity) of the electrons emitted from the cathode.

47.3. THE TRIODE AND ITS CHARACTERISTICS

1. A third electrode, called the *grid*, is inserted into a tube to control the electron current. Such a tube is called a *three-electrode electronic tube*, or *valve*, or simply a *triode*. The grid may have the form of a spiral wire located between the cathode and plate (Fig. 47.6a).

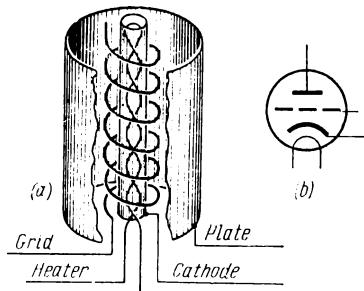


Fig. 47.6

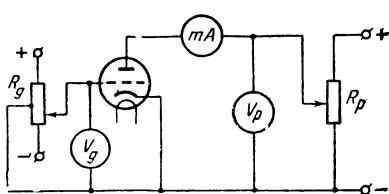


Fig. 47.7

The symbol used in diagrams for a triode with an indirectly heated cathode is shown in Fig. 47.6b. All quantities referring to the grid have the subindex g .

Let us wire the circuit illustrated in Fig. 47.7. Here resistor R_p is a potentiometer by means of which the plate potential is regulated and R_g is a potentiometer for varying the grid potential. If the sliding contact is set at the middle point of potentiometer R_g , the grid potential will equal the cathode potential, i.e. zero. When we move the sliding contact below the middle point we apply a negative potential over the grid; above this point the potential applied is positive.

2. Maintaining a certain plate potential we shall vary the grid potential from negative to positive values, recording the plate current each time. Using these measurements we shall plot the *grid characteristic* of the triode. It shows the dependence of the plate current on the grid potential at a given plate potential. Such a curve is shown in Fig. 47.8. An analysis of the characteristic leads to the following conclusions.

3. The characteristic is *nonlinear* and therefore Ohm's law does not hold here as well. However, we can neglect the nonlinearity as a first approximation over a short portion in the middle of the characteristic curve.

We see, further, that the current increases for positive grid potentials and decreases for negative potentials. This has the following explanation.

If the grid potential is zero, the current i_0 in the tube is determined only by the plate potential. Suppose a positive potential is applied on the grid. Then the field strengths of the plate and grid are added together, leading to an increase in the velocity of the electrons and to the dissipation of the electron cloud. As a result the current increases to the saturation current i_s . For a negative grid potential the reverse process is observed: the field is weakened, the velocity of the electrons decreases and the concentration of the electron cloud increases. As a result, the current decreases.

Finally, at a certain negative grid potential φ_c , depending on the plate voltage, the kinetic energy of the electrons is insufficient to overcome the repelling action of the grid and electrons no longer enter the region between the grid and plate. At this the tube is *cut off*, i.e. its current equals zero.

Thus, by varying the grid potential we can control the current in the tube. This is why we call it a *control grid*.

47.4. THE CATHODE-RAY TUBE

1. One extensively used electronic device is the *cathode-ray tube*. It is illustrated schematically in Fig. 47.9. Electrons emitted by the

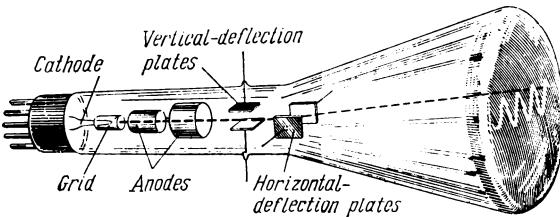


Fig. 47.9

indirectly heated cathode pass through the grid (control electrode) and are accelerated by two anodes. This system, called an *electron*

gun, serves to produce an electron beam focussed on a screen. This screen is coated with a *fluorescent substance* (see Sec. 79.2) which glows brightly when struck by the bombarding electrons.

The electron beam is focussed in the following manner. A negative potential (from -20 to -70 V) is applied to the control electrode (grid). The field of this electrode narrows the electron beam coming from the cathode. A positive potential of $+250$ to $+500$ V is applied over the first anode and one from $+1000$ to $+2000$ V over the second anode. Even higher potentials are employed in a television-picture tube, or kinescope. By varying the potentials of the control electrode and anodes we can change the focus and brightness of the spot on the screen of the tube.

2. Next the beam passes through a system of control electrodes. If we apply, for example, a positive potential over the upper vertical-deflection plate and a negative potential over the lower plate, the electron beam will be deflected upward. If the polarity is reversed the beam is deflected downward. Thus an oscillating potential applied to these plates causes vertical oscillation of the electron beam. In exactly the same way, oscillations of the potential applied to the horizontal-deflection plates leads to horizontal oscillations of the electron beam.

A tube in which the electron beam is deflected by changing the electric field between the deflecting plates is said to have *electrostatic control*. There are tubes with *electromagnetic control* which are used, for instance, in television sets. These tubes have no control electrodes but, instead, special coils are mounted on the neck of the tube. The current in these coils is varied as required. Changes in the magnitude and direction of the induction vector of the magnetic field set up by the coils correspondingly change the magnitude and direction of the Lorentz force that deflects the electron beam.

Chapter 48

CURRENT IN GASES

48.1. IONIZATION AND RECOMBINATION

1. For a gas to become a conductor it must be ionized, i.e. its neutral molecules (or atoms) must be converted into ions. In ionization certain kinds of rays detach electrons from the molecules which are thus converted into positive ions. In this way an electric current in gases is a counterflow of positive ions and free electrons.

Ionization may be effected by short-wave radiation, i.e. by ultraviolet, X- and gamma rays, and also by alpha, beta and cosmic

rays. Electrical conduction due to external radiation is sometimes called *nonsself-maintained conduction*.

2. Let us denote by N_0 the number of molecules of gas in volume V being investigated. The molecule concentration is $n_0 = N_0/V$. A part of the molecules are ionized. We denote the ions of like sign by N , their concentration is $n = N/V$.

The ratio of the number of ions to the total number of molecules is called the *ionization coefficient*, $\alpha = N/N_0 = n/n_0$. In nonsself-maintained conduction the ionization coefficient of gases is usually very low: $\alpha \cong 10^{-12}$ to 10^{-10} .

The number of neutral (nonionized) molecules is

$$N' = N_0 - N = N_0(1 - \alpha) \cong N_0$$

The same is true of the concentration: $n' \cong n_0$.

3. Suppose that the action of some kind of rays during the time Δt produces in a volume V a number of ions

$$\Delta N = \beta N' \Delta t \cong \beta N_0 \Delta t = \beta n_0 V \Delta t \quad (48.1)$$

where β is the ionization coefficient, which depends on the energy of the ionizer. The ionization process is accompanied by the reverse process, the *recombination of ions*. In the latter, a neutral atom or molecule is formed when a positive ion collides with an electron. The number of particles recombining in the volume V during the time Δt is proportional to both the number of positive ions and the number of electrons. Since each of these numbers is equal to N , the probability of recombination is proportional to N^2 . Besides, it is proportional to the time Δt and inversely proportional to the volume V . Indeed, the closer the particles of different signs are to one another, the greater the probability of their recombination. Hence

$$\Delta N_{rec} = \gamma \frac{N^2 \Delta t}{V} = \gamma n^2 V \Delta t \quad (48.2)$$

where γ is the *recombination coefficient*, which depends upon the kind of gas being ionized.

When a certain time has elapsed after the beginning of ionization, the rate of recombination $\Delta N_{rec}/\Delta t$ becomes equal to the rate of ionization $\Delta N/\Delta t$, and a dynamic equilibrium is established in the gas. This state corresponds to a definite ion concentration which can be found by equating expressions (48.1) and (48.2). Thus

$$n = \sqrt{\frac{\beta n_0}{\gamma}} \quad (48.3)$$

48.2. NONSELF-MAINTAINING DISCHARGES

1. To investigate this process let us consider the phenomena occurring in an *ionization chamber* of the type shown schematically in Fig. 48.1. The body of the vessel containing the gas usually serves

as the cathode. The anode is a rod which is insulated from the body by some good insulator made of amber, porcelain or a special plastic. Hard radiation, such as X-, gamma or cosmic rays, penetrates the walls of the chamber. Soft radiation, such as beta or alpha particles, is admitted through a special window.

We apply a substantial potential difference over the electrodes of the chamber. It ranges from several hundred to several thousand

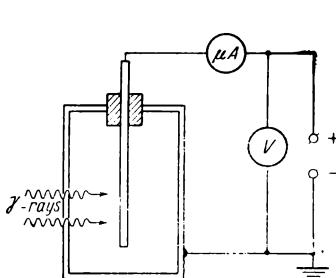


Fig. 48.1

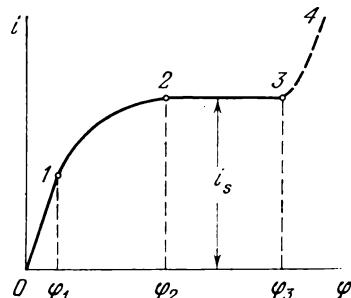


Fig. 48.2

volts depending upon the design of the chamber. Varying the potential difference and measuring the corresponding current we plot a *current-voltage characteristic* of the gas discharge (Fig. 48.2).

2. If we pay no attention to the current jump in portion 3-4, the rest of the characteristic curve, portion 0-1-2-3, closely resembles the current-voltage characteristic of a diode. Portion 0-1-2-3 is called the region of nonself-maintaining gas discharge; region 3-4 represents the transition to a self-maintaining discharge.

In the region of weak currents (portion 0-1) the characteristic curve is approximately linear. Therefore, Ohm's law holds here. In portion 1-2, Ohm's law no longer holds and the current increases at a lower rate than the potential. Finally, in portion 2-3, the current is independent of the potential. This is the so-called *saturation current* i_s .

3. To find why the discharge changes for different values of the potential, let us determine the current in the chamber. For the sake of simplicity we shall assume that the electrodes of the chamber are parallel flat plates arranged at a short distance d from each other. Here the field is uniform between the plates and the current density is the same at all points. The current density is expressed by equations (46.3) and (46.6) as for electrolytes. Thus

$$j = en(v_+ + v_-) = en(b_+ + b_-)E \quad (48.4)$$

where b_+ = mobility of the positive ions

b_- = mobility of the electrons

$n = \alpha n_0$ = ion concentration.

4. We obtained Ohm's law from equation (46.3) by analyzing the mechanism of the current in electrolytes because the concentration of ions in them is independent of the current. It is quite a different matter in gases. Here the ion concentration decreases with an increase in current, and the dependence of the current on the potential difference is a complex function.

Let us consider the changes in the condition of dynamic equilibrium between the ionization and recombination processes when we have a current. Owing to the current a definite number of ions is removed from the volume between the electrodes during the time Δt . Thus

$$\Delta N_{cur} = \frac{Q}{e} = \frac{i\Delta t}{e} = \frac{jA\Delta t}{e} \quad (48.5)$$

When there is a current the condition of dynamic equilibrium becomes

$$\Delta N = \Delta N_{rec} + \Delta N_{cur} \quad (48.6)$$

Substituting the values of the terms we have

$$\beta n_0 V = \gamma n^2 V + \frac{jA}{e}$$

Since $d = V/S$ and is the distance between the electrodes, we obtain, finally,

$$\beta n_0 = \gamma n^2 + \frac{j}{ed} \quad (48.7)$$

Hence, we find that the ion concentration is a function of the current density. Cancelling the ion concentration from the system of equations (48.4) and (48.7) we have a function expressing the dependence of the current density on the field strength. The graph of this dependence is the current-voltage characteristic (see Fig. 48.2).

5. Let us consider two limiting cases: when the number of ions carried away by the current is much less or is much more than the number of recombining ions.

First we shall suppose that $\Delta N_{cur} \ll \Delta N_{rec}$. Then the second term in the right-hand side of equation (48.6) can be neglected. We have returned to equation (48.3). Therefore, for weak currents the ion concentration is practically independent of the current, and the current density is thereby approximately proportional to the field strength. This corresponds to the almost linear portion 0-1 of the characteristic curve.

Now we shall suppose that the opposite is true: the current is so great that practically all the ions are carried away by it before any recombination can take place. Then $\Delta N_{cur} \gg \Delta N_{rec}$ and the first term can be neglected in the right-hand side of equations (48.6) and

(48.7). We have

$$\beta n_0 \cong \frac{j_s}{ed} \quad (48.8)$$

As we see, the current density is independent of the field strength. This is the *saturation current*.

6. The saturation current is

$$i_s = j_s A = \beta n_0 e V \quad (48.9)$$

It is proportional, as we see, to the volume of the ionization chamber, concentration of gas molecules (in other words, to the gas pressure) and to the ionization coefficient. In this way, by operating an ionization chamber at the saturation current, we can compare the ionizing capacity of various types of radiation. Ionization chambers are extensively employed for this purpose in nuclear physics (see Sec. 81.8).

48.3. COLLISION IONIZATION

1. It is evident from the curve in Fig. 48.2 that for a potential difference exceeding φ_3 , the current in the gas increases with a jump though it was constant in the large range of potential differences from φ_2 to φ_3 . An analysis of expression (48.4) for the current density leads us to the conclusion that the sudden large jump in current can be due only to the fact that the *ion concentration increases with a jump* at potential differences above φ_3 . The cause of the abrupt jump is *collision ionization* of the gas.

A certain energy E_{ion} of an amount depending on the chemical nature of the gas is required to ionize an atom or a molecule. This energy is $E_{ion} = 13.6$ eV for atoms of hydrogen and oxygen, 14.5 eV for nitrogen, 24.5 eV for helium and of the order of 4 or 5 eV for the alkali metals. If the kinetic energy of the particle colliding with an atom exceeds the ionization energy, the molecule may be ionized in an *inelastic collision*.

2. The kinetic energy K of an ionizing particle is due to the work W done by the electric forces acting on the ion in the electric field. Thus

$$K = W = F\lambda = eE\lambda \quad (48.10)$$

where E = field strength

λ = mean free path of the ion.

Both the positive ions and the electrons move in a field of the same strength, but the mean free path of the electron is much longer than that of the ion. Consequently, free electrons play the main role in collision ionization.

3. A second reason why electrons are more effective ionizers is that, in accordance with the law of conservation of momentum (see Secs. 21.7 and 23.2), only a part of the kinetic energy in a collision can be converted into internal energy. It was found that the smaller the mass of the ionizing particle, the greater the share of its kinetic energy that can be converted into internal energy and expended on ionization. Let the mass and momentum of the ionizing particle be m and p , and let the molecule of mass M be at rest before the inelastic collision. The change in internal energy is equal to the loss of kinetic energy. Thus

$$\Delta \mathcal{E}_0 = K - K_1 = \frac{p^2}{2m} - \frac{p^2}{2(M+m)} = \frac{p^2 M}{2m(M+m)} = K \frac{M}{M+m} \quad (48.11)$$

where $K = p^2/2m$ is the energy of the ionizing particle. The mass of the ion is equal to the mass of the molecule and therefore $\Delta \mathcal{E}_0 = K/2$. The mass of the molecule is several thousand times more than the mass of the electron. Consequently, in equation (48.11) the fraction $\frac{M}{M+m} \cong 1$ and $\Delta \mathcal{E}_0 \cong K$. Thus, even with the same kinetic energy, the electron is twice as effective as the ion for ionizing molecules.

4. With collision ionization and an external field of sufficient strength, an *ion avalanche* is produced in the gas. Secondary electrons, due to collision ionization, are also accelerated by the field and, in their turn, ionize the molecules they encounter. As a result of such a *chain reaction*, even the relatively small number of electrons produced in the process of external ionization is capable of causing a substantial discharge current.

However, this process cannot be called a self-maintaining discharge since the discharge ceases, if the external ionizer stops acting, as soon as all the ions are neutralized at the electrodes. To produce a well-developed self-maintaining discharge it is necessary for the positive ions as well to be capable of knocking free electrons either out of the gas molecules or out of the cathode. It follows from equation (48.10) that this can be accomplished either by increasing the field strength or by increasing the mean free path of the ion.

48.4. THE GEIGER-MÜLLER COUNTER

1. In collision ionization the number of ion pairs is sharply multiplied (from 10^4 to 10^6 times). Therefore if even a single ion pair (one positive ion and one electron) is produced in a volume of gas, their number will be increased millions of times by collision ionization. In this case, the ionizing chamber is said to be operating with *gas amplification*.

Gas amplification is applied in the Geiger counter, which is used to detect and count the separate particles of radioactive radiation. Under definite operating conditions such a device will count each separate particle.

2. Usually such a counter consists of a sealed airtight glass tube with cathode K , a thin-walled metal cylinder, fitted inside. Anode A is a fine wire stretched along the middle of the tube (Fig. 48.3). Counters used for soft beta and alpha particles differ in that the radiation is admitted into the working space in the tube through a thin film that closes one end.

The counter is connected to a registering circuit. A negative potential is applied to the body of the tube (cathode) and a positive

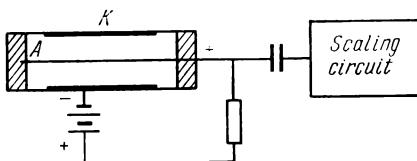


Fig. 48.3

potential to the wire. Connected in series with the counter is a resistor with a resistance of several megohms. The signals are transmitted from the resistor through a separating capacitor with a capacitance of several thousand picofarads to the input of the scaling circuit.

3. Suppose that a particle enters the tube and, along its path, produces at least one ion pair. Electrons moving toward the anode (fine wire) pass through a field with increasing strength. Their velocity quickly increases and they produce an ion avalanche along their path.

When they reach the wire the electrons reduce its potential and, as a result, a current is established in the resistor. This produces a voltage pulse at the ends of the resistor. This pulse passes through the capacitor to the input of the scaling circuit.

The high potential, which was formerly wholly over the anode, is switched over to the resistor, and the field strength in the counter is reduced, thereby reducing the kinetic energy of the electrons [see equation (48.10)]. This cuts off gas amplification. Thus the resistor operates as a *damping resistance*.

4. Owing to the fact that the mobility of the positive ions is less than that of electrons, and because these ions move toward the cathode where the field strength is much less than at the anode, the velocity of the positive ions is much less than that of electrons. As a result, the ions reach the cathode in a time not less than $\tau \cong 10^{-4}$ seconds after

the beginning of the discharge. During this time the counter is unable to register new particles entering it. Indeed, while the positively charged cloud of ions fills the working volume of the counter, the field strength is low and no collision ionization can take place. In this connection, the time τ is often called the *dead time* of the counter.

The *resolving power* of a counter indicates the number of particles the counter is capable of registering in unit time as separate pulses. Since the pulses from two successive particles are registered separately if they follow each other after a time interval not less than the dead time, the resolving power is about 10^4 pulses per second.

48.5. SELF-MAINTAINING DISCHARGES. PLASMA

1. As previously mentioned, the process of collision ionization is insufficient by itself to initiate a self-maintaining discharge. Such a discharge can be achieved only if processes occurring in the gas during the discharge continuously produce more electrons that participate in collision ionization after being accelerated. There can be several such processes. Sometimes they act simultaneously, sometimes one of them begins to play a predominant role. This depends upon the pressure of the gas, its temperature and the field strength. The most important of these processes are the following.

Thermionic emission from the cathode occurs when the cathode is at a high temperature. This is the process that initiates an *arc discharge*.

Secondary electron emission from the cathode is observed if the kinetic energy of the positive ions is sufficient to free electrons from the cathode by collision. This process provides for a *glow discharge* (see Sec. 48.6).

Thermal ionization of the gas takes place when the kinetic energy of the gas molecules exceeds the ionization energy. Thus

$$kT > \mathcal{E}_{\text{ion}} \quad (48.12)$$

It can readily be seen that the ionization of the gas upon thermal collisions of its molecules is possible only at very high temperatures ($\mathcal{E}_{\text{ion}} \cong 10$ eV). Thus

$$T > \frac{\mathcal{E}_{\text{ion}}}{k} \cong \frac{10 \times 1.6 \times 10^{-19}}{1.38 \times 10^{-23}} \cong 10^5 \text{ K}$$

Photoionization of the gas occurs as the result of short-wave ultraviolet or X-radiation. A molecule of the gas absorbs a portion of energy (a quantum of radiation, see Sec. 68.3) which is greater than the ionization energy. Then it ejects an electron and is converted into a positive ion.

2. At a high degree of ionization an ionized gas is actually a special state of matter, distinguished from the gaseous, liquid and crystalline states. This fourth state of matter is called a *plasma*.

Plasma produced by collision ionization of a gas is said to be *gas-discharge plasma*. Its special feature is that there is practically no energy exchange between the free electrons and the ions, and their average kinetic energies are considerably different. This last can be explained qualitatively by the fact that the mean free path of the electron is about 5 to 10 times longer than that of the ion. As a result, the electrons accumulate about the same number of times more energy in the electric field.

By analogy with the gas temperature which is found in formula $\bar{\epsilon} = \frac{3}{2} kT$ (see Sec. 26.5), we can introduce the *ion temperature* and the *electron temperature* from the equations

$$\bar{\epsilon}_{\text{ion}} = \frac{3}{2} kT_{\text{ion}} \quad \text{and} \quad \bar{\epsilon}_{\text{el}} = \frac{3}{2} kT_{\text{el}}$$

It was found that $T_{\text{el}} \cong 5000$ K at $T_{\text{ion}} \cong 300$ K.

Thus, a gas-discharge plasma is a nonequilibrium or, as they say, nonisothermal, state of matter.

3. The high-temperature plasma produced by thermal ionization is an equilibrium, or, in other words, *isothermal*, plasma. Its degree of ionization is very high and it is therefore a very good conductor. The conductivity of high-temperature plasma is comparable to that of metals.

The temperature of the surface of the sun and stars is several thousand degrees, and their interior is heated to millions of degrees. Hence, a considerable part of the matter in the universe is in the state of a high-temperature plasma.

48.6. GLOW DISCHARGES

1. This kind of discharge can be conveniently observed in a gas-discharge tube having two electrodes with a distance of about 0.5 m between them, and having a potential difference of about a thousand volts. We will find that at standard atmospheric pressure in the tube there is no discharge (we shall neglect very weak nonself-maintaining discharges). When we reduce the pressure in the tube to about 40 or 50 mm-Hg, a narrow glowing discharge is observed. At a pressure of about 0.5 mm-Hg, the discharge widens until it fills the whole tube, the positive column at the anode being divided into a number of alternating layers called striations. Finally, at a pressure near 0.02 mm-Hg, the glow in the tube becomes fainter and disappears, but the glass opposite the cathode begins to glow brightly.

The self-maintaining discharge initiated in the gas at low pressure is called a *glow discharge*.

2. Investigations showed that the potential of a discharge tube is very unevenly distributed, as is evident from Fig. 48.4. Near the cathode the field strength is considerably higher than in the glowing column at the anode. As a result the charges are subject to the action of the accelerating force practically only in the cathode dark space. They move at almost uniform velocity in the region of anode glow.

Electrons accelerated in the cathode space acquire a kinetic energy sufficient for collision ionization of the gas. As a result, a gas-discharge plasma is formed in the region of anode glow, the conductivity of this region is sharply increased and the potential difference

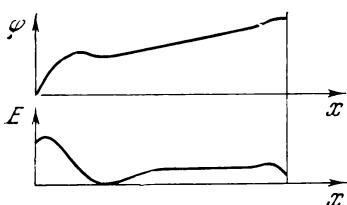


Fig. 48.4

is reduced. The glow of the gas is due to the energy evolved by its molecules in the recombination of the ions. The kind of glow (its colour) depends on the kind of gas filling the tube. In their turn, the positive ions, accelerated in the cathode dark space, bombard the cathode, tearing out electrons by their impact. Naturally, the kinetic energy of an ion must exceed the work function required to

liberate an electron from the cathode. According to equation (48.10) this can be accomplished either by increasing the strength of the electric field or by increasing the mean free path of the ion. This latter is achieved when the gas is rarefied.

A glow discharge is observed in fluorescent lamps. Neon, argon, krypton, mercury and sodium-vapour glow lamps are widely employed as special light sources in laboratories, and in advertising. Lamps of quartz glass, filled with mercury vapour, are used as sources of ultraviolet light.

When ions bombard the cathode they cause localized heating of the metal at projecting parts of the surface. The projections are melted and vaporized. The metal vapours thus produced are deposited on cold surfaces. This principle is used for the *cathode atomization of metals*, which enables thin, strong metallic coatings to be applied to various articles. High-quality mirrors, half-silvered glass plates and other items are manufactured by this method.

3. At a gas pressure below 0.01 or 0.02 mm-Hg the glow of the gas ceases in the tube. This means that collisions of electrons with gas molecules, ionization of the molecules and the reverse process of recombination occur relatively seldom. Hence the main mass of electrons emitted by the cathode and accelerated in the cathode space fly further freely by inertia. This explains why an electron beam in a cathode-ray tube moves in a straight line and perpendicular to the cathode regardless of the position of the anode (Fig. 48.5). Historically, this was the first way in which a beam of free electrons

was obtained. Sir William Crookes (1832-1919) obtained such beams in the last quarter of the 19th century. They were named *cathode rays*. Today this name is preserved only by tradition and they are called electron rays.

If the cathode is placed inside the discharge tube and holes (canals) are drilled in the cathode, weakly glowing beams are observed behind the cathode (Fig. 48.6). Investigations show that these particles

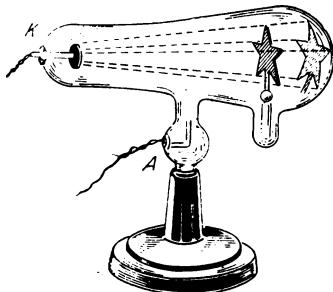


Fig. 48.5

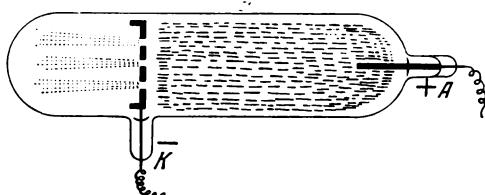


Fig. 48.6

are positive ions. They are accelerated by the field in the cathode space and slip through the canals by inertia into the space behind the cathode. They are called *anode, positive or canal rays*.

Canal rays are a source of positive ions of the gas that fills the tube. Such sources of positively charged ions are applied in mass spectrometers (see Sec. 41.8) and in accelerators (see Secs. 41.4, 41.5 and 41.6).

48.7. PLASMA IN A MAGNETIC FIELD

1. When not in a magnetic field plasma behaves like an ordinary gas. The reason is that plasma is *quasi-neutral*: even in quite small (but not microscopic) volumes the total charge of the electrons and positive ions equals zero. Therefore, in the absence of an external magnetic field, phenomena in plasma are described by the ordinary equations of fluid dynamics (see Ch. 30).

If the plasma is in a magnetic field it displays various features due to the action of the magnetic field on the moving charges. The branch of science dealing with the behaviour of plasma and other conducting liquids (for instance, liquid metals) in a magnetic field is called *magnetic hydrodynamics*, and for large Mach numbers (see Sec. 30.6), *magnetogas dynamics*. In this book we cannot devote space to a detailed discussion of this new and rapidly developing field. We shall consider only certain of its major concepts.

2. Suppose that a certain volume of plasma is moving at a velocity v across the lines of induction of a field of induction B . Then an induced emf will be set up in this volume just as in any conductor (see Sec. 43.2). If the characteristic dimension of the portion of plasma is l then, according to equation (43.5), $\mathcal{E} = vBl$. The resistance of this portion of plasma is $R = \frac{\rho l}{A} \cong \frac{l}{\gamma l^2} = \frac{1}{\gamma l}$. According to Ohm's law the current induced in the plasma is

$$i_{ind} = \frac{\mathcal{E}}{R} \cong \gamma v Bl^2 \quad (48.13)$$

In accordance with Lenz's law (see Sec. 43.8), the induced current interacts with the field so that the force of interaction opposes the movement of the plasma. Thus, in addition to ordinary hydrodynamic forces, the plasma is subject to electromagnetic forces as well.

3. Precise calculation of this interaction is associated with extremely complex mathematics. But the role of the various forces can be assessed by making use of certain dimensionless criteria similar to the Reynolds number (see Sec. 11.8). *The magnetic Reynolds number Re_m is the ratio of the magnetic induction of the field set up by the induced currents to the induction of the external magnetic field.* Thus

$$Re_m = \frac{B_{ind}}{B} \quad (48.14)$$

To evaluate this quantity we use the fact that $B_{ind} = \mu_0 H_{ind} \cong \frac{\mu_0 i_{ind}}{l}$, where l is a characteristic dimension of the portion of plasma, and the induced current is expressed by equation (48.13). Substituting, we have

$$B_{ind} \cong \mu_0 \gamma v Bl \quad (48.15)$$

$$Re_m = \mu_0 \gamma v l \quad (48.16)$$

Large magnetic Reynolds numbers are obtained either for a high electrical conductivity of the plasma, or for considerable characteristic dimensions and velocities. The last two are frequently observed on the astronomic scale and are of significance in astrophysics.

4. At large magnetic Reynolds numbers ($Re_m \gg 1$) the motion of the plasma in the magnetic field should set up a very strong induced magnetic field, many times stronger than the external magnetic field. This requires energy which is made available only at the expense of the kinetic energy of the plasma. Consequently, the induced currents, interacting with the external magnetic field, oppose the motion of the plasma across the field.

At $Re_m \gg 1$ it may turn out that the plasma practically cannot move with respect to the field. In this case the *magnetic field is said*

to be frozen into the plasma. Here any motion of the plasma is accompanied by corresponding changes in the magnetic field so that the plasma does not cross its lines of induction. Inversely, if the external magnetic field is changed then, at $\text{Re}_m \gg 1$, the plasma moves accordingly so as to retain the freezing-in condition. This feature is resorted to for "pinching" and heating the plasma by a rapidly increasing magnetic field. This will be discussed in the next section.

5. The second characteristic criterion in magnetic hydrodynamics is the *Alfvén number* Al , equal to the ratio of the energy density of the magnetic field $w_m = \frac{B^2}{2\mu_0}$ to the kinetic energy of unit volume of the plasma, i.e. to the density of the kinetic energy $w_k = \rho v^2/2$. Thus

$$\text{Al} = \frac{w_m}{w_k} = \frac{B^2}{\mu_0 \rho v^2} \quad (48.17)$$

The Alfvén number can also be interpreted as the ratio of the pressure exerted by the magnetic field $p_m = \frac{B^2}{2\mu_0}$ to the dynamic pressure (or head) $p_{dyn} = \rho v^2/2$ (see Sec. 30.16).

6. At low magnetic Reynolds numbers the plasma can move with respect to the field. This gives rise to magnetic forces which can be evaluated by Ampère's law [equation (41.15)]. Substituting into this formula the value of the induced current from equation (48.13), we have

$$F_m = iBl \cong \gamma v B^2 l^3 \quad (48.18)$$

To assess this force it is compared either to the friction drag $f \cong \eta lv$ or to the pressure drag $R \cong \rho v^2 l^2$. We then obtain two new criteria:

the Stuart number

$$N = \frac{F_m}{R} = \frac{\gamma B^2 l}{\rho v} = \text{Al} \times \text{Re}_m \quad (48.19)$$

and the Hartmann number

$$\text{Ha} = \sqrt{\frac{F_m}{f}} = Bl \sqrt{\frac{\gamma}{\eta}} = \sqrt{N \times \text{Re}} \quad (48.20)$$

The significance of these criteria can be understood from the following example. If a liquid flows along a pipe across a magnetic field the type of flow will be only weakly influenced by the field at low Hartmann or Stuart numbers, and the resistance to motion is due chiefly to the viscosity of the liquid. At high Hartmann or Stuart numbers, the viscosity of the liquid becomes a secondary factor and the resistance to motion is primarily due to the interaction between the liquid and the magnetic field.

48.8. PINCHING AND CONFINING A PLASMA

1. One of the most vital problems that can be solved by using a plasma is the realization of controlled thermonuclear fusion (see Sec. 82.12). Thermonuclear reactions can begin at temperatures not less than a hundred million degrees. It is obvious that the high-temperature plasma obtained at such temperatures cannot be held in any vessel. The plasma will either burn through the walls or transmit its energy to them and be cooled. This plasma can be confined only by means of a magnetic field. Indeed, if the pressure exerted by the magnetic field $p_m = \frac{B^2}{2\mu_0}$ is greater than the pressure of the gas $p = nkT$, the magnetic field will confine the plasma. The condition of confinement is

$$\frac{B^2}{2\mu_0} \geq nkT \quad (48.21)$$

The concentration of particles in a thermonuclear plasma is $n \cong 10^{22} \text{ m}^{-3}$. Hence the required induction of the confining magnetic field is

$$B \geq \sqrt{2\mu_0 nkT} \cong \sqrt{2 \times 4\pi \times 10^{-7} \times 10^{22} \times 1.38 \times 10^{-23} \times 10^8} \cong 6 \text{ T}$$

A steady field with such an induction has not yet been obtained, but even stronger pulsed fields are set up.

A confining magnetic field can be set up either by means of internal currents in the plasma itself, or by an external magnetic field produced by current in conductors. As an example, let us consider the idea for one experiment of this kind.

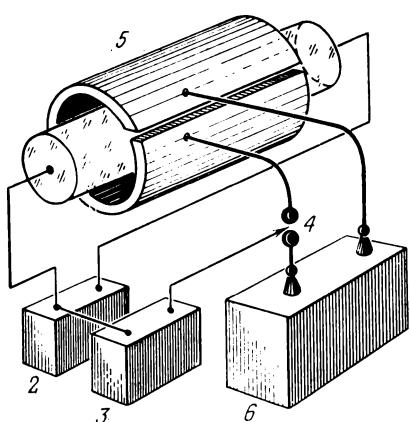


Fig. 48.7

tary 6 of capacitors connected in parallel. When the capacitors discharge, enormous currents, of the order of a million amperes, are established in the turn. As a result, a practically uniform

2. Illustrated schematically in Fig. 48.7 is a device for confining a plasma filament by a rapidly increasing magnetic field. This phenomenon is called *theta-pinch*. Gas in quartz tube 1 is ionized with the aid of auxiliary high-frequency generator 2 by establishing a glow discharge in the gas. When plasma has been produced in the tube, discharge device 3 ignites a spark in discharge gap 4 as a result of which turn 5 is connected to a bat-

magnetic field is set up inside the turn. Its lines of induction are along the tube (longitudinal field). The induction of this field can be evaluated by means of expression $B \cong \frac{\mu_0 i}{2a}$ which is obtained from equation (40.21) under the condition that $r_0 = a$ is the radius of the turn. Putting $a \cong 0.1$ m and $i \cong 10^6$ A we obtain

$$B \cong \frac{4\pi \times 10^8}{2 \times 10^7 \times 0.1} \cong 6 \text{ T}$$

The rapidly growing magnetic field pinches the plasma into a very narrow filament. The process of pinching the plasma takes place so rapidly that the resulting shock wave heats the plasma to a temperature over 10 million degrees, i.e. to the highest temperatures ever obtained in the laboratory.

3. One of the drawbacks of theta-pinch is the leakage of plasma at the ends of the tube. End leakage of the plasma can be prevented by several methods.

The tube, for example, can be fashioned in the form of a closed ring, or *toroid*, which has no ends. Such installations have been built in several countries: Alpha in the USSR, Zeta in England and others. But the toroidal devices also have their shortcomings; here the plasma is pushed by the magnetic field to the outer wall of the tube.

End leakage of the plasma can also be prevented by employing *magnetic plugs*. This is the name given to regions set up at the ends of the tube with increased magnetic induction, in other words, with a high concentration of lines of induction. In the magnetic field the ions and electrons that make up the plasma move along helixes (see Sec. 41.7) winding around the lines of induction. Near the magnetic plug the velocity of longitudinal motion is reduced, the velocity of rotation increases and the ions, reflected from the magnetic plug, are returned to the middle part of the chamber. Magnetic plugs are used in the Soviet device Ogra, one of the largest installations of its kind in the world.

4. The chief shortcoming of all the existing installations is the shortness of the time they can confine the hot plasma. This results from various kinds of *instability* of the plasma, due to which the plasma filament begins to oscillate and bend. When it touches a wall, the plasma gives up a part of its energy and begins to cool. In addition, a part of the plasma always leaks through the magnetic plugs.

Today, various methods of stabilizing high-temperature plasma are being intensively investigated in some of the largest laboratories of the world. Quite a large degree of success has been achieved. But it is still impossible to obtain a plasma with a temperature $T \geq 10^6$ K, a concentration $n \geq 10^{20} \text{ m}^{-3}$ and a time of confinement $\tau \geq 0.1$ s.

Of interest is the fact that some of these parameters have been reached. Thus, in Ogra-1 a plasma with $T = 9 \times 10^6$ K and $\tau =$

$= 0.1$ s has been obtained but only for a low concentration: $n \approx 10^{14} \text{ m}^{-3}$. In devices with focussed pinch (in the USSR and the USA) a plasma with $T = 2 \times 10^5 \text{ K}$ and $n = 10^{26} \text{ m}^{-3}$ has been obtained but with a short time of confinement: $\tau \approx 2 \times 10^{-7} \text{ s}$. Evidently, these instabilities will be overcome in the course of time, and a plasma with the required parameters will be obtained. But when this will come about, and how, are matters that are far from clear today.

48.9. THE MAGNETOHYDRODYNAMIC GENERATOR

1. The process of converting the energy of fuel into the energy of electric current is quite a complicated one even in the most up-to-date steam electric power stations, and the efficiency is not over 30 per cent. In the firebox and boiler the energy of the fuel is converted into steam energy, and in the steam turbine this is further

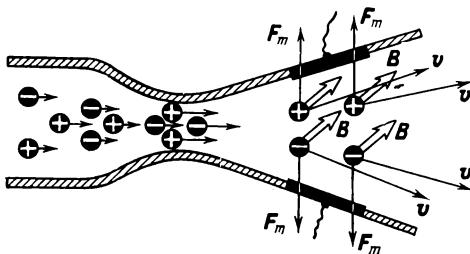


Fig. 48.8

converted into kinetic energy of the turbine rotor. The turbine rotor is coupled to the generator rotor so that a current is established in the stator of the generator. The chain of conversions of energy, firebox-boiler-turbine-generator, is a complicated one. The low efficiency is due chiefly to the turbine in which the temperature of the working medium (steam) is relatively low ($\leq 750^\circ\text{C}$), and also to the losses at all stages in the conversion of energy.

However, if the incandescent, and thereby ionized, gaseous products of combustion are made to move across a magnet field, an induced emf will be set up in this plasma, as in any conductor. This emf can be taken off the plasma by means of electrodes of proper design.

2. This idea lies at the basis of the *magnetohydrodynamic generator*. Its principle of operation is shown schematically in Fig. 48.8. Gas produced in the combustion of the fuel passes through the Laval nozzle (see Sec. 30.13) as a result of which its internal energy is converted into kinetic energy. As the gas moves through a strong cross magnetic field, an induced emf is set up. The positive ions move

toward the upper electrode, the free electrons toward the lower electrode. If we connect the electrodes to an external load, we obtain a current in the circuit.

To analyze the effectiveness of the generator we must evaluate its emf, and its electrical and thermal efficiencies.

3. The electrical efficiency is the ratio of the power consumed by the external load to the total power. Thus

$$\eta_{el} = \frac{i^2 R}{i^2 (R+r)} = \frac{R}{R+r}$$

where R = resistance of the external load

r = internal resistance of the generator.

The internal resistance of the generator can be reduced by increasing the electrical conductivity of the plasma. This can be done by raising its temperature. Another such measure is to make additions to the gas of the vapours of alkali metals, which have a low electron work function. Potassium salts are commonly used for this purpose; cesium is less frequently used because it is much more expensive.

The working temperatures in a magnetohydrodynamic generator are above 2000 K. At lower temperatures the conductivity is so low that the generator will not work.

4. An increase in the gas temperature has a favourable effect on the generator's thermal efficiency (see Sec. 29.5) which increases with the difference in the temperatures of the working medium and the environment. At temperatures of about 2000 K, the efficiency approaches 90 per cent (see Sec. 29.6). In addition, an increase in the gas temperature enables its velocity in the Laval nozzle to be increased. As a result, the emf is raised. But the wear of the walls of the nozzle is also sharply increased because at such high temperatures any material is intensively attacked.

5. The emf can also be increased by an increase in the magnetic induction of the field. For this purpose the magnetohydrodynamic generator is located in the gap of a powerful electromagnet. This magnet has an iron core and cooled windings. It can set up a field with an induction of 2 T, which corresponds to the magnetic saturation of iron.

An idea that shows promise in this field is to employ a magnet without an iron core, but with superconducting windings (see Sec. 75.10). The current density in such windings reaches 2×10^8 A/m² and the field induction may be as high as 6 T. For the time being, the application of superconducting windings is very complicated because it requires that the temperatures of liquid helium be maintained in large regions of space. Magnetohydrodynamic generators are still in the experimental stage. A semicommercial plant was developed in the Soviet Union in 1968. This magnetohydrodynamic generator provided a current for several hours. Evidently, such generators will find wide application in the future.

INDEX

- Abscissa, 21
Acceleration, 43
 angular, 210
 average, 43
 centripetal, 49
 free fall, 58
 of gravity, 58
 instantaneous, 43
 normal, 49
 radial, 49
Accelerator, 440
ADAMS, J. C., 78
Addition, vector, 37
 associative law of, 37
 commutative law of, 37
Adiabat, 268
 Poisson's quasi-static, 315
 shock, 315
Adiabatic process, 267
Adsorption, 365
Airfoil, 323
Airspeed, 324
Akulov-Bitter method, 463
Alfvén number, 531
Alpha thermonuclear installation, 637
Amorphous solid, 339
Amount of substance, 306
Ampere, 416
AMPERE, A. M., 453
ANDERSON, C. D., 439
Andrews isotherm, 373
ANDREWS, T., 373
Angström, 237
Anion, 509
Anisotropy, 336
Antiferromagnetism, 470
Archimedes' principle, 109
Area, midsection, 106
ARISTOTLE, 29
Astronomical unit, 83
Athodyd, 321
Atmosphere
 physical, 246
 standard, 246
 technical, 245
Atomization, cathode, 528
Attack, angle of, 323
AVENARIUS, M. P., 374
Avogadro's law, 254
Avogadro's number, 255, 294
Axis
 of coordinates, 21
 crystal, 336
 optic, 336
 rotational, 19
Bar, 244
Barkhausen effect, 465
BARNETT, S. J., 467
Barrier layer, 498
BERNAL, J. D., 355
BERNOULLI, D., 307
Bernoulli's equation, 307
BOGUSLAVSKY, S. A., 516
Bohr magneton, 454
BOHR, NIELS, 133
Boiling, 388
Boltzmann's constant, 249, 254
BOLTZMAN, L., 285
Bond, metallic, 346
Boundary layer, 105
Boyle-Mariotte law, 244, 373
Boyle's law, 224
BRAHE, TYCHO, 79
Brownian motion, 235, 290, 291
BROWN, R., 235
Bubble chamber, 389
Calorie, 154
Camber, 324
Capacitance, 400
Capacitor, parallel-plate, 399
Capillary tube, 364
Carnot cycle, 299
CARNOT, S., 299
Catalyst, 367
Cathode-ray tube, 518
Cathode rays, 529
Cation, 607, 509
Causality relationship, 149
CAVENDISH, H., 81
Chain reaction, 524
Characteristic
 current-voltage, 515, 521
 grid, 517
Charles' law, 264
Charge
 electric, 88
 elementary, 91, 406
 point, 88
 polarization, 407
Charge-to-mass ratio, 447
Chemical equivalent, 511
Chemical reaction, 187
Clausius principle, 289, 298
CLAUSIUS, R., 285
Coercivity, 462
Collision, 158
 central, 161
 completely inelastic, 158

- elastic, 160
 inelastic, 523
 Collision cross section, effective, 237
 Compensating process, 289
 Compressibility, 325
 isometric, 328
 Compression curve, 296
 Compression shock, 313
 Condensation centre, 387
 Conductance, 418
 Conduction
 heat, 191
 nonself-maintained, 520
 Conductivity, thermal, 503
 Conductor, 402
 Conservation
 of angular momentum, 207
 of electric charge, 88
 of energy, 177
 of energy, law of, 195
 of linear momentum, 139
 of mass, 195
 Constraint, 56
 Contact angle, 363
 Contact-potential difference, 499
 Continuity, equation of, 305
 Convection, 191
 Cooler, 296
 Coordinate, initial, 23, 71
 Coordinate-time plane, 25
COPERNICUS, NICOLAUS, 29
 Corkscrew rule, 428
 Correspondence principle, 125
 Coulomb, 91
COULOMB, C. A. de, 88
CROOKES, W., 529
 Cryogenics, 392
 Crystal, 336
 atomic, 345
 globular, 352
 ionic, 345
 metallic, 346
 molecular, 345
 Curie's law, 458
 Current
 conduction, 414
 direct, 416
 saturation, 516, 523
 Cycle, 296
 closed, 296
 Cyclotron, 440
- D**ebye law, 413
DEBYE, P. W., 502
 Debye temperature, 502
 Dee, 440
 Defect
 packing, 340
 substitutional, 342
 Deformation
 elastic, 51
 plastic, 51
 Degeneration of gases, 253
 Density, 60
 of charge, surface, 399
 critical, 374
 current, 416
 energy, 401
 Density gradient, 240
 Derivative, 28
 Deuterium, 163, 242
 Dew point, 376
 Dewar flask, 385
 Diamagnetic materials, 452
 Diamagnetism, 455
- Dielectric, 402, 406
 Dielectric constant, 406
 Diesel engine, 267
 Diffraction, X-ray, 339
 Diffusion, 238, 278, 341, 342, 358
 coefficient of, 240
 law of, 240
 Diode, 515
 Dipole, 92, 397
 Dipole arm, 92
 Dipole moment, induced, 330
 Discharge
 arc, 526
 glow, 526, 527
 self-maintaining, 526
 Dislocation, 340
 Displacement
 domain wall, 464
 electric, 408
 root-mean-square, 292
 Dissociation, electrolytic, 508
 Distribution
 barometric, 259
 molecular speed, 233
 Domain, 551, 463
DOPPLER, C. J., 133
 Doppler effect, 133
 Drag
 pressure, 107
 wave, 316
 Drude-Lorentz theory, 489
DRUDE, P., 489
 Dulong and Petit law, 501
DULONG, P. L., 501
 Dynamics
 basic problem of, 70
 fundamental law of, 62
 magnetogas, 529
 Dynamometer, 51
 Dyne, 69
- Efficiency, 297, 300
EINSTEIN, ALBERT, 32
 Einstein and de Haas experiment, 446
 Einstein's formula, 294
 Elasticity, coefficient of, 52
 Electric component, 426
 Electrification, 88
 Electrochemical equivalent, 511
 Electrolysis, 508
 Electrolyte, 508
 Electromagnetic interaction, 87
 Electromotive force, induced, 474
 Electron, 404
 collectivized, 346
 valence, 346
 Electron cloud, 514
 Electron drift, 487
 Electron gas, 489
 Electron gun, 518
 Electron volt, 154, 175
 Electrostatic image, 499
 Electrostatic shielding, 403
 Electrostatic unit, 91
 Energy
 activation, 357
 of configuration, 172
 of evaporation, 368
 internal, 184
 kinetic, 149
 mutual, 172
 potential, 172
 rest, 149
 surface, 362
 total, 148

- total mechanical, 177
 zero-point, 254
 Energy equation, 307
 Entropy, 285
 Entropy increase principle, 298
 Equation
 of motion, 22
 of state, 250
 Equivalence, principle of, 222
 Erg, 154
 Error, relative, 128
 Evaporation, 388
 Event, 121
 Expansion, thermal, 334
 Expansion curve, 296
 Expansion engine, 391

 Farad, 400
 FARADAY, MICHAEL, 393
 Faraday's constant, 511
 Faraday's law of induction, 479
 FEDOROV, E. S., 339
 Ferrimagnetism, 472
 Ferrite, 472
 Ferromagnetic material, 451
 Ferromagnetism, 459
 Field
 central, 224
 Coulomb, 414
 electric, 94
 electrostatic, 393
 gravitational, 84, 224
 uniform, 395
 Field strength, 393
 electric, 94
 magnetic, 433
 Field strength gradient, 398
 Filament, 513
 Five-fold symmetry, 355
 Flow, 305
 nonsteady, 305
 steady, 305
 Flow rate, mass, 305
 Fluctuation, 290
 Fluid, 304
 compressible, 306
 ideal, 306
 nonviscous, 306
 stagnant, 324
 Flux, mass, 305
 Forbiddenness principle, 212
 Force, 50
 Ampere, 449
 balancing, 54
 buoyant, 108
 centripetal, 62
 coercive, 462
 conservative, 171
 Coulomb, 90
 covalent, 345
 dissipative, 172
 elastic, 53
 electric, 88
 electromotive, 415
 equilibrium, 54
 external, 53, 139
 friction, 97
 generalized Lorentz, 437
 of gravitational attraction, 80
 of gravity, 56
 homopolar, 345
 inertial, 217
 internal, 139
 Lorentz, 436
 magnetic, 423
 molecular, 328
 moment of, 204
 propulsive, 143
 short-ranged, 332
 thermoelectromotive, 497
 van der Waals, 345
 viscous, 103
 Fourier's law of heat conduction, 503
 Frame, reference, 21
 FRANZ, R., 507
 Free fall, 57
 Free space, 20
 FRENKEL, Ya. I., 354
 Frequency, 48
 Friction
 angle of, 100
 dry, 97
 external, 97
 fluid, 97, 104
 internal, 97, 103
 kinetic, 97, 101
 rolling, 97, 102
 sliding, 97, 101
 static, 97
 viscous, 97
 Friction drag, 107
 Fuel, rocket, 144
 Fusion, heat of, 384

 GALILEI, GALILEO, 29
 GALLE, J. G., 78
 Galvanic cell, 512
 Gas
 electron, 346
 ideal, 247
 monatomic, 264
 Gas constant, universal, 256
 Gas-expansion machine, 391
 Gas mixture, separation of, 240
 Gauss, 428
 Gay-Lussac, law of, 265
 Geodesic, 225
 Geometry, non-Euclidean, 221
 GERLACH, W., 467
 Geiger counter, 524
 GLASER, D. A., 389
 G-load, 65
 GOUDSMITH, S. A., 469
 Gram, 69
 Graph, 23
 velocity, 45
 Gravitational constant, 81
 Gravitational prospecting, 86
 Gravity, force of, 56
 Grid, control, 518

 HAAS, W. J. de, 466
 Hall effect, 486
 HALL, E. H., 486
 Hartmann number, 531
 Heat
 of fusion, 356
 of melting, 384
 quantity of, 192
 Heat engine, 295
 Heat exchange, 191, 279
 convective, 191
 Heat flow, 503
 Heat pump, 303
 Heat sink, 296
 Heater, 296
 Heavy water, 163
 HEISENBERG, W., 134
 HELMHOLTZ, H. von, 404

- Henry, 480
 HERSCHEL, W., 77
 Hooke, R., 52
 Hugoniot equation, 315
 Humidity
 absolute, 376
 relative, 377
 Hydrodynamics, magnetic, 529
 Hygrometer, 376
 Hypersonics, 505
 Hysteresis, 461, 465
- Impact, 158
 Incidence, angle of, 323
 Incompressible fluid, 306
 Inductance, 480
 Induction
 electromagnetic, 473
 magnetic, 427
 Inertia
 centre of, 145
 law of, 30
 rotational, 200
 Instability, plasma, 533
 Institute of High-Pressure Physics, 350
 Intensity, gravitational field, 84
 Interaction, 50
 electromagnetic, 87
 strong, 213
 weak, 213
 International System of Units (SI), 68
 IOFFE, A. F., 344
 Ion avalanche, 524
 Ionization
 collision, 523
 thermal, 526
 Ionization chamber, 520
 Ionization coefficient, 520
 Isobaric process, 265
 Isochoric process, 264
 Isotherm, 268
 Isothermal process, 243, 266
 Isotope, 242
 Isotropy, 338
 Isovolumic process, 264
- Jet engine, 321
 Jet propulsion, 143
 Joint Institute for Nuclear Research, 444
 Joule, 154, 192, 421
 Joule-Lenz law, 421
- KAPITZA, P. L., 391
 KAUFMANN, W., 404
 Kelvin, 383
 KELVIN, Lord, 298
 KEPLER, J., 79
 Kilogram, 78
 Kilogram-metre, 154
 Kilogram-metre², 200
 Kilomole, 255
 Kurchatov Institute of Atomic Energy, 498
- LAMMERT, B., 231
 LANDAU, L. D., 379
 LANGMUIR, I., 516
 Lattice, crystal, 339
 Lattice constant, 339
 LAUE, M. von, 339
 Laval nozzle, 321
 LAWRENCE, E. O., 440
 Left-hand rule, 436
- LENARD, P., 404
 Length, proper, 121
 Leningrad Institute of Semiconductors, 498
 LENZ, H. F. E., 479
 Lenz law, 479
 LEVERRIER, U. B., 78
 Lifetime, residence, 356
 LIFSHITZ, E. M., 463
 Lift, dynamic, 323
 Line of force, 393
 Line of induction, 428
 Linear expansion, coefficient of, 336
 Liquation of gases, 391
 Liquid
 supercooled, 386
 superheated, 386, 389
 Liquid state, 354
 LOBACHEVSKY, N. I., 222
 Looping the loop, 179
 LORENTZ, H. A., 119
 Loschmidt's number, 254
 LOWELL, P., 78
 Lubrication, 105
- Mach cone, 313
 MACH E., 313
 Mach number, 311
 McMILLAN, E. M., 446
 Magnetic component, 427
 Magnetic plug, 533
 Magnetic Reynolds number, 530
 Magnetization, 454
 remanent, 462
 Magnetohydrodynamic generator, 534
 Manometer, inclined tube, 245
 Mass, 59
 centre of, 145
 measurement of, 142
 molar, 255
 proper, 126
 relative, 256
 relativistic, 126
 rest, 126
 Maxwell, 478
 MAXWELL, J. C., 234
 Mendeleev-Clapeyron equation, 256
 Metastable state, 361, 385
 Metre, 68
 MICHELSON, A. A., 111
 MILLIKAN, R. A., 404
 MILLIONSHCHIKOV, M. D., 498
 Mobility
 charge, 419
 particle, 487
 Moderating ratio, 162
 Modulus, bulk, 327
 Mole, 255
 Molecular beam, 232
 Molecular force graph, 331
 Molecular statistics, 279
 Molecule
 giant, 353
 polar, 411
 Molecule concentration, 237
 Moment, electric dipole, 92
 of force, 204
 of inertia, 200
 magnetic, 430
 spin magnetic, 467
 Moment arm, 204
 Momentum, 62, 126
 angular, 206
 orbital angular, 453
 sideways, 424
 Momentum equation, 306

- Monocrystal, 336
 Monomer, 351
Motion
 accelerated, 44
 bulk, 43
 compound, 43
 harmonic, 77
 mechanical, 19
 rectilinear, 22
 relative, 43
 rocket, 143
 rotational, 19
 translational, 19
 transport, 43
 uniform, 23
 uniform circular, 48
 variable, 24
Motion equation, 22
Neutron moderation, 162
Newton, 68
NEWTON, ISAAC, 29
Newton-metre, 204
Newton's laws of motion, 30, 53, 56, 62
Nozzle, 318
 converging-diverging, 321
Nuclear reaction, 188
Oersted, 433
Ogra thermonuclear device, 533
Ohm, 418
Ohm's law, 418, 419, 510
Oil-drop experiment, 404
Open cycle, 296
Ordinate, 21
Packing
 gas-crystal 353
 hexagonal close, 347
Parallelogram method, 38, 54
Paramagnetic materials, 452
Paramagnetism, 458
Particle, 19
 localized, 155
Pascal, 244
Pascal-second, 105
Path, mean free, 236
Peltier effect, 498
PELTIER, J.C.A., 498
Performance, coefficient of, 303
Period, 48
Permeability, magnetic, 455
Permeability constant, 427
Permittivity, relative, 406
Permittivity constant, 91
Perpetual motion, 298
PERRIN, J. B., 294
PETIT, A. T., 501
Phase transition, 378
 heat of, 384
Phonon, 505
Photoelectric emission, 500
Photoionization, 526
Pitot-Prandtl tube, 325
PLANCK, M., 285
Planck's constant, 134
Plasma
 gas-discharge, 527
 isothermal, 527
 ultradense, 376
Point
 fixed, 252
 ice, 252
lattice, 339
mass, 19
material, 19
reference, 252
steam, 252
triple, 383
Point mass, 19
Poise, 105
Poisson's equation, 269
Polarizability, 330
 deformation, 410
 orientational, 412
Polarization, 407
Poly crystal, 338
Polygon, force, 54
Polygon method, 38, 54
Polymer, 351
 bundle, 352
 network, 353
Polymorphism, 348
Positron, 438
Potential, 393
 cut off, 516
 electric, 174
Potential barrier, 182
Potential difference, 174
Potential energy curve, 180, 333
Potential energy well, 181
Potential trough, 500
Potentiometer, 515
POUND, R. V., 227
Power
 average, 153
 instantaneous, 153
Pressure, 164
 critical, 374
 full ram, 325
Pressure shock, normal, 314
Probability
 molecular distribution, 258
 thermodynamic, 281, 283, 284
Process, reversible, 277
Processes, thermodynamic, 264
Propulsion by reaction, 143
Protium, 242
Proton synchrotron, 444
Pseudonucleus, 356
Psychrometer, 378
PTOLEMEY, 31
Quasi-static processes, 276
Radar, 130
Radiation, 191
Ramjet engine, 321
Rays, anode, 529
Reaction engine, 321
Rebinder effect, 367
REBINDER, P. A., 366
REBKA, G. A., Jr., 227
Recoil phenomena, 141
Recombination, 509
 ion, 520
Recombination coefficient, 520
Reference frame, 21
 geocentric, 31
 heliocentric, 32
 inertial, 30
 noninertial, 64, 216
Refrigerator, 303
 thermoelectric, 498
Relativity
 principle of, 113
 Einstein's principle of, 32

- Galileo's principle of, 32
 Resistance, 418
 pressure, 106
 wave, 317
 Resistivity, 418
 Reverse Carnot cycle, 301
 Reynolds number, 108, 326
 Richardson-Dushman equation, 516
 RIECKE, E.V.C., 444
 Rolling friction, coefficient of, 102
 Romashka nuclear reactor, 498
- Scalar, 35
 SCHMIDT, O., 86
 Second, 68, 69
 SEEBECK, T., 497
 Self-diffusion, 239
 Self-induction, 481
 Semiconductor, 492
 Short-range order, 354
 Simultaneity of events, 113
 Six-fold symmetry, 346
 Sliding friction, coefficient of, 101
 SMOLUCHOVSKI, M., 236
 Solenoid, 434
 Solvation sheath, 509
 Source, heat, 296
 Space, "curved", 222
 Specific heat, 264, 501
 molar, 264
 Specific-heat ratio, 266
 Spectrum, mass, 449
 Speed
 most probable, 233
 root-mean-square, 243
 Spin, 469
 Stability, phase, 446
 Statcoulomb, 91
 State, two-phase, 379
 Static friction, coefficient of, 99
 Statistics
 molecular, 279
 Steiner's parallel-axis theorem, 202
 Stern-Gerlach experiment, 467
 STERN, O., 231
 STEWART, T. D., 484
 STOKES, G., 108
 STONEY, G. J., 404
 Streamlining, 107
 Stress, 327
 tangential, 99, 104
 Structure
 domain, 463
 ice, 351
 polycrystalline, 338
 Stuart number, 531
 Sublimation, 368
 Substance, working, 296
 Surface, equipotential, 394
 Surface-active substance, 365
 Surface potential jump, 499
 Susceptibility
 electric, 408
 magnetic, 455
 Symmetry, space-time, 213
 Synchronization, condition of, 442
 Synchrophasotron, 444
 Synchrotron, 444
 System
 adiabatically isolated, 194
 closed, 139, 194
 solar, 139
- Temperature
 absolute, 249
- critical, 360, 374
 Curie, 460
 degeneracy, 253
 ion, 527
 Néel, 471
 Temperature gradient, 503
 Tension, surface, 362, 389
 Tesla, 428
 Thermal process, 186
 Thermionic emission, 500, 513
 Thermocouple, 497
 Thermodynamic state variables, 276
 Thermodynamics, 193
 Thermoelectrons, 514
 Thermometer, gas, 250
 Thermopile, 497
 Theta-pinch, 532
 THOMSON, J. J., 404
 THOMSON, W., 298
 Thought experiment, 29
 Three-halves power law, 516
 Thrust, 143
 Time, 21
 dead, 526
 mean free, 291, 491
 proper, 122
 relaxation, 276, 422, 484
 Time constant
 capacitive, 422
 inductive, 484
 Time dilation, 229
 TOLMAN, R. C., 484
 TOMBAUGH, C. W., 78
 Topaz nuclear reactor, 514
 Torsion balance, 88
 Transformation
 Galilean, 33, 112
 Lorentz, 119
 Transition diagram, 379
 Triode, 517
 Tritium, 242
 Tsolkovsky formula, 144
 Turbocompressor jet engine, 322
 Twin paradox, 228
- UHLENBECK, G. E., 469
 Unit
 derived, 68
 fundamental, 67
 Unit cell, 339
 Units
 cgs electrostatic system of, 91
 engineering system of, 69
 physical system of, 69
 system of, 67
 Uncertainty relation, 134-
 Universal gravitation, law of, 80
- Van de Graaff electrostatic generator, 403
 Vaporization, 367
 Vapour
 saturated, 369
 supercooled, 386, 388
 superheated, 373
 supersaturated, 388
 Vector, 36
 displacement, 41
 radius, 36
 position, 36
 unit, 436
 VEKSLER, V. I., 446
 Velocities
 addition of, 42
 relativistic addition of, 118

- Velocity
 angular, 50
 average, 26, 41
 circular orbital, 144, 179
 escape, 145, 179
 exhaust, 143
 initial, 44, 71
 instantaneous, 28, 41
 of light, 110, 118
 of sound, 311
 space, 178
 terminal, 109
 wavefront, 308
- Velocity filter, 448
- Velocity gradient, 104
- Velocity range
 Newtonian, 129
 relativistic, 129
 ultrarelativistic, 129
- VERESHCHAGIN, L. F., 350
- Viscosity, 105, 359
- Volatilization, 368
- Volt, 175
- VOLTA, A., 495
- Voltaic cell, 512
- Volume, molar, 255
- Volume expansion, coefficient of, 335
- Watt, 155
- Wave
 bow shock, 313
 of elastic disturbances, 308
 oblique shock, 314
 shock, 309
- Wavefront, 308
- Weber, 478
- Weight, 56
- Weightlessness, 66
- WEISS, P., 463
- Wetting angle, 363
- White dwarfs, 375
- Wiedemann-Franz law, 507
- WIEDEMANN, G. H., 507
- Wilson cloud chamber, 142, 388
- WILSON, C.T.R., 388
- Work
 element of, 152
 path-independent, 175
- Work function, 499
- Working medium, 297
- Zero, absolute, 253
- Zero-point vibrations, 254
- Zeta thermonuclear installation, 533

TO THE READER!

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make.

Our address is:

USSR, 129820, Moscow I-110, GSP

Pervy Rizhsky Pereulok, 2

Mir Publishers

Printed in the Union of Soviet Socialist Republics